Marine Corps Operational Test and Evaluation Activity

Ground Combat Element Integrated Task Force

Experimental Assessment Report

August 2015, Original

Marine Corps Operational Test and Evaluation Activity
2032 Barnett Avenue
Quantico, VA 22134-5014

MOORE.KEITH.MICHAEL.1 (b)(6)

Digitally signed by MOORE.KEITH.MICHAEL (b)(6)
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USMC, cn=MOORE.KEITH.MICHAEL (b)(6)
Date: 2015.08.14 11:07:48 -04'00'

Approved: ______________________ 14 Aug 2015

K. M. Moore — Date
Colonel, USMC
Director, MCOTEA

~~FOR OFFICIAL USE ONLY~~
~~This document constitutes pre-decisional, deliberative opinion and recommendations.~~
~~Not releasable under FOIA pursuant to 5 U.S.C. 552 (b) (5).~~

THIS PAGE LEFT INTENTIONALLY BLANK.

Purpose

This Experimental Assessment Report (EAR) formally records MCOTEA's assessment of experimental results.

Background

The Commandant of the Marine Corps authorized the formation of a Ground Combat Element Integrated Task Force (GCEITF), and the Assistant Commandant of the Marine Corps assigned the Marine Corps Operational Test and Evaluation Activity (MCOTEA) the responsibility of conducting an experiment on the GCEITF. The task force trained female Marine volunteers in closed military occupational specialty (MOS) skills and integrated them into a combat arms unit, while a dedicated research team observed the unit's performance in an operational environment.

Scope

The task force was built around an infantry battalion minus (-) reinforced with attachments in the Battalion Landing Team model as a notional ground combat element component. This unit formed in July 2014 and remained in place until July 2015. MCOTEA used data obtained from GCEITF experimental events and non-experimental event periods to answer objectives identified in the Experimental Assessment Plan.

Conclusions

(~~FOUO~~) The female Marines integrated into the closed MOS units demonstrated that they are capable of performing the physically demanding tasks, but not necessarily at the same level as their male counterparts in terms of performance, fatigue, workload, or cohesion.

(~~FOUO~~) Integrated units, compared with all-male units, showed degradations in the time to complete tasks, move under load, and achieve timely effects on target. The size of the differences observed between units and tasks varied widely. The more telling aspect of the comparisons is the cumulative impacts. The pace, timing, and accuracy of any singular task is not necessarily important, but taken together, and in the context of actual combat operations, the cumulative differences can lead to substantial effects on the unit, and the unit's ability to accomplish the mission.

(~~FOUO~~) Gender and MOS type are the best predictors of occupational injuries. In particular, we found that females are more likely to incur occupational injuries, resulting in reduced readiness compared to their male counterparts. Males, on the other hand, are more likely to incur non-occupational injuries. Additionally, Marines in vehicle MOSs tended to have lower injury rates than those in MOSs that march (i.e., foot mobile) or Artillery MOSs.

(~~FOUO~~) No clear conclusions can be drawn from the Proficiency and Conduct ratings of the GCEITF volunteers.

THIS PAGE INTENTIONALLY LEFT BLANK

Experimental Assessment Report

Ground Combat Element Integrated Task Force

1. Purpose

This Experimental Assessment Report (EAR) formally records MCOTEA's assessment of experimental results.

2. Background

The expansion of roles for women in the armed forces has evolved over decades. Department of Defense (DOD) policy restricting women from serving in ground combat units was modified in 1994, and again in 2013. Under the 1994 policy, women could not be assigned to units below the brigade level whose primary mission was to engage in direct combat on the ground. This meant that women were barred from infantry, artillery, armor, combat engineer, and special operation units of battalion size or smaller. On 24 January 2013, then-Secretary of Defense Leon Panetta rescinded the Direct Combat Exclusion Rule on women serving in previously restricted occupations (i.e., combat). Today, women can serve in any military unit.

On 20 February 2014, the Commandant of the Marine Corps authorized the formation of a Ground Combat Element Integrated Task Force (GCEITF), and the Assistant Commandant of the Marine Corps assigned the Marine Corps Operational Test and Evaluation Activity (MCOTEA) the responsibility of conducting an experiment on the GCEITF. The task force trained female Marine volunteers in closed military occupational specialty (MOS) skills and integrated them into a combat arms unit, while a dedicated research team observed the unit's performance in an operational environment.

3. Scope

The task force was built around an infantry battalion minus (-) reinforced with attachments in the Battalion Landing Team model as a notional ground combat element component. This unit formed in July 2014 and remained in place until July 2015.

This report supports the assessment of the GCEITF from its inception through the final experimental exercises. MCOTEA used data obtained from experimental events (references (a) and (b)) and non-experimental event periods to answer objectives identified in the Experimental Assessment Plan (EAP) found in reference (c).

The event was limited to the ground combat element mission performed by Marines in closed and open MOSs. Closed MOSs consist of combat arms fields, including infantry (03xx), armor (18xx), and artillery (08xx). Open MOSs are all others.

4. Objectives

Standing up a GCEITF was intended to allowed us to evaluate the physical performance of individual Marine volunteers in the execution of individual and collective tasks in an operational environment. The purpose of the experiment was to estimate the effect of gender[1] integration in closed and open MOSs, in closed MOS units, and on various measures of readiness and mission success for these MOSs (e.g., physical capacity, fatigue, workload, tasks, etc.). The results of this assessment are reported here in the Experimental Assessment Report. The objective of this assessment was to address the following research questions:

4.1 Objective 1 (Closed MOSs)

In support of establishing gender-neutral occupational standards, we will measure mission effects at various levels of gender integration. We will also investigate:

4.1.1 Sub-objective 1.1

Whether the relationship between gender integration and mission effects varies by MOS.

4.1.2 Sub-objective 1.2

How gender integration affects gender distributions of fatigue and workload during collective tasks.

4.1.3 Sub-objective 1.3

Whether it is possible to isolate quantifiable individual physical characteristics that correlate to collective mission effects.

4.2 Objective 2 (Open MOSs in Closed MOS Units)

In support of establishing gender-neutral occupational standards for open MOSs in closed MOS units, we will measure closed MOS mission (i.e., provisional rifle company mission) effects at various levels of gender integration. We will focus on closed MOS units with male and female volunteers currently serving in open MOSs. We will also consider:

4.2.1 Sub-objective 2.1

How gender integration affects gender distributions of fatigue and workload during collective tasks.

[1] The term gender used throughout this report refers to the sex of the individual vice the spectrum of masculine and feminine traits as the modern use of the word has been referenced to.

4.2.2 Sub-objective 2.2

Whether it is possible to isolate quantifiable individual physical characteristics that correlate to collective mission effects.

4.3 Objective 3 (Personnel Readiness, Proficiency and Conduct)

In support of establishing gender-neutral occupational standards, we will work to identify physical characteristics that correlate to individual readiness, proficiency, and conduct. We will also study:

4.3.1 Sub-objective 3.1

How individual readiness, proficiency, and conduct correlate to mission effects.

5. Deviations

A number of deviations occurred from the published EAP. MCOTEA recognizes the following deviations from the EAP resulting from the GCEITF experiment conducted from 2 March 2015 to 18 May 2015:

5.1 Reduction of Treatment Groups

Marines participated in this experiment voluntarily and could leave the experiment at any time, for any reason, making the total number of human subjects available highly volatile. Trials were not executed if there were not enough participants to carry out the assigned mission and attrition affected the integration levels for the experiment. As a result, MCOTEA reduced the number of experimental treatment groups in cases where the total number of human subjects was not sufficient to conduct realistic operational missions at the small-unit level with all specified groups. The treatment groups were affected in the following MOSs:

- Infantry Rifleman (MOS: 0311)
- Infantry Mortarman (MOS: 0341)
- Infantry Assaultman and Infantry Antitank Missileman (MOS: 0351/0352)[2]
- Provisional Infantry Machine Gun (PIMG)
- Mountaineering (Closed MOSs)
- Mountaineering (Open MOSs)
- M1A1 Tank Crewman (MOS: 1812)

5.2 Proficiency and Conduct Analysis

The proficiency and conduct portion of objective 3 cannot be answered with any meaningful analysis of the data gathered from the GCEITF experiment. Objective 3 of

[2] From this point forward the 0351 and 0352 MOSs will be combined and referred to as 035X.

the research included an examination of proficiency and conduct information from volunteers who participated in the experiment in an attempt to correlate to individual proficiency and conduct to mission effects. However, the protections afforded to the volunteers in the plan bias the information in such a way as to make its use questionable at best. According to the EAP:

> *"Male and female open MOS volunteers participating in the provisional rifle company experiment as well as female volunteer participants for closed MOSs will not be performing in their normal MOS functions for the duration of their assignment to the GCE ITF. Duty proficiency marks will be assigned in accordance with the Individual Records Administration Manual (HQMC, 2000) and the commanding officer of the GCE ITF will provide due allowance for Marines who will be filling billets inconsistent with their grade and normal MOS skills."*

In practical terms, participation in the experiment cannot harm the participants, including their chances for promotion or retention. The special allowances necessary to protect the volunteers provided by the commanding officer of the GCEITF in the issuance of the proficiency and conduct ratings precludes their use in any meaningful analysis.

5.3 Modifications to Analysis

Although readiness for each Marine can be summarized as a ratio of the number of full duty days to the number of total duty days, we had much more detail on individual volunteer outcomes and we discuss the analyses of these outcomes. The total summary measure of readiness may obscure some of the more detailed outcomes, thus, we do not analyze it as a stand-alone result.

The original plan included analysis by integration level only; however, some tasks were performed by one or two persons who dominated the physically demanding aspect of the task. For those tasks we provide additional analysis by the gender.

5.4 Correlating Readiness to Mission Effects

In practical terms, when individuals became injured during the conduct of the experiment they were removed from trials and as such, we did not directly correlate the impact of readiness to mission effects.

6. Assumptions

Conclusions depend on the validity of assumptions. Assumptions will be validated throughout the evaluation process whenever possible. These assumptions include:

6.1 Physical Demands of Experimental Tasks

The collective tasks selected for the Marines in this assessment were considered to most physically demanding tasks that Marines could reasonably be expected to perform on a frequent basis. If gender integrated groups perform at least as well as all-male groups, then in theory, they would be capable of performing all MOS tasks.

6.2 Generalizability to Larger Units

Impacts on small (subordinate) unit effectiveness (i.e., Squads, Fireteams, Vehicle Crews, Gun Crews, etc.) are generalizable to larger unit effectiveness.

6.3 System and Equipment Representativeness

The systems and equipment used during the conduct of the experiment were representative (quality and type) of systems used by the Marine Corps in the conduct of its missions.

6.4 Volunteer Representativeness

Marines conducting the experiment were representative of the total population within the Marine Corps. This allows inferences of results and conclusions to be applied to future physical, physiological, and performance standards.

7. Limitations

7.1 Planned versus Actual Trials

The total number of Marine volunteers during the experiment was highly volatile for selected MOSs. Throughout the experiment, Marines were able to self-select to withdraw from the experiment (Drop on Request) or were dropped involuntarily for medical, legal, or command-related incidents. Trials were not executed if there were not enough participants to carry out the assigned mission. As a result, many tasks saw a reduction in sample size, some only running one third of the planned trials. Whenever reasonable, a comparison group (usually, the low-density group) was eliminated from consideration by the research team, splitting the smaller sample size between two different concentrations (all-male and high-density) rather than three to maintain better experimental power to detect differences. Still, with variation in some tasks particularly high among integrated groups, a larger sample size would have resulted in more precise estimates and more confident statements about observed differences. Each MOS-specific annex details the numbers of planned versus actual trials.

7.2 Relative Difficulty of Experimental Cycle

The assessment was designed to gather data associated with some of the most physically demanding tasks within each MOS. However, these tasks in isolation did not fully replicate life experienced by a Marine during a typical field exercise (FEX) or a combat environment and did not obtain the maximum cumulative workload that could be placed on a Marine. With limited time available, only selected tasks were assessed. Other tasks/duties outside the assessment were minimized due to specific experimental constraints and human factors. During a typical FEX or in a combat environment, it is common for Marines to conduct 24-hour operations that include day and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. Outside the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

7.3 Other Combat Arms MOSs

This event did not address the potential mission-effectiveness impacts of assigning women to the Reconnaissance Man (0321) and Critical Skills Operator (0372) MOSs. Given the particularly physically demanding nature of these MOSs, future studies should to examine the physically demanding tasks associated with these MOSs and conduct a similar assessment.

7.4 Limited Number of Tasks and Missions

This event did not address all potential tasks or missions of all closed MOSs, nor did it address all provisional rifle company tasks that female Marines in open MOSs assigned to previously closed units could perform. Instead, this event focused on what were considered the most physically demanding tasks that Marines could reasonably be expected to perform on a frequent basis. Additionally, tasks were chosen that were of a limited duration (maximum of several hours to complete) due to the requirement that tasks be repeatable in the form of experimental trials. In the event that more physically demanding tasks are identified in the future, additional studies may be required to draw appropriate inferences regarding the physical capabilities necessary to conduct the collective tasks.

7.5 Quantity of Volunteers in Select MOSs

The volunteers for the experiment were drawn from the active and reserve components of the Marine Corps. An examination of individual characteristics indicates they are a representative sample. Their representativeness, however, does not mitigate the fact that in some MOSs there is only a small quantity of males and females in these subgroups. In the extreme cases of the experiment, there were no more than three males (i.e., PIMG) and females (i.e., tanks) completing the entire experimental phase. Caution should be used when considering the generalizability of findings in these MOSs.

7.6 Volunteer Characteristics

The volunteer population was sourced and selected from the operating forces. Our sourcing of volunteers from the operating forces means accepting variations in some important respects, such as Time in Service, Time in MOS, training levels, and physiological development. We cannot be certain that male and female participants were totally equitable in these characteristics. All of the female Marines who participated in the Closed MOSs & 1371s had no operating force experience in ground combat units. Even with the training period prior to the experimental phase designed to mitigate differences in training and physiological development, some differences likely remained. Future studies with longer time horizons would benefit from recruiting participants at the same entry point, such as at the Marine Corps Recruit Depots.

7.7 Volunteer Selection Bias

The GCEITF experiment is likely to suffer from selection bias – the idea that the Marines' decision to participate (volunteer) in the experiment may be correlated with the specific traits that affect the study itself. This could manifest itself in the volunteers' motivation, performance, survey answers, and other aspects of the experiment. To the extent that we can, we try to ensure the sample was representative of the population of Marines, with respect to observable physical characteristics that the USMC collects and records.

7.8 Explaining Group Outcomes with Individual Characteristics

The experiment was designed to detect differences in performance of gender-integrated and all-male units. We also tried to determine how *individual* characteristics contribute to group outcomes in our attempt to inform gender-neutral standards. We do this using data on each individual volunteer in the group, sometimes combining variables, when sample size permitted. Ultimately, however, our models assume a very specific functional form: we assume that the group outcome depends on some linear combination of all participants' personnel variables. Intuitively, this need not be the case, and the poor fit of many of our models confirms that we gain little explanatory power by including personnel variables. This is not to say that the variables do not have an impact on the outcome, just that more complex modeling may be needed to uncover these relationships.

7.9 Statistical Independence

An important underlying assumption of some of the statistical analysis techniques we used, basic inferential statistics, is that the outcomes of each trial are statistically independent of each other. For many of the tasks, this assumption is justified because the number of volunteers was sufficient and their roles within tasks had little relationship to each other. Thus, when the same individual was sampled multiple times for the task but was performing different functions or was binned with different people, the data are considered sufficiently independent for the purposes of the experiment. In the case of provisional machine guns, tanks, and assaultman/antitank missileman that assumption is questionable. In these three cases, we had only three males (PIMG) and three females (tanks & 035X) who were able to complete the entire experiment. With so few people from a specific gender representing an MOS, we cannot be certain that each trial is necessarily independent. Caution should be used when considering the generalizability of results in these MOSs!

8. Methods

The specific methods for conduct of the experiment can be found in the EAP, reference (c). Annex R of this report describes the analysis methods used in the construction of the reported results. What follows in this section is a timeline and description of major experimental events and milestones as shown in Figure 1. This section also contains an explanation of the types of groups that formed the basis for comparison.

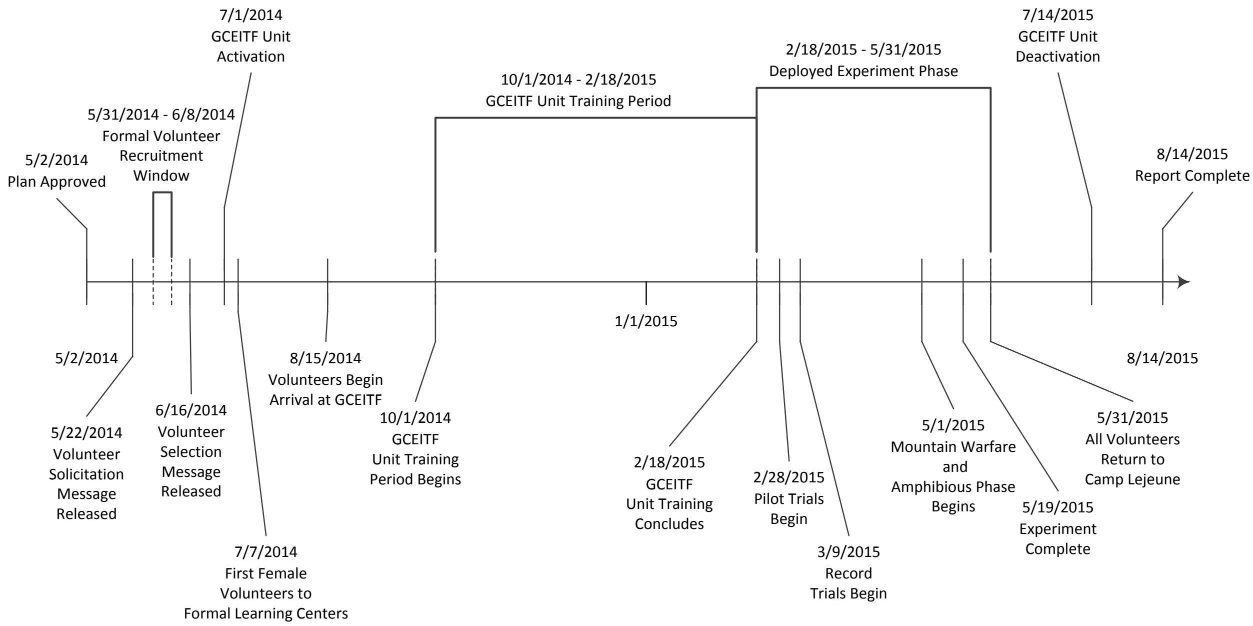


Figure 1. Timeline of Major Experimental Events and Milestones

8.1 Recruitment and Selection

The majority of recruitment, selection, and notification activities occurred during May through June 2014. Some recruitment took place after that timeframe, but prior to December 2014, to backfill volunteers who prematurely withdrew. Recruitment was done through formal in-person briefs, via phone, and web-based recruitment. Selection and notification occurred through formal message traffic.

8.2 GCEITF Unit Activation and Formal Schools

The GCEITF was formally activated on 1 July 2014 at Camp Lejeune, NC. During the activation period, the unit obtained facilities, equipment, and materiel necessary to conduct unit training prior to the experiment phase. Concurrent with unit activation was female volunteer attendance at Formal Learning Centers for MOS training. MOS training for female volunteers for combat arms MOSs occurred in multiple locations beginning July 2014 and concluding mid-October 2014.

8.3 GCEITF Unit Training Period

The GCEITF unit training period formally began on 1 October 2014 and continued until mid-February 2015 when the GCEITF arrived at the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, CA. During the unit training period, volunteers learned and rehearsed the live-fire and non-live-fire tasks necessary for experimental conduct.

8.4 Experiment Phase

The experiment phase began on 18 February 2015 with the arrival of the GCEITF at MCAGCC. The first 2 weeks allowed for volunteer acclimatization, range setup, and rehearsals of experimental tasks and data collection. Beginning 28 February 2015, volunteers began pilot trials. The pilot trials were identical to record trials with one exception; they were not used in the analysis. Pilot trials allowed the volunteers to work through learning-curve issues associated with performing the task in a new environment and in the presence of data collection. By 9 March 2015, record trials began for the experiment and trials continued throughout March, April, and into mid-May. During the conduct of record trials, various MOSs completed their formal experimentation phase at MCAGCC while others transitioned from MCAGCC to the Mountain Warfare Training Center (MWTC) and Marine Corps Base Camp Pendleton by 1 May 2015. All record trials were completed by 19 May 2015 at all locations and the GCEITF transitioned all personnel back to the unit's home station at Camp Lejeune, NC.

8.5 GCEITF Unit Deactivation and Reporting Period

Upon return to Camp Lejeune, the active duty volunteers were issued follow-on orders to new units. Volunteers from the reserve component were deactivated and returned to their home of record. The GCEITF also divested the remainder of equipment and materiel, and formally deactivated on 14 July 2015. Data reduction and analysis took place during the final deactivation and volunteer transition period with formal reporting of results completed by mid-August.

8.6 Small Unit Group Compositions

In support of research objectives, the MCOTEA team took measurements as participants of the GCEITF performed operationally relevant tasks in groups. The experimental plan specified that for each MOS, one experimental group would be all-male (control group), one would contain a low-density of females, and the third would be high-density. However, as volunteers left the experiment, some concentration levels had to be changed or dropped from evaluation altogether. For each MOS, trials were run for at least the control group and one of the integrated concentrations. The result of each trial for each of these units constitutes one observation for the collective tasks analysis. The final male-female concentrations, per MOS, are displayed in Table 1.

Table 1. Group Size and Female Concentration Levels by MOS

MOS	*Total in Group*	*Number of Females in Low-Density Group*	*Number of Females in High-Density Group*
0311	12	NA	2
0313	3	1	2
0331	3	1	2
0341	4	NA	2
035X	4	NA	2,3,4
PI	12	2	4,5
PIMG	3	NA	3
0811	6	1	2
1812	2	1	NA
1833	3	1	2
1371	8	2	4
Mountaineering Closed MOS	12	NA	6
Mountaineering Open MOS	12	NA	6

8.7 Small Unit Leadership

Each small group defined in Table 1 had some level of military leadership. The Marines in leadership positions were non-volunteers of sufficient rank and experience to ensure tactically sound and safe operations during the conduct of the experiment. As an example, the 0311 MOS had 12 Marine volunteers and a squad leader who was not a volunteer for a total of 13 Marines in the rifle squad.

The non-volunteer leadership Marines did not materially participate in the conduct of tasks. Even with non-participation, the influence of leadership can have an impact on the outcome of military effects. To mitigate this effect, we applied appropriate randomization of volunteers under different leaders as a control measure during conduct of the experiment.

9. Results

~~(FOUO)~~ This section of the report covers results for each experimental objective. Contained within each objective are the MOS tasks that begin with an operational view and task description followed by the results observed. The results are presented in terms of whether or not statistically significant differences are observed between groups, and occasionally by gender. In the majority of the results, the numerical details are left to the Annexes of this report, except where it was necessary to illustrate performance compared to an existing standard or to illustrate trends observed, as a results of basic inferential statistics. Unless otherwise stated in the results, we consider significant differences observed and reported to have at least some relevant operational impact, either by themselves or when taken together with other tasks in the context of actual combat operations.

9.1 *Objective 1*

9.1.1 Infantry

(FOUO) The infantry assessment was a 2-day scheme of maneuver intended to replicate offensive and defensive tasks. The infantry assessment included Infantry Rifleman (MOS 0311), Machine Gunners (MOS 0331), Mortarman (MOS 0341), and Assaultman/Antitank Missileman (MOS 035X). All Marines, regardless of task, wore the fighting load (35 lb) and carried a personal weapon (6.4 lb, 7.9 lb, or 10.5 lb, depending on weapon type).

9.1.1.1 Trial Cycle Day One

(FOUO) The first day of the infantry assessment was a coordinated, live-fire squad attack employing the rifle, machine gun, mortar, and assault squads. The coordinated attack allowed for simultaneous employment of all squads in close proximity on the same range for operational realism, but with offsets in battlespace geometries of fire to allow for independent observation and measurement of mission effects, fatigue, and workload. In general terms, the coordinated attack began with an initial movement, followed by an enemy engagement and consolidation on the objective, and concluded with a casualty evacuation. Day one was a fluid scenario that allowed the Marines to flow from beginning to end without interruption or breaks. Figure 2 is an illustration of the infantry live-fire attack. At the conclusion of the casualty evacuation the Marines operated under the guidance of their Company leadership, performing tasks that required minimal physical demands, and then bivouacking in the field, sleeping in 2-man tents.

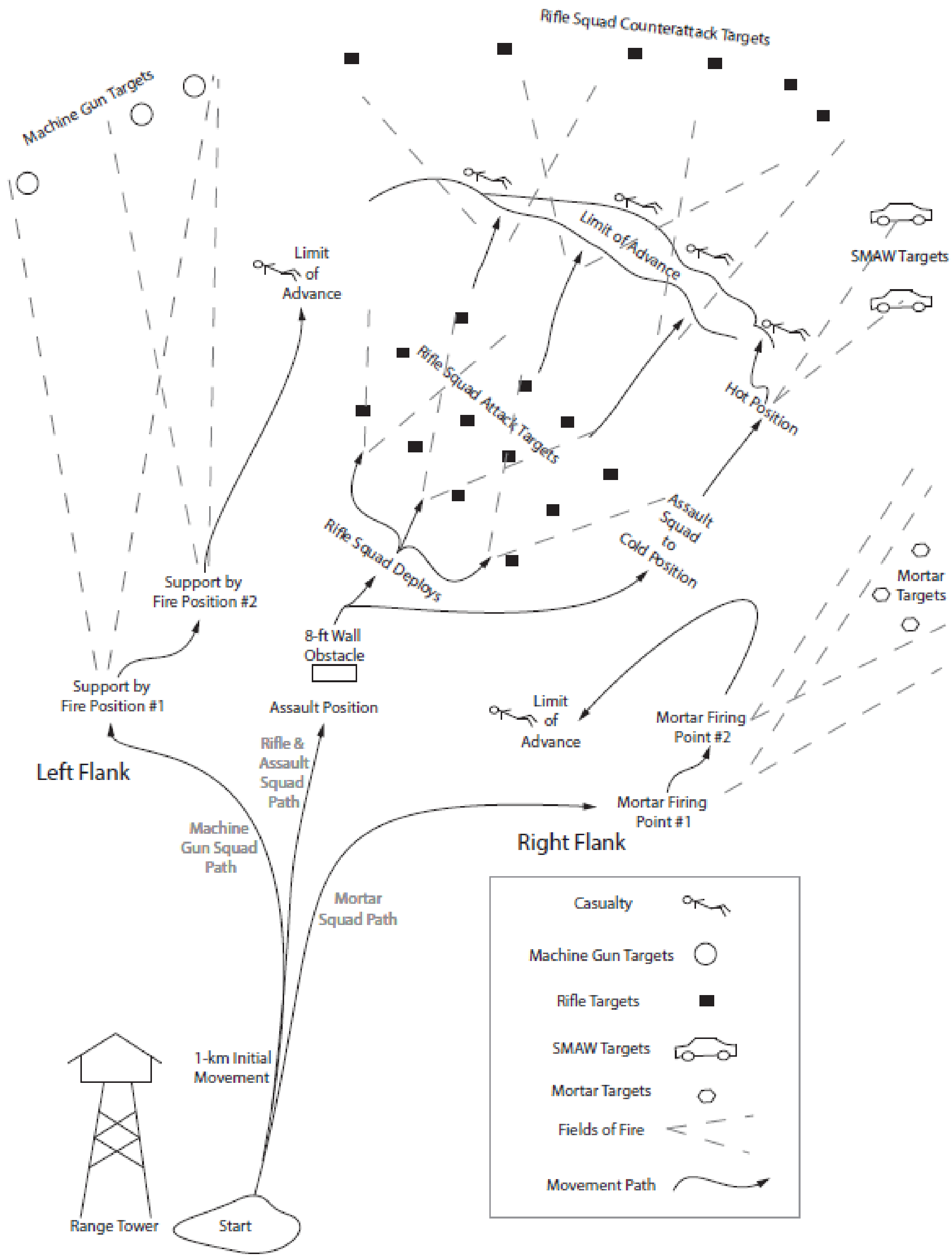


Figure 2. Infantry Day-One Scheme of Maneuver

9.1.1.1.1 Initial Movement

~~(FOUO~~) The assault began from an assembly area with a rapid movement to assault position. The initial movement was approximately 1 km to the mortar firing point (0341s), support by fire position (0331s), and assault position (0311s & 035Xs). Each squad moved this distance as quickly as possible while carrying an assault load (30 lb). In addition to the assault load, the following additional equipment was carried by non-0311 squads:

- 0331s: The M240B machine gun, tripod, spare barrel, and ammunition (600 rounds of 7.62 mm) divided among the three members of the squad (additional 28 – 40 lb per person)
- 0341s: Two 60-mm mortar tubes, two M8 auxiliary baseplates, and six 60-mm mortar rounds divided among the four members of the squad (additional 28 – 35 lb per person)
- 035Xs: Two SMAW launchers and four rockets divided among the four members of the squad (additional 26 – 42 lb per person)

The all-male squads, regardless of MOS, were significantly[3] faster than the integrated squads. The differences observed between all-male and integrated squads were more pronounced in non-0311 squads that carried the assault load plus the additional weight of crew-served weapons and ammunition. Figure 3 shows the percent increase in average times when comparing integrated groups to all-male groups. Only the 0331 squads had low- and high-density squads to compare to all-male squads. The remaining MOSs had only enough volunteers for high-density squads to compare to all-male squads.

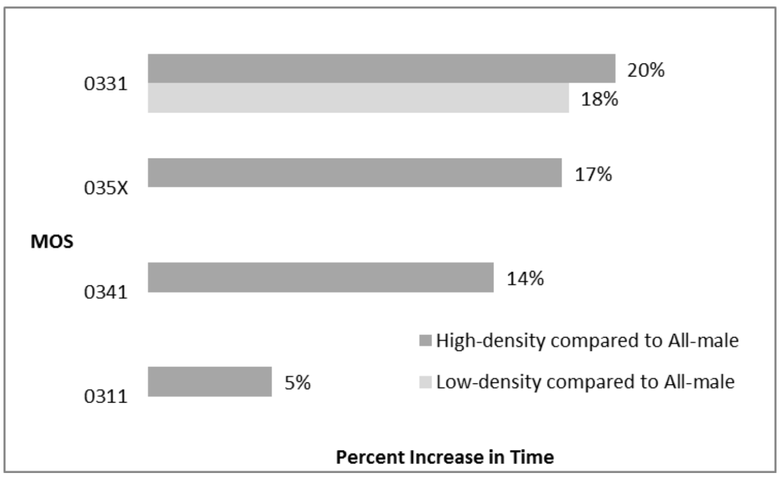


Figure 3. Percent Increase in Initial Movement Times for Integrated Squads Compared to All-male Squads

9.1.1.1.2 Negotiate Obstacles

~~(FOUO)~~ Upon arrival at the assault position, the 0311s and 035Xs Marines were required to negotiate an 8-ft wall by getting all Marines and equipment over as quickly

[3] In this report the term “significant” is shorthand for statistically significant. In layman’s terms, we say that something is statistically significant if it has a low likelihood of being due to random chance or error. The cut-off point for that likelihood is 10%. In technical terms, that means statistically significant with at least a one-sided p-value ≤ 0.10.

as possible. The squads were allowed to negotiate the obstacle any way they chose, as long as they maintained the appropriate tactical posture. The differences that resulted when comparing 0311s and 035Xs are largely a result of the technique used and the quantity of individuals within a squad (12 for Rifle Squads, 4 for Assault Squads).

When comparing integrated 0311 squads to all-male 0311 squads we saw no significant differences in times. There was, however, an issue with getting assault packs over the wall. Prior to negotiating the wall, 0311 Marines doffed their assault packs and individually threw them onto the top of the 8-ft wall prior to climbing. Females in integrated squads were noted as having required assistance from male squad members to get their packs onto the wall. Male Marines in the squad did not require the same assistance.

Unlike the 0311 integrated squads, the integrated 035X squads took significantly longer to negotiate the wall obstacle than all-male 035X squads. The assault squads were formed by combining teams, each with two Marines. When both teams in a squad were all-female, the 035X squad used their belts as a ladder to help the last Marine over the obstacle. When there was at least one male in the high-density squad, this technique was not required.

9.1.1.1.3 Fire and Movement

~~(FOUO~~) While the 0311s and 035Xs were negotiating the wall obstacle, the 0341s and 0331s were occupying their respective mortar firing points and a support-by-fire position to suppress enemy targets and prepare the objective for the rifle squad.

9.1.1.1.3.1 Indirect Fire - Mortar Squad

~~(FOUO)~~ The 0341s on the right flank typically began the attack by using indirect fire to suppress the enemy and to prepare the objective for the rifle squad to locate, close-with, and destroy the enemy by fire and maneuver. When the attack was conducted, 60-mm mortars were employed in the handheld method. Each two-Marine mortar team engaged one target from two different mortar-firing points with three rounds per point and finally moved to a limit of advance. In this attack, the integrated 0341 squads were significantly slower than the all-male 0341 squads at providing indirect fires. The attack started with integrated and all-male 0341 squads performing similarly at the first mortar-firing point, but as the scenario continued, the differences among group results emerged, with the integrated groups falling behind (i.e., taking longer) at the second firing point and on the movement to the limit of advance.

9.1.1.1.3.2 Suppressive Fire - Machine Gun Squad

~~(FOUO~~) The 0331 machine gun squad operated on the left flank and their timing for movement closely followed the 0341 squads. The machine gun squad occupied a position of overwatch and provided support-by-fire for the rifle squad. From the support-by-fire position, the machine gun squad had to rapidly get its gun into action, acquire targets, and accurately engage the enemy.

The movement and emplacement of the machine gun took significantly longer with both low- and high-integrated groups when compared to all-male groups; but there was no significant difference in effects (i.e., hits and near misses) on target once emplaced.

9.1.1.1.3.3 Rifle Squad Attack and Counterattack

(FOUO) The mission of the rifle squad is "to locate, close-with, and destroy the enemy by fire and maneuver, and repel the enemy's assault by fire and close combat." Having negotiated the 8-ft wall obstacle, the rifle squad deployed into an on-line formation to maximize firepower in the direction of the enemy and then moved with weapons at the ready expecting imminent contact. Upon contact with the enemy (targets being presented), the Marines established a heavy volume of fire to gain fire superiority. They then began buddy rushes to close-with the enemy objective. Each squad conducted approximately 300 meters of buddy rushes, engaging a total of 12 targets with direct-fire weapons and three machine gun bunkers with the M16A4 with M203 grenade launcher attached. After all targets were destroyed (no targets remaining), the squad moved another 125 meters to a limit of advance and prepared for an enemy counterattack.

(FOUO) At the conclusion of every assault, the attacking unit must be prepared for the enemy to regroup and organize a counterattack. Upon consolidation at the limit of advance, the squad oriented its weapons in the direction of the enemy's retreat. Upon target presentation, the counterattack commenced for 90 seconds.

(FOUO) The principle measurement of the squad's ability to destroy the enemy by fire during the attack and counterattack was hits on target. When comparing integrated 0311 squads and all-male squads, the integrated 0311 squads had a significantly lower probability of hit. The squad attack involved all 12 Marines engaging targets. Each Marine can, and often does, engage some of the same targets because of overlapping sectors of fire. To uncover whether or not a gender effect existed in accuracy at the individual level, individual shots recorded by targets were correlated back to shooters. This enabled an investigation of gender differences by weapon system (i.e., M4, M27, and M16A4/M203).

(FOUO) Analysis of shot accuracy by gender and weapon reveals that there is a significant difference between genders for every weapon system within the 0311 squads, except for the probability of hit & near miss[4] with the M4.

[4] Hits are shots that impact the E-silhouette target. Near misses are shots that are detected within 1-meter of outside edge of E-silhouette target that may have some suppressive effect on the enemy.

(~~FOUO~~) Figure 4 and Figure 5 depict accuracy results for males versus females by weapon type. The overall accuracy declined and the percent difference increased as the weight of the weapon system increased. The M4 was the lightest weapon and yielded the most accurate results, while the M16A4/M203 was the heaviest weapon and yielded the least accurate results.

(~~FOUO~~) One might think that experience level of the Marines (male vs. female) influenced the results observed in the 0311s. However, the male Provisional Infantry (PI) results do not support this conjecture. PI males had no infantry experience and, in this regard, were similar to female 0311s prior to the experiment. The PI males had a further disadvantage because they had not completed the 0311 MOS course. Despite these differences, the PI males had higher hit and near-miss percentages than the 0311 females, as shown in Figure 4 and Figure 5.

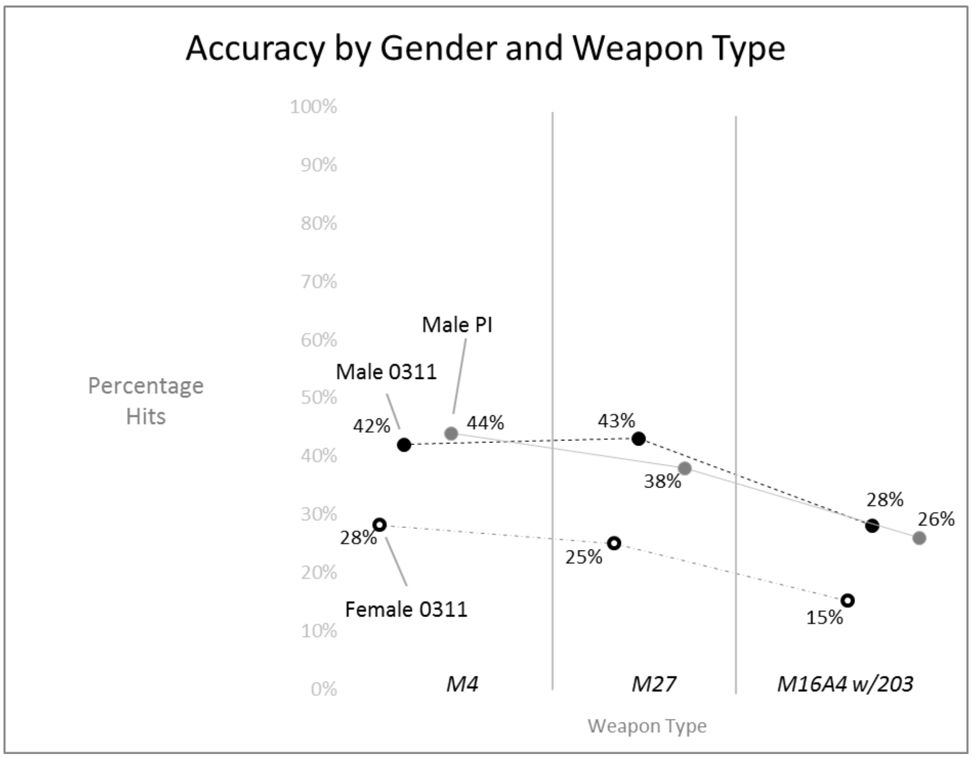


Figure 4. Accuracy Comparison by Gender and Weapon Type for Hits

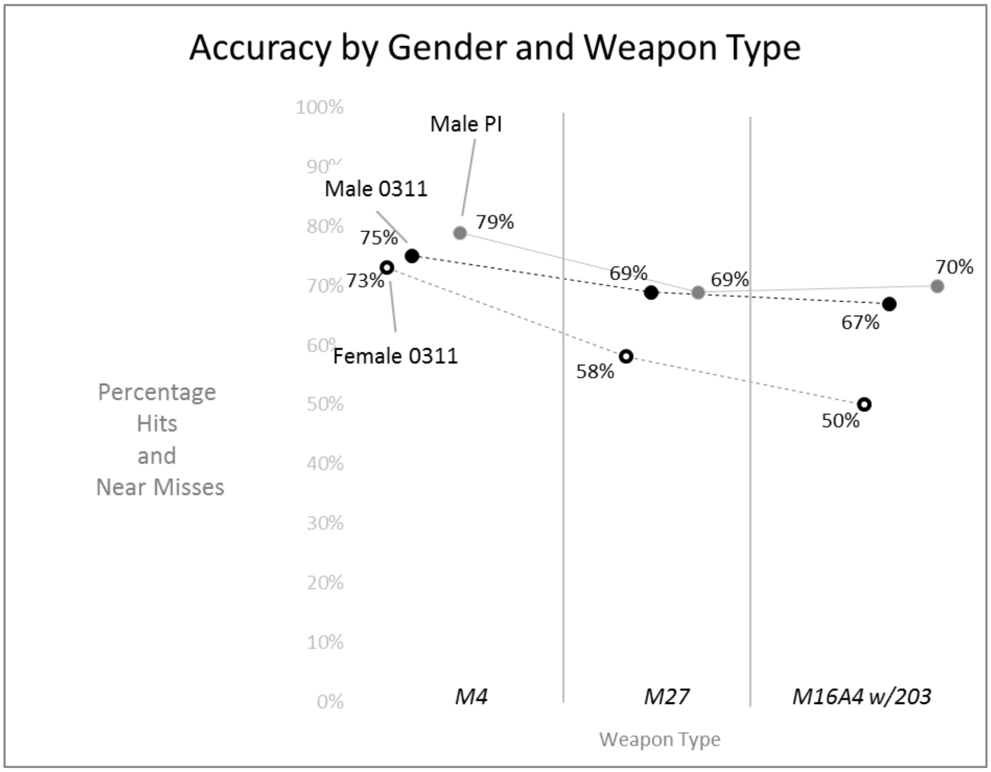


Figure 5. Accuracy Comparison by Gender and Weapon Type for Hits and Near Misses Combined

9.1.1.1.3.4 Assault Squads

(FOUO) An assault squad often accompanies a rifle squad into an attack. The assault squad provides high-explosive rocket fires to reduce fortifications or destroy other designated targets. When employing the Shoulder-launched Multipurpose Assault Weapon (SMAW), the squad prepared their rockets from a covered position, commonly referred to as a cold position. Then they moved to an exposed position, approximately 50 meters, with a good line of sight to the enemy, commonly referred to as a hot position. Once at the hot position, they engaged an armored vehicle, conducted a rapid reload, engaged a second armored vehicle, and displaced back to the cold position.

(FOUO) The integrated 035X squads took significantly longer to engage targets and got significantly fewer hits on target as compared to the all-male squads during the engagements. The combination of the two results makes the integrated squads more vulnerable and less lethal than their all-male counterparts.

9.1.1.1.4 Movement to Limit of Advance

(FOUO) The ability to close with the objective after conducting a live-fire attack is crucial to maintaining momentum during offensive operations. At the conclusion of the live-fire attack, the 0311 squads moved 125 meters to a limit of advance. The 0331 and 0341 squads also conducted a displacement to a limit of advance, but the distance was 300 meters. The integrated 0311, 0331, and 0341 squads took significantly longer than their all-male counterparts to complete their movement to the limit of advance. The late arrival reduces the time to prepare for, or respond to, the counterattack launched by an enemy force.

9.1.1.1.5 Casualty Evacuation

(FOUO) Casualties are an inevitable part of conducting combat operations. Units train to become proficient in the triage, handling, and transport of a casualty. After a Marine is injured, it is essential to move the casualty to the appropriate level of care as quickly as possible. Each 0331, 0341, and 035X squad was assigned one casualty to evacuate upon conclusion of the live-fire attack and counterattack. Rifle squads (0311) had 12 Marines in a squad as compared to 4 or 3 for the other squads. Because the rifle squads had three times the number of Marines, each rifle squad had three casualties to evacuate (i.e., one casualty for every three to four Marines).

(FOUO) The casualty was a dummy[5] wearing combat equipment that had to be moved a distance of 100 meters to a casualty collection point along with the weapons and gear of the Marines performing the transport. The Marines could use a variety of techniques for transport, but had to carry the dummy off the ground (not drag any part) and could not drop the casualty.

(FOUO) The differences between integrated squads and all-male squads vary substantially by MOS (Figure 6). The 0341s showed no significant differences in evacuation times while the 035Xs showed the largest significant observed difference in the entire experiment. The only commonality across the casualty evacuation for the four MOSs was the casualty itself. Each MOS had one or more techniques for carrying out the evacuation that depended upon the abilities of the Marines carrying out the task.

[5] The dummy was a full-scale anthropomorphic test device that simulated the dimensions, weight proportions, and articulations of the human body. The dummy's weight was equivalent to an average Marine (174 lb). When combat equipment (Marine Corps Combat Utility Uniform, boots, Kevlar helmet, plate carrier vest with small arms protective insert plates, and M16A4) was added to the dummy, the total weight increased to 220 lb.

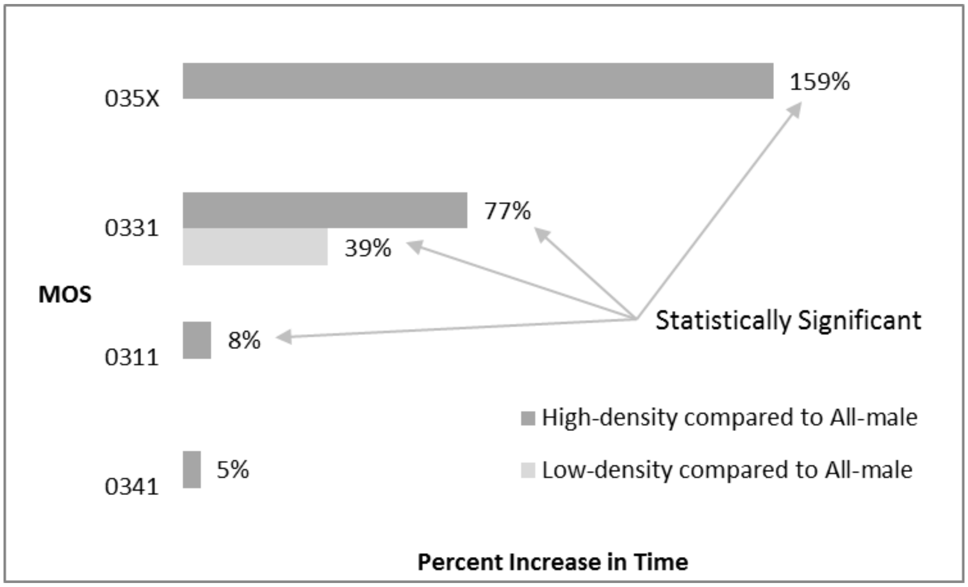


Figure 6. Percent Increase in Casualty Evacuation
Times by MOS and Integration Level

9.1.1.1.5.1 Mortar Squad Casualty Evacuation

(~~FOUO~~) The 0341s had the smallest difference in average time between integrated and all-male squads, and the difference in time was not significant. However, the method of employment used by this group explains why differences in times may not be significant. The 0341s largely employed a single-Marine fireman's carry to move the casualty. Furthermore, in a majority of the integrated trials, the person carrying the casualty using the single fireman's carry technique was male, not female.

9.1.1.1.5.2 Rifle Fireteam Casualty Evacuation

(~~FOUO~~) The 0311 results were analyzed by 4-person fireteams instead of by squad for this task because the effort was conducted independently between fireteams. The integrated fireteams were significantly slower than the all-male fireteams. At the discretion of each fireteam, Marines used a 2-Marine, 3-Marine, or 4-Marine carry to move the casualty.

9.1.1.1.5.3 Machine Gun Squad Casualty Evacuation

(~~FOUO~~) The all-male 0331 squads were significantly faster than high-density integrated squads, but not significantly faster than low-density integrated squads. The 0331 squads used a mix of techniques to carry the casualty, which explains some of the variability in the observed results. The most prominent explanation of why differences may not be significant between all-male and low-density 0331 squads is the use of the single fireman's carry. The top one-third of results of the low-density squads were almost exclusively male fireman's carry results.

9.1.1.1.5.4 Assault Squad Casualty Evacuation

(~~FOUO~~) The largest difference observed in the experiment occurred in the 035X casualty evacuation where the all-male 035X squads were significantly faster than the integrated squads. The 035X squads used a mix of techniques to include a 2-Marine, 3-Marine, and 4-Marine carry. In this task, like other 035X tasks, the all-male squads were almost always compared to all-female squads.

9.1.1.1.6 Infantry Day One Fatigue and Workload

9.1.1.1.6.1 Comparing Male Responses to Female Responses

(~~FOUO~~) Two self-report surveys (fatigue and workload) were administered to all volunteers at the end of trial activities on day one of the trial cycle. The surveys were intended to provide insights into Marines' perceptions of their own level of fatigue and workload level during the course of offensive trials.

(~~FOUO~~) Overall, males tended to report lower levels of fatigue and workload when compared to females in the 0311, 0331, and 035X squads. In 0341 squads, females actually reported lower levels of fatigue and no differences in workload when compared to males.

9.1.1.1.6.2 Comparing Male Responses in Integrated Squads to All-male Squads

(~~FOUO~~) In addition to comparing male and female fatigue and workload self-reports, we also examined whether responses from males remained consistent at different integration levels.

(~~FOUO~~) The reported fatigue levels of males did not appear to change depending on integration of the squad with the exception of 0331 males. Males in high-density integration squads actually reported lower levels of fatigue than when they participated in all-male or low-density integrated squads.

(~~FOUO~~) Male opinions on workload were more mixed than male fatigue results. Males in 0311 and 0331 squads reported lower workload results when in integrated squads when compared to all-male squads. Males in 0341 and 035X squads reported no differences based on integration level.

9.1.1.2 Trial Cycle Day Two

(~~FOUO~~) The second day of the infantry assessment focused on an approach march for all squads, defensive actions of the rifle squad, and heavy weapons employment of the machine gun, mortar, and anti-armor squads, as shown in Figure 7. At the conclusion of the defense and weapons employment, the Marines operated under the guidance of their Company leadership, performing tasks that required minimal physical demands, and then bivouacking in the field, sleeping in 2-man tents.

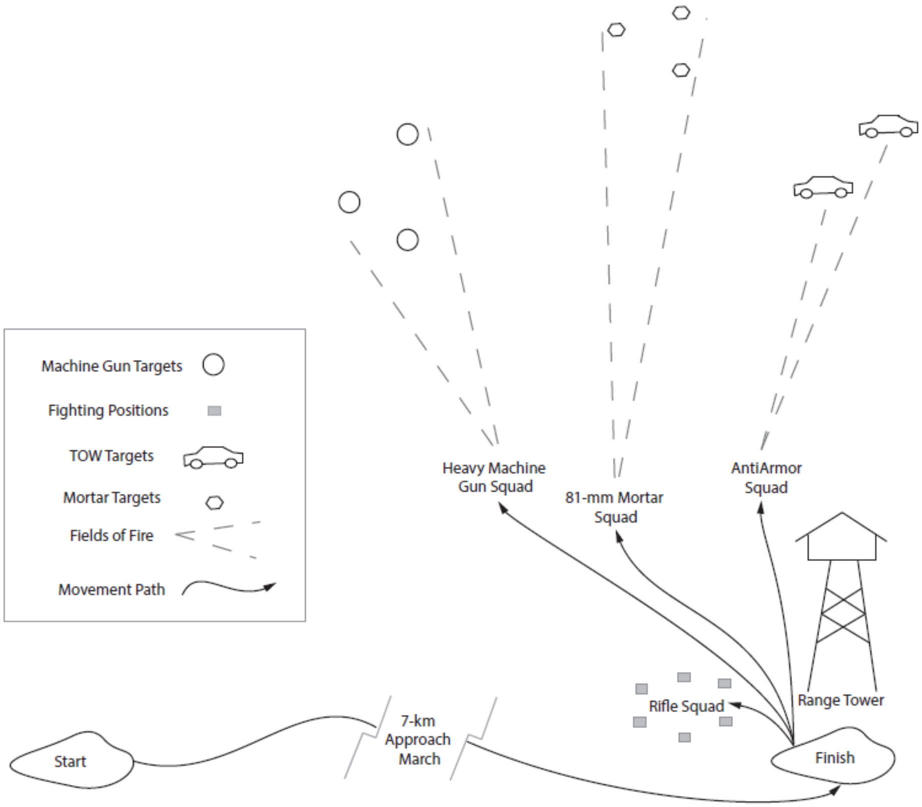


Figure 7. Infantry Day-Two Scheme of Maneuver

9.1.1.2.1 Approach March

(FOUO) The beginning of day two started with a march carrying the approach load (55 lb)[6]. Moving under a load is one of the most fundamental tasks of an infantry unit; it is both physically and mentally demanding. Units train by conducting tactical marches with an approach load at increasing distances. The Infantry Training & Readiness Manual states: "the approach march load will be such that the average infantry Marine will be able to conduct a 20-mile hike in 8-hours with the reasonable expectation of maintaining 90% combat effectiveness." Each squad moved a distance of 7.20 km carrying the approach load and personal weapons. In addition to the approach load, the following additional equipment was carried by non-0311 squads:

- (FOUO) 0331s: The M240B machine gun, M122A1 tripod, spare barrel, and ammunition (4 cans of 7.62 mm) divided among the three members of the squad (additional 28 – 40 lb per person).

[6] The approach load is in addition to the fighting load (35 lb) and personal weapon (6.4 lb, 7.9 lb, or 10.5 lb depending upon weapon type assigned).

- (FOUO) 0341s: 81-mm mortar tube, baseplate, bipod, and box sight divided among the four members of the squad (additional 28 – 35 lb per person).
- (FOUO) 035Xs: Two SMAW launchers and four rockets divided among the four members of the squad (additional 26 – 42 lb per person).

(FOUO) The all-male squads, regardless of MOS, were significantly faster than the integrated squads. The differences observed between all-male and integrated squads are more pronounced in non-0311 squads, which corresponds to an increase in carried weight of crew-served weapons and ammunition in addition to the approach load. Figure 8 shows the percent increase in average times when comparing integrated groups to the all-male groups.

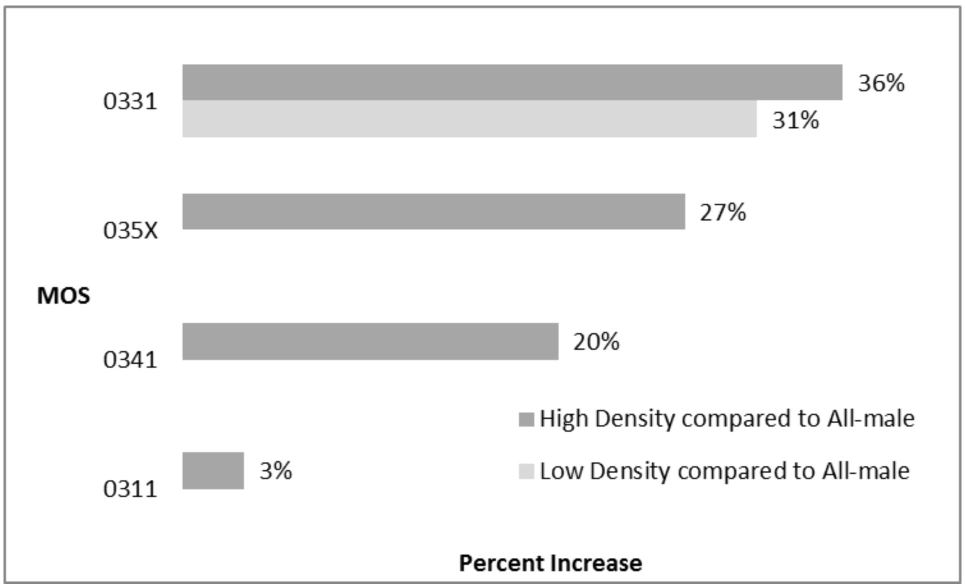


Figure 8. Percent Increase in Approach March Times for Integrated Squads Compared to All-male Squads

(FOUO) The differences illustrated in Figure 8 follow the same pattern observed in the initial movement. The march with approach load has an applicable standard that can be used to provide a second frame of reference. Movement over 20 miles in 8 hours equates to a pace of 4 kilometers per hour (km/h). Figure 9 shows a conversion of the average squad times over the 7.2-km distance to pace. This conversion enables a comparison to the 4-km/h required to meet the standard. As shown in the figure, the only integrated squad to achieve the standard is the integrated (high-density) 0311 squad. Regardless of MOS, the all-male squads surpassed the 4-km/h standard. The figure is also illustrative of the impact that additional weight has on pace. There is a corresponding decrease in the pace as weight of the carried load increases.

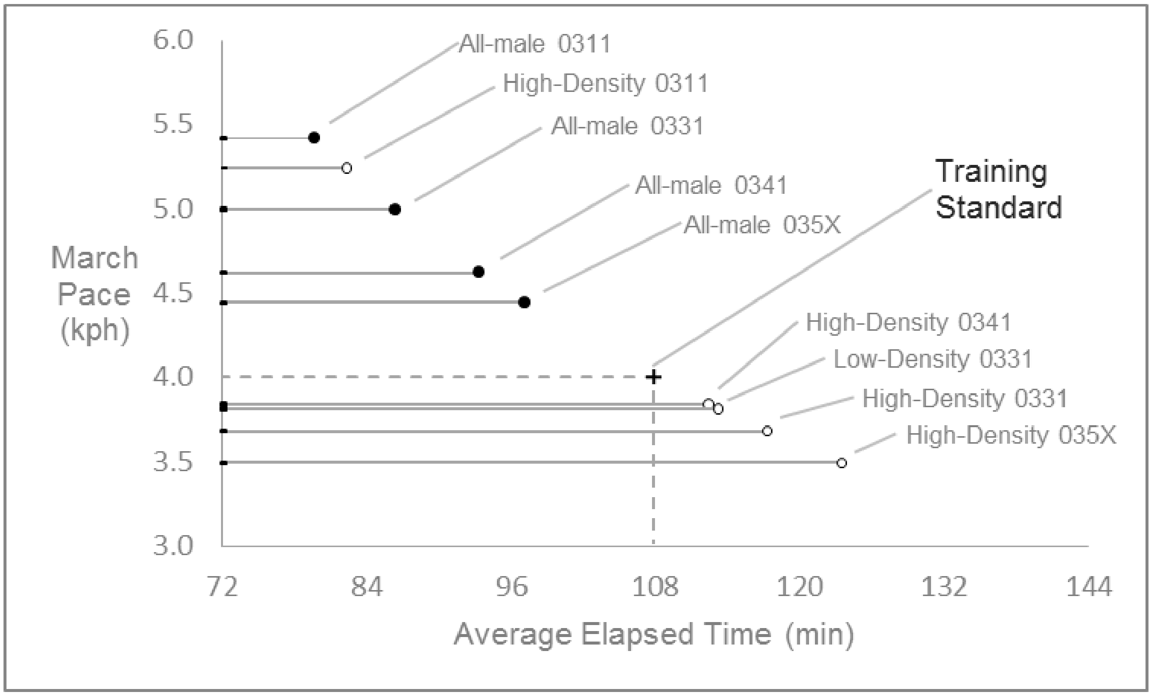


Figure 9. Conversion from Average Elapsed Time for the 7.2-km March to Pace for the 03XX Approach March

9.1.1.2.2 Defensive Fighting Positions & Heavy Weapons Employment

(FOUO) Upon completion of the approach march and after a brief rest period, the infantry moved into the defense. The 0311 Infantry Marines moved into the preparation of defensive fighting positions while the other MOSs (0331, 0341, and 035X) moved into heavy weapons employment.

9.1.1.2.2.1 Fireteams Preparing Defensive Fighting Positions

(FOUO) Protection is vital to infantry units, especially in a defensive posture. Infantry units commonly construct fighting positions to conceal positions and minimize exposure to enemy fire. There are many characteristics to constructing a doctrinal fighting position, but the most physically demanding aspect is digging out the fighting hole. The terrain where the fighting holes were dug consisted of hard, compact sand and rocks. Each 0311 fireteam dug two 2-man fighting holes in a time limit of 2 hours, maintaining 50% security, meaning two Marines dug while two provided security in the prone position. The Marines swapped positions every 15 minutes. During the course of this task, the earth was scooped into buckets and weighed.

(FOUO) The average weight of earth moved by the integrated 0311 fireteams was not significantly different than the weight moved by the all-male groups.

9.1.1.2.2.2 Squad Employment of M2 Machine Gun

(FOUO) Providing defensive fires with the M2 heavy machine gun entails moving the system to a position of advantage, engaging the enemy, and conducting a rapid displacement. After completing the approach march, the 0331 squads dismounted the M2 heavy machine gun from a vehicle platform, moved and emplaced the weapon, engaged targets, displaced, and mounted the weapon onto a vehicle platform.

(FOUO) To dismount the weapon system, the squad worked together to manually disassemble and lower each component to the ground from a High-Mobility Multi-Wheeled Vehicle (HMMWV). To mount the weapon system on the HMMWV, the process was reversed. This task determined the time for a squad of three Marines to fully dismount and mount the M2 from a tactical vehicle and required the strength to lift, manipulate, and lower heavy components.

(FOUO) The integrated 0331 squads' average time was not significantly different than the average time for all-male groups, for either mounting or dismounting the M2 heavy machine gun.

(FOUO) During the assessment, each machine gun squad emplaced the M2 on a tripod at a specified firing location. The squad engaged targets as they began to be exposed while hits on, and around, the target were recorded by sensors. The assessment began with targets exposed and engaged by the squad.

(FOUO) The high-density integrated 0331 squads had significantly more hits on target than the all-male squads. The low-density 0331 squads also had more hits on target than the all-male squads, but the difference was not significant. Figure 10 shows the percentage of hits and near misses[7] on the machine gun targets. The finding that integrated groups have a higher percentage of hits on target is somewhat unusual given the findings from 0311 hit percentages. To see if there is a consistent finding, the results of provisional machine gunners were compared to the 0331 squads. As shown in Figure 10, the trend of integrated groups outperforming all-male groups is repeated and consistent.

[7] For experimental purposes, a hit or near miss for a machine gun target is defined as any detectable round within a 3-meter radius from the center of the target.

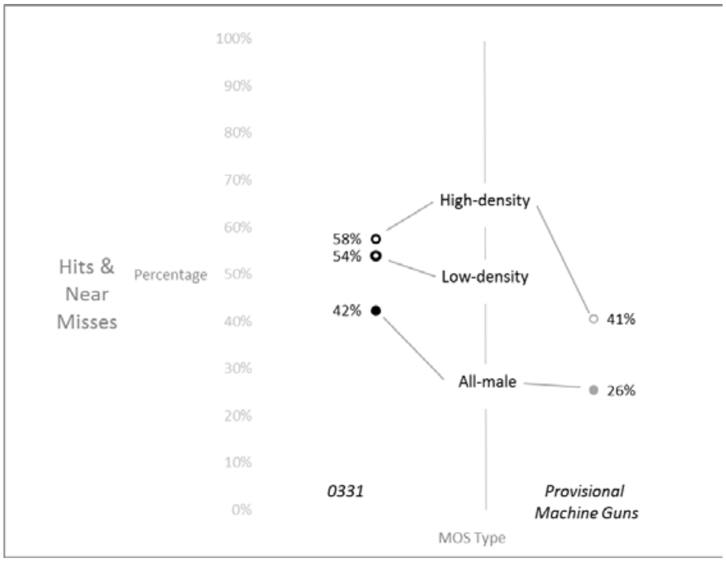


Figure 10. M2 Heavy Machine Gun Hits and Near Misses by MOS and Integration Level

(~~FOUO~~) After an engagement, a machine gun squad must be able to break down their weapon system and move to a position of cover as rapidly as possible. Immediately upon completing target engagement, the 0331 squad took the gun out of action and displaced 100 meters from the firing line to a designated location, moving a heavy load a short distance, manipulating the weapon while fatigued.

(~~FOUO~~) The high-density integrated 0331 squad took significantly longer to displace from the firing line compared to both the all-male and low-density integrated 0331 squads.

9.1.1.2.2.3 Squad Employment of 81mm Mortar

(~~FOUO~~) Indirect fire is used to suppress the enemy from outside the direct-fire weapon range and prepare the battlefield to allow the rifle squad to complete its mission of locate, close-with, and destroy the enemy by fire and maneuver. After completing the approach march, the 81-mm mortar squad conducted an emplacement of the crew-served weapon, engaged two different targets with five 81-mm rounds each, and conducted a displacement of the mortar system. The mortar team engaged the targets using the direct-lay method and visually acquired and adjusted all rounds during each fire mission.

(~~FOUO~~) The focus of the mortar squad is their ability to move an 81-mm mortar system and 10 mortar rounds approximately 100 meters from an operational rally point to a mortar firing position (MFP), get the mortar system fire capable (FIRECAP), and conduct a displacement back to the rally point. Although the mortar squads engaged two targets with five rounds each after the mortar system was FIRECAP, accuracy was not assessed because of the heavy dependence on skill and meteorological conditions.

(FOUO) The integrated 0341 squads had no significant differences compared to the all-male 0341 squads with respect to emplacement and displacement times. What the times did not capture was a masking effect that occurred within the emplacement portion. The initial movement portion of this task (a physical-based capacity) was masked by the time it took to assemble the mortar and sight in on the target (a skill-based capability).

(FOUO) For example, it was observed that when slower members of the squad fell back during the initial movement, their delay was masked (not captured) by the fact that the rest of the team began emplacing the 81-mm mortar system concurrently. By the time the weapon system was FIRECAP, all members had arrived at the MFP. Therefore, a squad that moved quickly versus a squad that got spread out was diffused by the time it took to sight-in on the target because the movement distance was only 100 meters.

9.1.1.2.2.4 Squad Employment of Tube-launched, Optically tracked, Wire-guided (TOW) Missile System

(FOUO) The TOW missile fired out of the Saber system is generally employed from a mounted platform, such as an HMMWV. Members of an anti-armor squad must mount and dismount the Saber system to an HMMWV. During the mounting of the Saber system, the calibration process was not performed due to its technical nature and the time required. The mounting and dismounting tasks were selected due to the physical strength required to lift, curl, and press the Saber components up and lower them down by members of the squad without damaging the system.

(FOUO) The challenging aspect of mounting and dismounting the Saber system is the weight of the components. The integrated 035X squads performed the mount and dismount significantly more slowly than the all-male 035X squads. When mount and dismount were examined separately, the integrated 035X squads were significantly slower in both tasks.

(FOUO) The more challenging method of employing the TOW missile is from a tripod to provide defensive fires. During the assessment, a precalibrated Saber system was mounted on a tripod at a designated firing position. Each anti-armor squad moved four TOW missiles to the firing line and engaged two designated targets, firing each missile in sequence. One missile reload per team was conducted between the first and second shots.

(FOUO) This task was chosen due to the strength required to move each missile and the ability to employ the TOW while fatigued. The integrated 035X squads took significantly longer to engage targets when compared to the all-male 035X squads, but the hit percentage showed no significant differences. The combination of increased engagement times, but similar hits on target makes the integrated squads more vulnerable than their all-male counterparts.

9.1.1.2.3 Infantry Day Two Fatigue and Workload

9.1.1.2.3.1 Comparing Male Responses to Female Responses

(~~FOUO~~) Two self-report surveys (fatigue and workload) were administered to all volunteers at the end of trial activities on day two. The surveys were intended to provide insights into Marines' perceptions of their own level of fatigue and workload level during the course of defensive trials.

(~~FOUO~~) Overall, males tended to report lower levels of fatigue and workload when compared with females, with the exception being 0331 and 0341 squads. In the 0331 squads, males and females reported similar fatigue levels. In the 0341 squads, males and females reported similar fatigue and workload results.

9.1.1.2.3.2 Comparing Male Responses in Integrated Squads to All-male Squads

(~~FOUO~~) In addition to looking at how fatigue and workload self-reports compared between males and females, we also examined whether males responses remained consistent at different integration levels.

(~~FOUO~~) Overall, the reported fatigue and workload levels reported by males tended to decrease when in integrated squads. Initially, this may appear to be a counterintuitive finding, until the results are considered in the context of the task performance. On day two, the common task for all 03XX squads is the 7-km approach march. When males were in integrated squads, they tended to move at a slower pace when compared to all-male squads. The slower pace for males in integrated squads was likely a factor for reduced fatigue and workload.

9.1.1.3 Mountain Warfare (Infantry)

(~~FOUO~~) The mountaineering assessment replicated a logistical resupply of a forward-staged squad while moving in a mountainous environment. Each 12-Marine squad departed an assembly area located at a lower base camp where they hiked 4.6 km with a 75-lb pack and personal weapon (M4) to an objective rally point. The route for this movement was on an unimproved, hilly surface; the terrain was hard and rocky, and uphill, with an approximate gain in elevation of 175 meters, as shown in Figure 11.

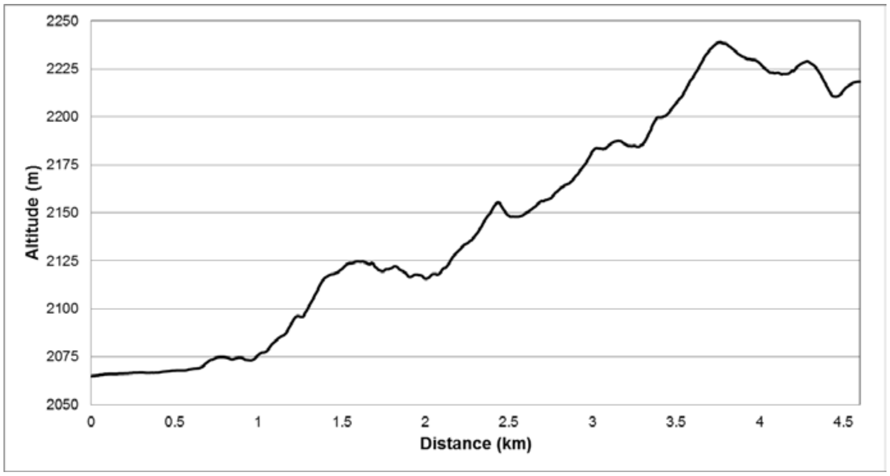


Figure 11. Elevation Change in Meters over 4.6-km Hike Distance

(~~FOUO~~) From the objective rally point, the squad crossed a 200-ft gorge via two single-rope bridges. Once across, the squad pulled their packs across and made a short movement to a location requiring a cliff ascent. Each Marine executed a 40-ft cliff ascent using two climbing lanes. The final action was to conduct a return foot-march to lower base camp. The return march was 5 km while wearing the 75-lb pack and personal weapon across terrain similar to the ascent, but downhill. The entire resupply mission took approximately 5 hours to complete.

(~~FOUO~~) The mountaineering tasks were non-MOS-specific tasks. Due to the non-MOS-specific nature of the tasks, all Marines from 0311, 0331, 0341, 0351, and 0352 MOSs were combined for the purposes of forming 12-Marine squads. Two types of squads were formed each day, all-male and high-density integration squads[8].

(~~FOUO~~) Every Marine received 1 day of recovery after each execution day over the course of 15 days. At the end of a day and during off-days, the Marines lived in squad bays at lower base camp and operated under Company leadership, performing tasks that required minimal physical demands.

(~~FOUO~~) The all-male squads were significantly faster than integrated squads on hikes, gorge crossings, and cliff ascents.

(~~FOUO~~) Fatigue and workload self-report surveys were also administered at the end of mountain warfare activities, in particular after the ascent hike and upon return to lower base camp. When comparing males to females, males reported lower fatigue and workload. Males also reported lower fatigue and workload when in integrated squads compared with when they were in all-male squads. This finding is similar to the infantry day two survey results, where movement under load dominated the task activities.

[8] High-density integrated squads had six females per squad.

9.1.1.4 Infantry Summary

Figure 12 presents a summary of the differences observed when comparing all-male squads to integrated squads. The horizontal bars depict the percentage change observed when comparing the average all-male squad result to an average integrated squad result. The numerical percentage presented adjacent to the bar is indicative of statistical significance. Statistical significance is related to the size of the difference, variation, and the number of trials for that task. This explains why some larger bars show no significance.

When a bar shifts to the right of the centerline the all-male group's average is better than the integrated group's. When a bar shifts to the left the integrated group's average is better than the all-male's. Most of the horizontal bars are shifted to the right for each MOS indicating that all-male groups typically outperform integrated groups.

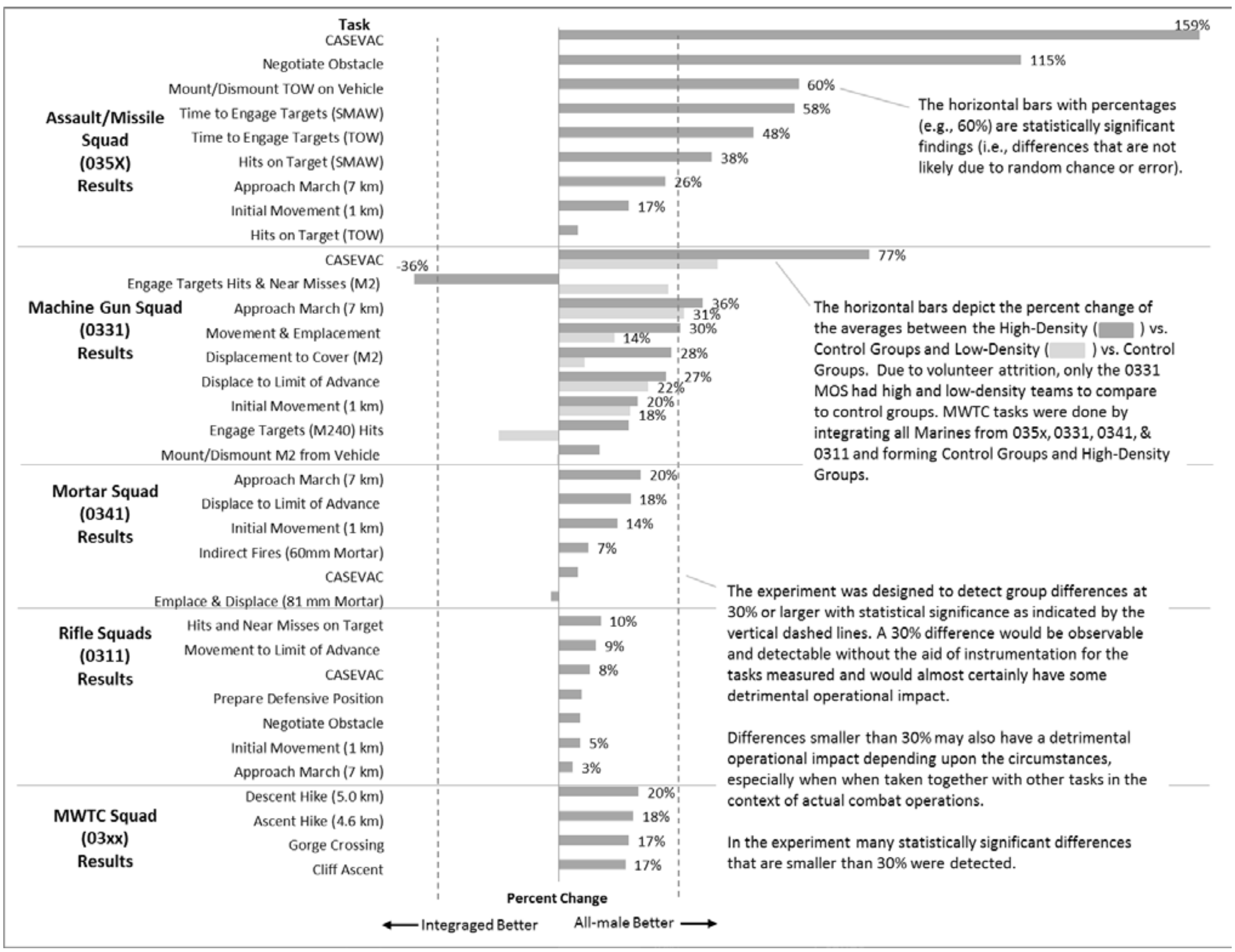


Figure 12. Infantry Summary Comparison of Percentage Change of All Tasks

9.1.2 Artillery

(FOUO) The Artillery assessment was a 1-day scheme of maneuver intended to replicate fire support to a maneuver element.

9.1.2.1 Scheme of Maneuver

(FOUO) The daily scheme of maneuver was designed around the mission of Marine Artillery and its function in a tactical environment. The mission of Marine Artillery is to furnish close and continuous fire support by neutralizing, destroying, or suppressing targets that threaten the success of the supported unit. The design was to mirror an artillery section, providing timely, accurate, and continuous-fire support.

(FOUO) The artillery Marines operated on a 6-days-on/1-day-off schedule. During the 6 days they were executing the daily scheme of maneuver, the Marines bivouacked in the training area, staying in 2-man tents. All Marines, regardless of task, wore the fighting load (35 lb). On the 1 day off, the Battery redeployed to base camp for 24 hours. All normal battery operations, such as logistics resupply and dunnage removal, were completed by the cannoneers.

9.1.2.2 Howitzer Emplacement

(FOUO) Emplacement is the method by which a howitzer is changed from its towed configuration while attached to a prime mover truck, to a detached position where the howitzer is seated on the deck and capable of firing. Emplacement involves dismounting the prime mover, placing the trident bar (58 lb) on the deck, manually manipulating the trail arms (204 lb), entrenching the spades, and setting up the aiming stakes 100 meters away. The Marines performed two emplacements per day during the scheme of maneuver.

(FOUO) The data show that there is no significant difference on the first or second emplacement between the all-male gun crews and the integrated gun crews.

9.1.2.3 Ammunition Preparation

(FOUO) Prior to commencing fire missions, the Marines must offload ammunition from either the prime mover truck or the ammunition truck, and move it to the ammunition pit. One such offload is for emergency fire missions that require an immediate response after a hasty emplacement. In this offload, two cannoneers offloaded three M795 High Explosive Rounds (105 lb per round) from the Loose Projectile Restraint System (LPRS) in the bed of the prime mover. The #3 cannoneer was located in the back of the prime mover and moved the rounds from the LPRS to the #4 cannoneer on the ground. The #4 cannoneer then moved the round from the back of the prime mover to the ammunition pit 10 yards away.

(FOUO) The data show that there is no significant difference between the all-male gun crews and the integrated gun crews when this task was analyzed by integration level, and no significant difference when it was analyzed by the gender of the Marines performing the critical billets (Cannoneer #3 and #4).

(FOUO) The second type of ammunition offload is more deliberate, entails more rounds, and is sourced from the ammunition truck vice the prime mover truck. The ammunition truck offload consisted of four cannoneers offloading 20 M795 High-Explosive rounds and 6 M825 Smoke Rounds from the bed of the ammunition truck to an ammunition pit on the ground 10 yards away. The M795 and M825 are ballistically and dimensionally similar, and offer the same experience in moving the projectile manually.

(FOUO) The data show that there is no significant difference between the all-male gun crews and the integrated gun crews when deliberately offloading ammunition.

9.1.2.4 Fire Missions

(FOUO) Five fire missions were executed by a howitzer section during each 1-day cycle. The missions were: a 3-round low-angle High-Explosive (HE), 3-round high-angle white-phosphorus (WP) smoke, 9-round low-angle HE, 3-round low-angle smoke out of traverse limits, and finally another 3-round low-angle HE from the supplemental position. There is no ostensible difference at the gun-section level between a HE and WP mission.

(FOUO) Each fire mission began with the Fire Direction Center (FDC) starting the event. Procedures for each fire mission were largely the same, regardless of shell and fuze combination or angle of fire. While the gun was receiving the firing data from the FDC, the crews were simultaneously aiming the howitzer by traversing left or right and adjusting the quadrant elevation up and down. The loader, or #4 cannoneer, picked up a projectile from the ammunition pit, held it to allow the section chief to inspect the round, and then placed the round on to the loading tray. Cannoneer #4 then joined the driver and rammed the round into the breech. The #2 cannoneer then loaded the powder increments into the breech powder chamber and closed the breech. Finally, the howitzer was fired by the #1 cannoneer on the section chief’s command.

(FOUO) In the case of a high-angle mission, the loading of the projectile occurred prior to fully elevating the howitzer, a minor departure in procedure from the low-angle missions. The tube would be depressed between shots to facilitate loading.

(FOUO) For all five fire missions, the data show that the all-male gun crews were faster and the difference when compared to the higher-integration gun crews (4 male Marines and 2 female Marines) was significant, but not when compared to low-integration gun crews.

9.1.2.5 Ammunition Resupply

(~~FOUO~~) A significant portion of a cannoneer's time in the field is spent moving ammunition. This task is physically demanding, as it requires the Marines to pick up a 155-mm round, weighing 105 lb, from the ground to their shoulder and maintain positive control while transporting the round. In a field environment, the ammunition resupply truck is often centrally located within a section of howitzers. The typical distance from the ammunition truck to an individual howitzer is approximately 100 meters. In the ammunition resupply task, all volunteers were required to carry two individual projectiles for 100 meters.

(~~FOUO~~) The data show the all-male gun crews were faster and the difference when compared to the higher-integration gun crews (4 male Marines and 2 female Marines) was significant, but not when compared to low-integration gun crews.

9.1.2.6 Speed Shifts

(~~FOUO~~) Deployed howitzer sections often support multiple units in the battlespace. Once the howitzer is emplaced, the crew's ability to respond to a fire mission is restricted by the left and right traverse limit forming a 45-degree cone of fire. To engage targets outside this cone, the crew performs a speed shift by pumping up the suspension and pivoting the howitzer on its wheels by pushing the tube left or right as a team. Once aligned in the direction of the target, the crew emplaces the howitzer by digging in the trail arms and preparing for the fire mission.

(~~FOUO~~) This task occurred twice during a cycle. Furthermore, while all six cannoneers performed this task, the task was analyzed both by integration level and by critical billet. The preponderance of the work during this task falls to the #1 and #2 cannoneers, who are responsible for manually pumping up the howitzer's suspension. While there was no significance between the all-male gun crews and the integrated gun crews, the analysis by critical billet revealed that the differences became significant when the #1 and #2 cannoneer were assigned as a Male-Female and Female-Female combination. This proved to be the same for both speed shifts.

9.1.2.7 Howitzer Displacement

(~~FOUO~~) Quick displacement enables a battery to deliver fire support to maneuver units and move to the next firing point before the enemy has the opportunity to conduct counterbattery fire. During this task, which occurred twice per day, volunteers dismantled the ammunition pit, attached the trident bar, pumped the howitzer to ride height, and dislodged the howitzer's dug-in trail arms. The howitzer was transitioned from a firing configuration to a towed configuration.

~~(FOUO)~~ The data show the all-male gun crews were faster, and that that the difference when compared with the higher-integration gun crews (4 male Marines and 2 female Marines) was significant.

(~~FOUO~~) Also required was an ammunition upload. Ammunition upload requires Marines to pick up a 155-mm round, weighing 105 lb, maintain positive control, and push the ammunition up (62 inches) onto the bed of the prime mover truck to stow it for transport. Unlike the offload task, this task works against gravity.

(~~FOUO~~) The uploading of ammunition took significantly longer with both low- and high-integrated crews when compared with all-male crews.

9.1.2.8 Artillery Fatigue and Workload

(~~FOUO~~) Two self-report surveys (fatigue and workload) were administered to all volunteers at the end of the day's trial activities. The surveys were intended to provide insights into Marines' perceptions of their own level of fatigue and workload during the course of offensive trials.

(~~FOUO~~) Overall, males tended to report lower levels of fatigue and workload when compared to females. However, the reported fatigue and workload levels of males do not appear to change depending on integration of the gun crew. This means that the presence of females on the gun crews did not affect males' perceptions of workload and fatigue levels. This is likely a consequence of the individualized duties of each cannoneer and the inability to spread load tasks between individuals on a crew.

9.1.2.9 Artillery Summary

Figure 13 presents a summary of the differences observed when comparing all-male gun crews to integrated gun crews. The figure *does not* illustrate the differences observed when comparing the males to females in critical billets. The horizontal bars depict the percentage change observed when comparing the average all-male crews result to an average integrated crew result. The numerical percentage presented adjacent to the bar is indicative of statistical significance. Statistical significance is related to the size of the difference, variation, and the number of trials for that task. This explains why some larger bars show no significance.

When a bar shifts to the right of the centerline the all-male group's average is better than the integrated group's. When a bar shifts to the left the integrated group's average is better than the all-male's. Most of the horizontal bars are shifted to the right for each MOS indicating that all-male groups typically outperform integrated groups.

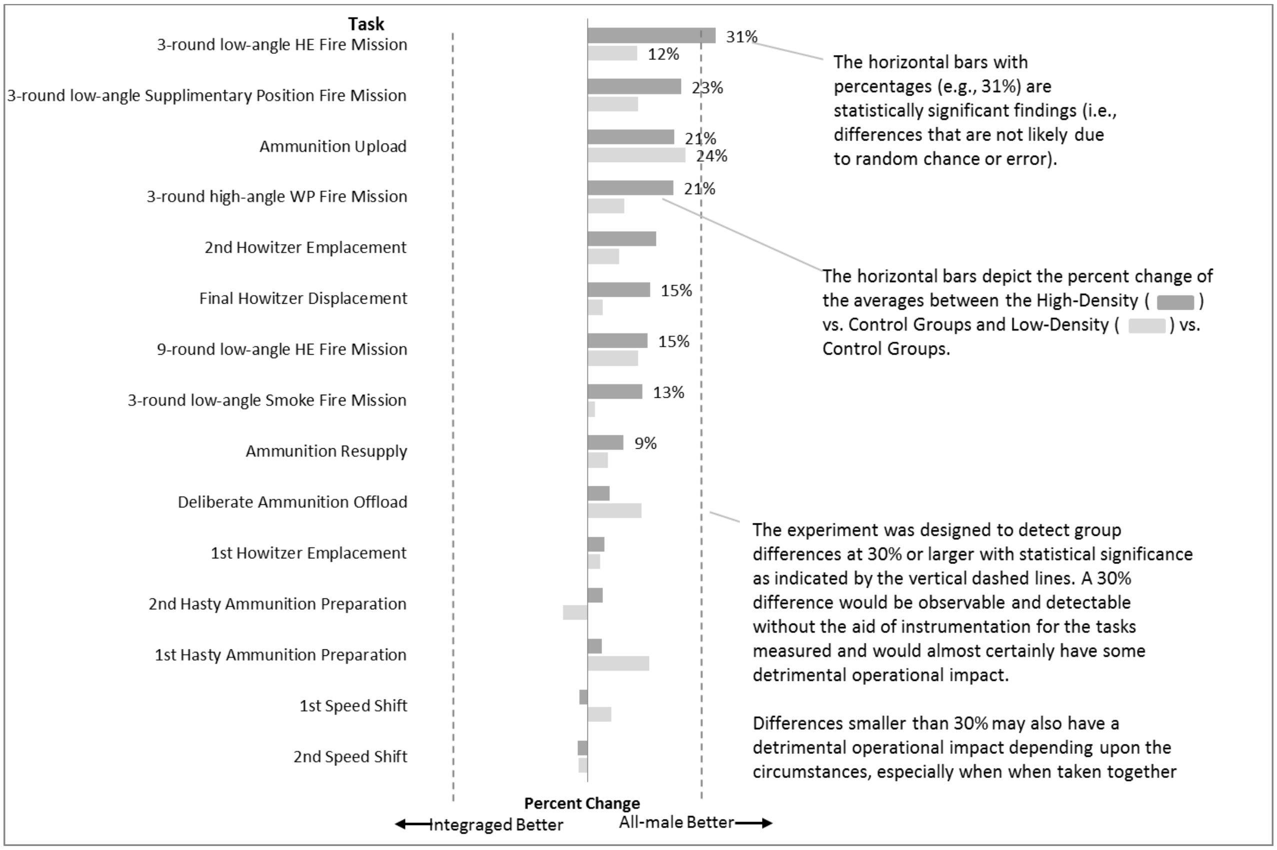


Figure 13. Artillery Summary Comparison of Percentage Change of All Tasks

9.1.3 Combat Vehicles

(~~FOUO~~) The combat vehicle assessment includes M1A1 Tank Crewmen (MOS 1812), Assault Amphibious Crewmen (MOS 1833), and Light-Armored Vehicle Crewmen (MOS 0313). The combat vehicle assessment was a 3-day scheme of maneuver. Two of the three days included live-fire tasks and non-live-fire tasks where crew performance was measured. The remaining day included no measured tasks, but was required for routine preventive maintenance and repairs to ensure that the unit's vehicles and equipment remained fully operational and to improve the operational realism of the experiment.

(~~FOUO~~) The measured tasks for combat vehicles were intended to replicate the tasks required to prepare for combat, engage the enemy, conduct casualty evacuations, perform field maintenance, and, ultimately, perform follow-on maintenance actions. The combat vehicle units participating in the assessment received 1 rest day after completing 4 cycles, resulting in an operational tempo of 12 days on and 1 day off. The majority of the tasks took place aboard the MCAGCC, Twentynine Palms, CA with follow-on tasks for assault amphibious crewmen conducted at Camp Pendleton, CA.

(FOUO) The combat vehicle crews all wore a standard vehicle fighting load while conducting all tasks. This included a combat vehicle crewman (CVC) protective suit, an improved plate carrier (IPC) with small-arms protective insert (SAPI) plates, a Kevlar or CVC helmet, and an M4 Carbine. This gear weighed approximately 35 lb. Prior to any trial day with observed tasks, the Marines bivouacked at a platoon position on their vehicles.

9.1.3.1 Prepare for Combat

(FOUO) The capability of combat vehicles to respond to a mission depends on the ability of the crews to rapidly prepare vehicles for combat operations. Physically demanding aspects of preparing vehicles for combat include loading and stowing ammunition, readying weapons stations for operation, and securing/unsecuring vehicles. Readying a weapon station for operation, although discussed as a preparatory task, is also important because it is similar to what would be required to troubleshoot and correct a malfunction that occurs while employing the weapon in a combat situation.

(FOUO) To prepare vehicles for combat crews must:

- (FOUO) Tank - Crews must prepare the commander's M48 .50-caliber machine gun (84 lb) and upload a full complement of ammunition onto the vehicle from a field ammunition supply point.[9]
- (FOUO) LAV - Crews must prepare the M240 coaxially mounted machine gun (26 lb), the M242 Bushmaster Cannon (263 lb), and load 630 rounds of 25-mm ammunition (720 lb) into the vehicle with upload to the feed chutes.
- (FOUO) AAV - Crews must load and install the M2 .50-caliber machine gun (84 lb) and the MK-19 40-mm grenade launcher (78 lb) with ammunition. In addition to the weapons, the crews loaded two cans of .50-caliber ammunition (35 lb per can) and three cans of 40-mm grenades (45 lb per can) via the lowered ramp of the AAV. Loading and installation, depending on the situation, can be done externally or internally. When deploying from naval shipping, the crews also secure and unsecure vehicles in preparation for amphibious operations.

(FOUO) Preparing the Tank Commander's M48 weapon and the LAV M240/M242 were tasks performed almost exclusively by individuals vice the entire crew. Due to the individual nature of these tasks, gender of the Marine in the critical billet was used for comparison instead of crew integration level. There was no significant difference between male and female tank crewmembers when preparing the M48, but male LAV

[9] The tank ammunition resupply was not evaluated (see deviations in Annex I) due to substantial participation by non-volunteer Marines.

crewmembers performed M240 and M242 preparatory tasks significantly faster than female crewmembers. Loading and stowing ammunition, including upload into feed chutes, in the LAVs was a crew-level task that showed no differences when comparing integrated crews with all-male crews.

(FOUO) In the AAVs, mounting weapons and loading ammunition was combined into an overarching task performed by the entire crew. In this task, all-male crews were significantly faster at external loading of weapons and ammunition compared with integrated crews. All-male crews were also faster at internal loading of the M2 .50 caliber machine gun when compared to high-density integrated crews, but not when compared to low-density integrated crews.

(FOUO) The AAV crews also had to perform secure/unsecure tasks as part of preparing for combat, as would be required when deploying from naval shipping. In the conduct of this task, the crews had to move eight sets of chains and dogging brackets from its stowage area and secure the AAV to the deck of a naval vessel. Upon arrival at the objective, the crews reversed the process by unsecuring the AAV and returning the chains to their appropriate stowage areas.

(FOUO) There were no significant differences in securing and unsecuring when comparing all-male crews with integrated crews.

9.1.3.2 Engage Enemy Targets

(FOUO) Combat vehicle crews must be capable of acquiring and engaging targets in both offensive and defensive engagements under operational conditions and environments. To assess crews' ability to engage enemy targets, gunnery qualification tests were adapted for the assessment and included offensive and defensive engagements. The engagements included reloading, manual manipulation in the event of malfunction, and remedial actions necessary to keep the weapon stations in operations.

9.1.3.2.1 Tank Engagements

9.1.3.2.1.1 Load and Arm Main Gun

(FOUO) Tank crews engaged a series of main gun targets and machine gun targets from a defensive fighting position. Crews also engaged gun targets while in the offense, similar to the defense, except that the tank was maneuvering downrange while engaging the targets rather than engaging targets from a stationary defensive fighting position.

(FOUO) With the exception of the loader's M240, accuracy of engagements was not the focus. The focus for the assessment was on driver and loader, the more physically demanding positions.

(FOUO) The loader has a vital role in the lethality and survivability of the tank. During main gun engagements, the loader must remove a round weighing 52.8 lb from a stowage compartment approximately 3 to 5 feet off the ground. Then the loader must flip over the round and load it into the breach while balancing inside a stationary or moving vehicle. Because the tank is manually reloaded, the rate of fire is limited by the speed at which the loader can manually reload the main gun. In a static environment, prior to entering the live-fire range, male and female loaders performed the task of loading and arming the main gun. This was done as an individual effort for comparison because no assistance from other crewmembers is possible inside the tank.

(FOUO) In the static environment, female loaders were significantly slower than their male counterparts. The average time for males (6.45 seconds) was within the prescribed standard of 7 seconds, but the average time for females (7.45 seconds) was not, by a fraction of a second.

(FOUO) To take a more operational view of loading and arming the main gun, this reloading task was measured on the live-fire range in both offensive and defensive engagements. To minimize the rest time for the loader between reloads under strenuous operational conditions, targets were chosen at close ranges and azimuths that facilitated quick acquisition and engagement of subsequent targets. Because vehicle movements play a role in engagements, the task was examined by crew integration level and the gender of the loader.

(FOUO) Regardless of offensive or defensive engagements, all-male crews and male loaders were significantly faster than their integrated crews, and female loaders performed similarly to what was observed in the static environment.

9.1.3.2.1.2 Manual Turret Engagement

(FOUO) The turret and main gun of the M1A1 are primarily operated by electric and hydraulic systems. As a backup, they can be manipulated manually using crank handles located in the gunner’s station. These backups provide a way for the crew to continue the fight in the event of a hydraulic or electrical malfunction. This task required a single crewmember to manually traverse and elevate the turret.

(FOUO) Male crewmembers performed this task significantly faster than female crewmembers.

9.1.3.2.1.3 Ammunition Transfer

(FOUO) When the ready ammunition stowage compartment is low on ammunition, the loader transfers ammunition into it from the semi-ready ammunition stowage compartment. The loader is required to transfer 16 rounds, each weighing 52.8 lb from the semi-ready ammunition stowage compartment to the ready ammunition stowage compartment. Transfer, like reloading, is an individual task.

(~~FOUO~~) In this task, male loaders performed the task significantly faster than female loaders.

9.1.3.2.1.4 Employing Loader's M240

(~~FOUO~~) The loader's weapons system was an M240 medium machine gun pintle-mounted adjacent to the loader's hatch.

(~~FOUO~~) The loader's M240 is primarily used for close defense of the tank. This task required the loader to engage a group of four man-sized targets at a range of 250 meters, pausing to reload the weapon during the engagement. The targets were set to continuously reappear so that the loader would run out of ammunition and be forced to conduct the reloading portion of the task.

(~~FOUO~~) Employing secondary weapons and reloading were individual tasks where male crewmembers were compared to female crewmembers. Male crewmembers took significantly less time to get first hits on target, and got significantly more total hits on target; however, female crewmembers were able to reload the M240 significantly faster than male crewmembers. The loader's M240 on a M1A1 takes a great deal of physical strength to employ accurately. Unlike the common M240B variant, the loader's M240 has dual-spade grips and a butterfly trigger instead of a buttstock and pistol grip. It is also pintle-mounted on a skate ring and has no bipods for support. While these modifications greatly increase the quick maneuverability of the weapon and enable the loader to cover a 180-degree sector of fire, they also necessitate the application of a significant amount of force to keep the weapon on target while firing.

9.1.3.2.1.5 Reloading Commander's Weapon Station

(~~FOUO)~~ The Commander's weapon system is the M48 heavy machine gun that can be operated inside the tank, but must be reloaded from an exposed position outside the tank armor. Stationary tank engagements with the heavy machine gun forced a reload of the Commander's M48 heavy machine gun during live fire.

(~~FOUO~~) Male crewmembers reloading the commander's M48 were not significantly faster than their female crewmembers.

9.1.3.2.2 AAVs

9.1.3.2.2.1 Offensive and Defensive Engagements

(~~FOUO~~) The AAV crews engaged multiple targets ranging from 400 to 1500 meters from multiple defensive and offensive positions. Employing both weapons systems in an AAV turret is a physically strenuous task for the Turret Gunner, especially in the case when immediate or remedial action is required. Although the other crew members are not directly involved in the employment of the Up-Gunned Weapons Station (UGWS), the driver's performance spotting targets and maneuvering the vehicle to allow the turret gunner to safely engage those targets can have a significant effect on the Turret Gunner's ability to identify and engage targets.

(~~FOUO~~) The potential interdependency of driver and turret gunner necessitates an examination by integration level, first by comparing integrated crews with all-male crews. Integrated crews showed no significant difference in offensive or defensive effects when compared to all-male crews. A second examination was done using the gender of the turret gunner as the comparative factor. In both offensive and defensive engagements, male turret gunners performed significantly better than female turret gunners. Experience with the weapon station may have played a causal role in the disparity, as could the interdependency between the driver and turret gunner.

9.1.3.2.2.2 Interior and Exterior Reloads

(~~FOUO~~) In combat operations, faster reload times (even measured in fractions of a second) are highly desirable. Faster reloading increases the combat effectiveness and survivability of an AAV crew. Interior and exterior reloads require the AAV Turret Gunner and Rear Crewman to conduct a reload of 200 rounds of .50-caliber ammunition and 64 to 96 rounds of 40-mm grenade rounds. A loaded can of .50-caliber ammunition weighs approximately 35 lb and a loaded can of 40-mm grenade rounds weighs approximately 45 lb.

(~~FOUO~~) For interior reloads, the Rear Crewman assisted the Turret Gunner by moving boxes of ammunition from the appropriate troop-compartment stowage spaces and handing them to the Turret Gunner, who sat in the vehicle turret. Interior reloads were performed twice during an engagement.

(~~FOUO~~) Interior reloads showed no significant differences when comparing all-male to integrated groups, with the exception of all-males, who were faster than high-density integrated crews on the second interior reload. A gender comparison was done based on the gender of the crewmembers occupying the turret gunner and rear crewman positions leading based on the combinations in Table 2.

Table 2. Gender Combinations by Billet for Analysis

Critical Billet Combination	Rear Crewman Gender	Turret Gunner Gender
MM	Male	Male
MF	Male	Female
FM	Female	Male
FF	Female	Female

(FOUO) When examining the critical billet holders for interior reloads, there was no significant difference in reload times, even when comparing the all-male combination to the all-female combination of critical billet holders.

(FOUO) Exterior reloads showed a much different result than interior reloads. Exterior reloads were performed once during an engagement. The Rear Crewman assisted the Turret Gunner by moving boxes of ammunition from the appropriate troop-compartment stowage spaces to the top of the vehicle and into the turret from the exterior. All-male AAV crews were significantly faster at exterior reloads than high-density crews, but not low-density crews. Similar to interior reloads, analysis was performed by gender of the critical billet holders in the task using the combinations in Table 2. The all-female turret gunner/rear crewman combination is significantly slower than the all-male combination. In addition, any time a female served as a rear crewman lifting ammunition cans up to the turret gunner, there appeared to be a corresponding increase in time when compared to other combinations.

9.1.3.2.2.3 Manual Turret Manipulation

(FOUO) Similar to the tank crewmen, the AAV Turret Gunner must be able to manually traverse, elevate, and depress the vehicle turret UGWS, using the manual traverse and elevation controls within the turret in the event of a loss of electrical power supply. Manual manipulation of the turret allows the turret gunner to continue to scan and engage targets in the event of degraded system operation.

(FOUO) At the crew level, this task only appears to show significance between all-male and high-density integrated crews. This task, however, is an individual task. Examining the task by gender reveals that male crewmembers are significantly faster than female crewmembers at manual manipulation of the turret.

9.1.3.2.3 LAVs

9.1.3.2.3.1 Offensive and Defensive Engagements

(FOUO) The LAV crews engaged multiple vehicle and personnel targets using the main gun and the coaxially mounted machine gun from both offensive and defensive positions. The final engagement was a manual engagement that required the gunner to use the manual traverse and elevation controls within the turret to acquire and destroy a vehicle target.

(FOUO) All offensive and defensive engagements, with the exception of the manual engagement, were evaluated at the crew level. The comparisons were made based on their time to engage and kill a target. In both offensive and defensive engagements, integrated crews performed similarly to all-male crews with respect to effects on target.

(FOUO) The most physically challenging of the engagements was manual engagement. Manual engagements were evaluated by time and percent kills. In manual engagements, there was no significant difference between male and female gunners on either the time, or the percent kills.

9.1.3.2.3.2 Main Gun Remediation

(FOUO) In the event of a misfire on the main gun, the gunner, with assistance from the vehicle commander and driver, must perform remedial actions to get the vehicle back in the fight. Remedial action of the main gun requires the crew to download the ammunition from the main gun, remove the feeder assembly, ensure that it was correctly timed and reinstalled, and upload ammunition.

(FOUO) The Vehicle Commander and Driver were able to assist in main gun remediation, but due to the nature of crew responsibilities and the position of the gun in relation to the crewmembers, the Gunner bore most of the physical workload. Comparisons were made based on gender of the gunner vice crew integration level. Male gunners were significantly faster than female gunners at both disassembly and reassembly with the larger difference on reassembly. Reassembly tends to take longer than disassembly due to the fact that the gunner must work against gravity to lift in place components of the weapon without damage.

9.1.3.2.3.3 Manual Turret Manipulation

(FOUO) Similar to the tank and AAV crewmen, the LAV Turret Gunner must be able to manually traverse, elevate, and depress the vehicle turret, using the manual traverse and elevation controls within the turret in the event of a loss of electrical power supply. Manual manipulation of the turret allows the turret gunner to continue to scan and engage targets in the event of degraded system operation.

(FOUO) Manually traversing the turret is an individual task. Comparisons were made by integration level of the crew, which showed all-male groups were faster than integrated groups. Given the individual nature of the task, however, the most appropriate comparison was by gender of the gunner. When comparing male gunners to female gunners, the male gunners were significantly faster at manually traversing and elevating/depressing the turret.

9.1.3.3 Crew and Casualty Evacuations

(FOUO) Personnel casualties should be an expected result of enemy engagements. All vehicle crews conducted casualty evacuations. Tanks and LAVs focused on a single incapacitated crewmember evacuation from the most difficult position and also exercised crew evacuations. AAV crews conducted three different types of incapacitated crewmember evacuations, one of them while at sea.

(FOUO) Each vehicle platform presents its own set of unique challenges and specific operational context, but the one commonality among the vehicles is the casualty dummy. Similar to the infantry casualty, the dummy weighed 174 lb, the average weight of a Marine. The casualty dummy was anatomically correct with natural joint articulation, and was dressed in combat vehicle clothing and equipment similar to the Marines bringing the total weight up to 205 lb.

(FOUO) In each case, at least one or more vehicle crewmembers were responsible for removing the casualty from a vehicle station (e.g., gunner's station). For Tanks and LAV crews, an additional step was required to move the casualty to a rally point that was between 25 and 50 meters from the vehicle. Performing this task quickly enables the casualty to receive medical attention as soon as possible.

9.1.3.3.1 Tanks

9.1.3.3.1.1 Crew Evacuation

(FOUO) A combat-loaded M1A1 contains several hundred gallons of flammable fuel and hydraulic fluid, and several hundred pounds of HE ordinance. In the event of a fire or other emergency, the survivability of the crew depends on a quick evacuation of the tank.

(FOUO) The crew evacuation task was performed with all crewmembers starting at crew stations with safety guards installed and hatches closed and locked. Upon receiving the command to evacuate, the crew exited their stations and moved to a rally point behind the vehicle. Opening the crew hatches required lifting a 40- to 65-lb hinged door in an upward motion, away from the crewman. All crewmen exited the vehicle safely and moved as a crew to the rally point 50 meters away. Marines wore fighting loads and carried individual weapons, as would be expected when evacuating a vehicle in a tactical environment. No significant difference in times was shown between integrated crews compared to all-male crews.

9.1.3.3.1.2 Casualty Evacuation

(FOUO) Onboard the tank, the casualty was staged in the gunner's station leaving only three crewmembers to participate in the evacuation. The gunner's station within a tank is the most difficult location from which to evacuate a Marine because of its inaccessibility and limited workspace. To perform the casualty evacuation, three crewmembers lifted the dummy out the loader's hatch and onto the roof of the turret, and carried the casualty to a rally point 50 meters behind the vehicle. No significant difference in times was shown between integrated crews compared to all-male crews.

9.1.3.3.2 AAVs

9.1.3.3.2.1 Evacuation on Land

(FOUO) The casualty onboard the AAV was staged in the turret for land evacuation leaving only two crewmembers to participate in the evacuation. The evacuations on land were conducted in two different ways to replicate different tactical scenarios. One method was the exterior evacuation and the other was the interior. For the exterior evacuation, the casualty was placed in the turret and had to be pulled out and placed on top of the vehicle. The interior evacuation also started with the casualty in the turret, with subsequent movement to the floor of the troop compartment while still remaining inside the vehicle.

(FOUO) The rear crewman began the turret evacuation with follow-on assistance provided by the driver after a 60-second delay that allowed the driver to radio a casualty evacuation report. If the crew had not evacuated the casualty after 6 minutes had passed, then a third crewman from a separate vehicle provided assistance.

(FOUO) Exterior casualty evacuations from the turret showed significant differences between all-male crews and high-density integrated crews with all-male crews performing the task faster than high-density crews, but not low-density crews.

(FOUO) To investigate further, the gender of the Marines occupying rear crewman and driver positions were analyzed. When a female occupied the rear crewman and driver positions simultaneously, the performance was significantly slower than when either one or both of those positions was occupied by a male. The position of a female in the two-person combination also explains why at the crew level, low-density integrated crews were not significantly different from all-male crews.

(FOUO) When a male occupied the rear crewman position, the evacuation times were not significantly faster than all-male crews. However, when a female occupied the rear crewman position, the difference was significant. The reason is explained by the fact that the rear crewman begins the evacuation without any assistance until the driver completes the casualty report. More progress is made in the evacuation when the male began the effort than when the female did.

(FOUO) The results for the interior evacuation are not as telling as the exterior evacuation. Interior evacuations showed no significant differences when comparing all-male to integrated crews. When examined by gender of the rear-crewman and driver, having a female begin the evacuation again played a role in longer evacuation times, but the result is less definitive. One reason for the differing results between exterior and interior evacuations relates to with the direction of the evacuation. For exterior evacuations, the casualty must be lifted out of the turret, whereas interior evacuations require the casualty to be pulled down into the troop compartment.

9.1.3.3.2.2 Evacuations on Water

(FOUO) Crew evacuations on water have an added emphasis and complexity. Operating on a platform that pitches and rolls with the waves makes the task more difficult. If the cause for the evacuation is a sinking vehicle, then time is of critical importance to the casualty's survival chances. For casualty evacuations on water, the turret gunner and driver were required to lift the casualty from the floor of the troop compartment to the top of the vehicle through the starboard, right cargo hatch.

(FOUO) Water evacuation results were similar to exterior evacuation results. In water evacuations, all-male crews were significantly faster than high-density integrated crews, but not significantly faster than low-density crews.

9.1.3.3.3 LAVs

9.1.3.3.3.1 Crew Evacuation

(FOUO) All crewmen were required to exit the vehicle safely and move as a crew to the rally point 25 meters away. Marines wore fighting loads and carried individual weapons, as would be expected when evacuating a vehicle in a tactical environment.

(FOUO) The all-male LAV crews evacuated the vehicle significantly faster than the high-density integrated crews, but not significantly faster than the low-density integrated crews. Operationally, the time difference for this task, taken in isolation, is not meaningful. This difference only becomes meaningful when combined with other tasks, such as follow-on casualty evacuations.

9.1.3.3.3.2 Casualty Evacuation

(FOUO) The LAV casualty was placed in the vehicle commander's station, the most difficult position to evacuate, leaving the two remaining crewmen to conduct the evacuation. The crewmen evacuated the casualty from the vehicle by extracting it from the turret and placing it on the front of the vehicle, then moving the casualty to the rally point 25 meters away. If removing the casualty took too long, other vehicle crewman from a different vehicle, or vehicle scouts, were pulled from a supplementary position and were allowed to assist in the evacuation.

(FOUO) LAV casualty evacuations from the turret showed significant differences between all-male crews and high-density integrated crews, but not low-density crews. High-density integrated crews were significantly slower at both extraction and movement to the rally point when compared with all-male crews.

9.1.3.4 Field Expedient Maintenance

(FOUO) In a combat environment, a quick recovery of a disabled vehicle denies the enemy the opportunity to maneuver or employ fires against a combat vehicle. In the event that a vehicle becomes disabled, crews are equipped and trained to perform actions, repairs, and self-recovery to enable missions to continue.

9.1.3.4.1 Tanks

9.1.3.4.1.1 Vehicle Recovery

(FOUO) In the event that a tank becomes disabled, tank sections are equipped and trained to perform self-recovery. Performing tank recovery involves crewmembers from each tank dismounting from their vehicles, removing the 300-lb tow bars from the stowage position on the operational tank, mounting them on the front of the disabled tank, and then holding the tow bars in place approximately 3 feet off the ground while the operational tank maneuvers into position to connect the tow bar's lunette eye to a pintle hitch.

(FOUO) All-male crews showed no significant difference in recovery times when compared with integrated crews.

9.1.3.4.2 AAVs

9.1.3.4.2.1 Manual Ramp Raise

(FOUO) A malfunction of the AAV ramp can leave a vehicle vulnerable and incapable of operations at sea. To correct malfunctions of an AAV ramp, crews are required to manually raise the ramp with a ramp jack.

(FOUO) The all-male AAV crews raised the vehicle ramp significantly faster than the high-density integrated crews, but not the low-density integrated crews. The difference in time to conduct this task, although statistically significant, is not operationally relevant because the time difference is small.

9.1.3.4.2.2 Vehicle Recovery on Water and Land

(FOUO) AAV crews performed water and land recovery of disabled AAVs with like vehicles. The crews of the recovery and disabled AAVs worked together for each portion of the recoveries. The water recovery required the Turret Gunner and the Rear Crewman from each vehicle to throw water-tow ropes to the recovery vehicle or use a boat hook to secure the thrown ropes. The crews then connected both vehicles stern-to-stern and towed the disabled vehicle in the water.

(FOUO) Once on land, towing required both crews to manipulate the two vehicles and the land tow bar, which weighed approximately 150 lb, and attach it to two points on the disabled vehicle. The operational vehicle then backed up until the tow pintle seated properly. Once the vehicle was attached, both crews loaded into the recovery vehicle and the recovery vehicle began to tow.

(FOUO) Neither water nor land recovery operations showed any significant differences in times when comparing all-male crews to integrated crews.

9.1.3.4.3 LAVs

9.1.3.4.3.1 Rig for Recovery and Tow

(FOUO) LAV crews performed recovery and towing operations with notionally mired and disabled LAVs. Rigging for recovery employed an LAV Logistics (LAV-L) variant connected via winch cable to a mired LAV-25 variant. The crews maneuvered the LAV-L variant 10 meters away, paid out the winch cable, and connected the two vehicles with two 75-lb snatch blocks to create a three-point recovery system.

(FOUO) With the mired vehicle recovered, the crew rigged the disabled vehicle for towing using the 175-lb tow bar located on the side of the LAV-L variant. With the two vehicles still spaced 10 meters apart, the crewmen detached the tow bar from the LAV-L and connected it to the tow points on the LAV-25. The crew ground-guided the Logistics variant and, with one crewman holding the tow bar, attached the eye of the tow bar to the tow pintle of the LAV-L variant.

(FOUO) There were no significant differences in the time to rig for recovery or towing when comparing all-male crews with integrated crews.

9.1.3.4.3.2 Remove and Install Tire

(FOUO) The ability to rapidly change a tire has critical implications to ensuring mission accomplishment. Depending on the terrain, it would not be unusual for a vehicle crew to change multiple tires throughout the course of an operation.

(FOUO) To perform a tire change, the LAV crew removed the spare tire, which weighed 200 lb, from the side of the vehicle. The crew jacked up the vehicle, removed a vehicle tire, mounted the vehicle's spare tire, and tightened the lug nuts. Finally, the crew returned the damaged tire to the side of the vehicle on the spare mounting bracket.

(FOUO) There were no significant differences in the time to change a tire when comparing all-male crews with integrated crews.

9.1.3.5 Follow-on Maintenance

(FOUO) Vehicle crewmen in the operating forces typically spend a significant percentage of available man hours performing routine maintenance. This work is often physically demanding because the parts, tools, and equipment that are organic to a combat vehicle unit are large and heavy. A proficient and effective vehicle crewman is capable of performing multiple hours of preventative maintenance or repairs each day. This is operationally realistic and an indispensable part of every vehicle crewman's job.

9.1.3.5.1 Tanks

(FOUO) The follow-on maintenance task for tank crewmen involved separating the tank track, removing and replacing a track section, and reinstalling the track. Maintenance tasks similar to this are commonly required to repair damage sustained during operations. When a tank sustains damage to its track or suspension during a mission, the crew will perform repairs encompassing some or all of the performance steps required for this task. In a combat environment, it is critical to perform repairs and regain mobility quickly, denying the enemy the opportunity to attack a stationary target.

(FOUO) In this task, all-male crews performed the repairs significantly faster than integrated crews.

9.1.3.5.2 AAVs

9.1.3.5.2.1 Break and Reassemble Track

(FOUO) The follow-on maintenance task for AAV crews involved separating the track, removing and replacing a track section, and reinstalling the track. Maintenance tasks similar to this one are commonly required to repair damage sustained during operations. When an AAV sustains damage to its track or suspension during a mission, the crew will perform repairs that encompass some or all of the performance steps required for this task. In a combat environment, it is critical to perform repairs and regain mobility quickly, denying the enemy the opportunity to attack a stationary vehicle target.

(FOUO) Breaking and reassembling track showed significant differences between all-male crews and high-density integrated crews, but not low-density crews. To investigate further, the gender of the Marines occupying the rear crewman and turret gunner positions was examined. When a female occupied the rear crewman and turret gunner positions, simultaneously, the performance was significantly slower than when both positions were occupied by males.

9.1.3.5.3 LAVs

9.1.3.5.3.1 Remove and Replace Armor Panels

(FOUO) Removing armor panels, while not conducted as regularly as other maintenance actions, is required should the panels be damaged in combat operations, or during wash-downs and cleaning after extended operations. The armor panels are composite-armor, mounted on the back and side of the vehicle. The two side-armor panels weigh 60 lb each and the two rear scout-hatch armor panels weigh 125 lb each.

(FOUO) Removing and replacing the lighter front armor panels showed no significant differences between all-male and integrated groups, but as the panel weight increased, so did the differences. The removal and replacement of the heavier scout-hatch panels took significantly longer with high-density integrated crews when compared with all-male crews, but not when compared with low-density crews.

9.1.3.6 Combat Vehicles Fatigue and Workload

(FOUO) Two self-report surveys (fatigue and workload) were administered twice to all volunteers. The first administration occurred at the conclusion of the tasks on non-live-fire days, which included crew and casualty evacuation, field expedient maintenance, and follow-on maintenance activities. The second administration was after the conclusion of live-fire tasks, which included preparing for combat and engaging enemy targets. The surveys were intended to provide insights into Marines' perceptions of their own level of fatigue and workload level during the course of offensive trials.

9.1.3.6.1.1 Non-Live-Fire Fatigue and Workload

(FOUO) Overall, males and females reported similar fatigue and workload levels across all three combat vehicle platforms with one exception: male tank crewmen. Male tank crewmen reported lower workload results than female tank crewmen.

(FOUO) The reported fatigue and workload levels of males do not appear to change depending on integration of the crew with the exception of male LAV crewmen. Male LAV crewmen report higher fatigue and workload levels when part of integrated crews. Non-live-fire days included maintenance tasks conducted by the entire crew, which allowed for more opportunities for compensation to occur.

9.1.3.6.1.2 Live-Fire Fatigue and Workload

(FOUO) On live-fire days, the fatigue and workload results are mixed. Male LAV crewmen reported lower levels of fatigue than female LAV crewmen, but the same cannot be said for Tank and AAV crewmen; male and female Tank and AAV crewmen reported similar fatigue levels on live-fire days. The reverse is true for workload. On live-fire days, male and female LAV crewmen reported similar workload levels while male Tank and AAV crewmen reported lower workloads than their female counterparts.

(FOUO) The reported fatigue and workload levels of males do not appear to change depending upon integration of the crew, with the exception of male tank crewmen. Male tank crewmen reported higher workloads when in integrated crews compared to when they were in all-male crews. There were a small minority of extreme male survey responses that impact this answer. When adjusting for a few extreme observations, the results change to being similar between all-male and integrated.

9.1.3.7 Combat Vehicles Summary

Figure 14 presents a summary of the differences observed when comparing all-male gun crews to integrated gun crews. The figure does not illustrate the differences observed when comparing the males to females in critical billets. The horizontal bars depict the percentage change observed when comparing the average all-male crews result to an average integrated crew result. The numerical percentage presented adjacent to the bar is indicative of statistical significance. Statistical significance is related to the size of the difference, variation, and the number of trials for that task. This explains why some larger bars show no significance.

When a bar shifts to the right of the centerline the all-male group's average is better than the integrated group's. When a bar shifts to the left the integrated group's average is better than the all-male's. Most of the horizontal bars are shifted to the right for each MOS indicating that all-male groups typically outperform integrated groups.

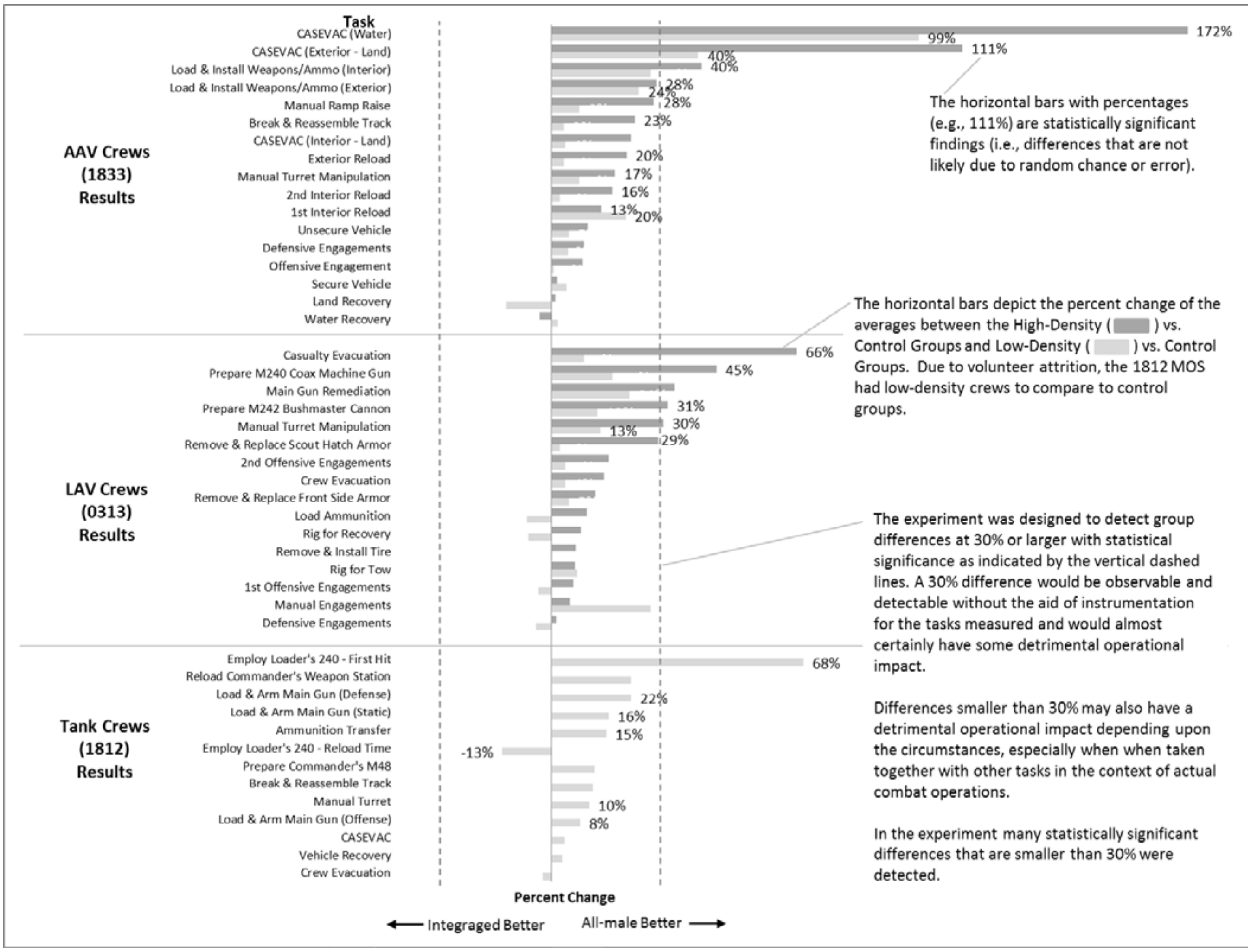


Figure 14. Combat Vehicles Summary Comparison of Percentage Change of All Tasks

9.2 Objective 2

9.2.1 Provisional (Infantry and Machine Guns)

~~(FOUO)~~ The provisional infantry assessment was a 2-day scheme of maneuver intended to replicate offensive and defensive tasks. The provisional infantry assessment included Provisional Infantry (PI) Riflemen and Provisional Machine Gunners (PIMG) who could be sourced from any MOS except 03XX, 08XX, or 18XX family of MOSs. All Marines, regardless of task, wore the fighting load (35 lb) and carried a personal weapon (6.4 lb, 7.9 lb, or 10.5 lb, depending upon weapon type).

9.2.1.1 SOM Overview

9.2.1.2 Trial Cycle Day One

~~(FOUO~~) The first day of the infantry assessment was a coordinated, live-fire squad attack employing the rifle and machine gun squads. The coordinated attack allowed for simultaneous employment of the squads in close proximity on the same range for operational realism, but with offsets in battlespace geometries of fire to allow for independent observation and measurement of mission effects, fatigue, and workload.

~~(FOUO~~) In general terms, the coordinated attack began with an initial movement, followed by an enemy engagement, consolidation on the objective, and, finally, a casualty evacuation.

~~(FOUO)~~ Day one was a fluid scenario that allowed the Marines to flow from beginning to end without interruption or breaks. Figure 15 illustrates infantry live-fire attack. At the conclusion of the casualty evacuation, the Marines operated under the guidance of their Company leadership, performing tasks that required minimal physical demands, and then bivouacking in the field, sleeping in 2-man tents.

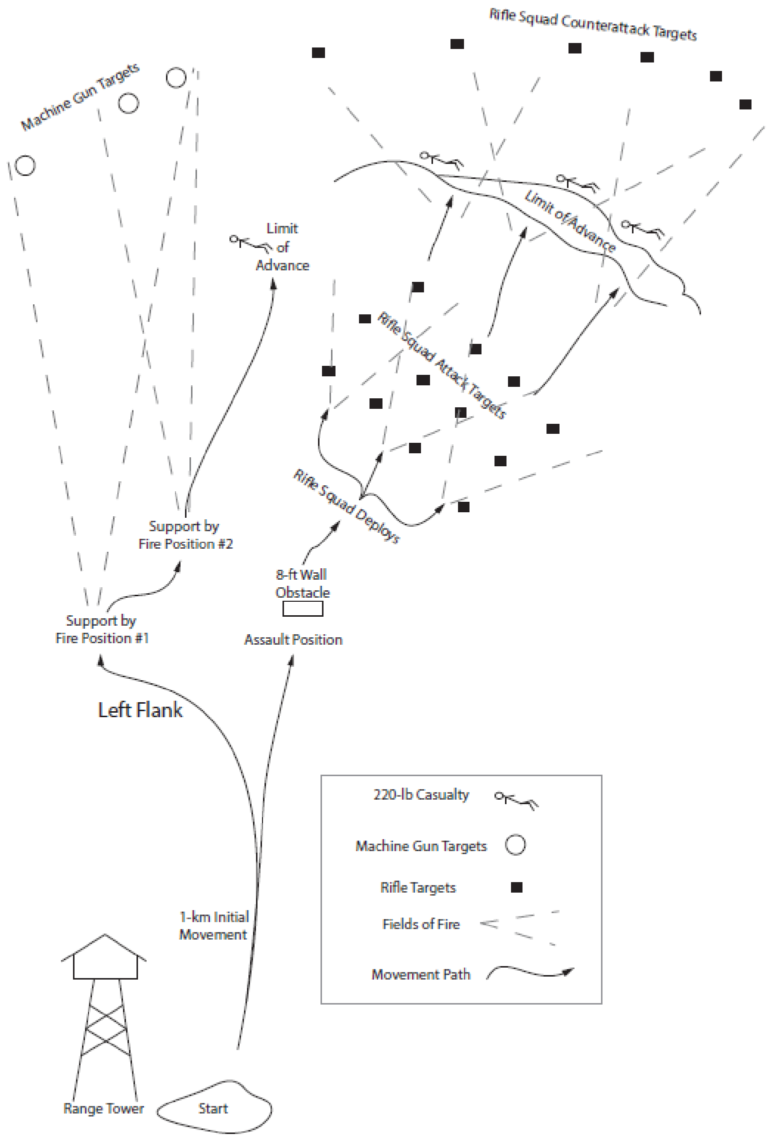


Figure 15. Provisional Infantry and Machine Guns Day-One Scheme of Maneuver

9.2.1.2.1 Initial Movement

(FOUO) The assault began from an assembly area with a rapid movement to assault position. The initial movement was approximately 1 km to the support-by-fire position (PIMG) and assault position (PI). Each squad moved this distance as quickly as possible while carrying an assault load (30 lb). In addition to the assault load, the following additional equipment was carried by PIMG squads:

- (FOUO) PIMG: The M240B machine gun, tripod, spare barrel, and ammunition (600 rounds of 7.62 mm) divided among the three members of the squad (additional 28 – 40 lb per person).

(FOUO) The all-male squads, regardless of PI or PIMG, were significantly faster than the integrated squads. The impact of additional carried weight was not as pronounced in PIMG over this initial movement like it was in the 0331 squads. The disparity grows with distance in the approach march.

9.2.1.2.2 Negotiate Obstacles

(FOUO) Upon arrival at the assault position, the PI squad was required to negotiate an 8-ft wall by getting all Marines and equipment over as quickly as possible. The squad was allowed to negotiate the obstacle any way possible so long as they maintained the appropriate tactical posture.

(FOUO) When comparing low-density integrated PI squads to all-male PI squads, we saw no significant differences in time. The high-density integrated squads were significantly slower than both the low-density integrated and all-male PI squads.

(FOUO) On this particular task, the results of PI squads should not be compared with 0311 squads. The PI squads were faster, but their speed was a function of technique, not strength. The PI squads sent all three fireteams over the obstacle simultaneously, as opposed to the 0311 squads, which went one at a time. In short, the PI squad's technique made them faster than the 0311s, but in the process, the PI squads sacrificed security. This made any comparisons of time problematic.

9.2.1.2.3 Fire and Movement

(FOUO) While the PI squad was negotiating the wall obstacle, the PIMG squad was occupying a support-by-fire position to suppress enemy targets and prepare the battlefield for the PI squad.

9.2.1.2.3.1 Suppressive Fire – Provisional Machine Gun Squad

(FOUO) Following close in timing to the PI squad on the left flank was the PIMG machine gun squad. The machine gun section occupied a position of overwatch and provided support-by-fire for the rifle squad. From the support-by-fire position, the machine gun squad had to rapidly get their gun into action, acquire targets, and accurately engage the enemy.

(FOUO) The movement and emplacement of the machine gun took significantly longer with both low- and high-integrated groups when compared to all-male groups. When the two tasks were broken out separately, the difference in time can be traced to the 100-meter hasty movement from the assault position to the support-by-fire position. Once at the support-by-fire position, there was no significant difference in the time to emplace the machine gun or effects on target (i.e., hits) when comparing integrated to all-male PIMG squads.

9.2.1.2.3.2 Provisional Infantry Rifle Squad Attack and Counterattack

(FOUO) The mission of the rifle squad is "to locate, close-with, and destroy the enemy by fire and maneuver, and repel the enemy's assault by fire and close combat." Having negotiated the 8-ft wall obstacle, the PI rifle squad deployed into an on-line formation to maximize firepower in the direction of the enemy and then moved with weapons at the ready, expecting imminent contact.

(FOUO) Upon contact with the enemy (targets being presented), the Marines established a heavy volume of fire to gain fire superiority. They then began buddy rushes to close with the enemy objective. Each squad conducted approximately 300 meters of buddy rushes, engaging a total of 12 targets with direct-fire weapons and three machine gun bunkers with the M16A4 with M203 grenade launcher attachment.

(FOUO) After all targets were destroyed (no targets remaining), the squad moved another 125 meters to a limit of advance and prepared for an enemy counterattack.

(FOUO) At the conclusion of every assault, the attacking unit must be prepared for the enemy to regroup and organize a counterattack. Upon consolidation at the limit of advance, the squad oriented its weapons in the direction of the enemy's retreat. Upon target presentation, the counterattack commenced for 90 seconds.

(FOUO) The principle measurement of the squad's ability to destroy the enemy by fire during the attack and counterattack is hits on target. When comparing integrated PI squads to all-male squads, the integrated PI squads had similar results for probability of hit.

(FOUO) The squad attack involved all 12 Marines engaging targets. Each Marine can, and often does, engage some of the same targets because of overlapping sectors of fire. To uncover whether or not a gender effect exists in accuracy at the individual level, individual shots recorded by targets were correlated back to shooters. This enabled an investigation of gender differences by weapon system (i.e., M4, M27, and M16A4/M203).

(FOUO) The shot accuracy by gender and by weapon reveals that there is a significant difference between genders for the M4 and M16A4/M203 weapon systems within the PI squads for the probability of hit and the probability of hit/near miss[10]. Figure 16 and Figure 17 depict accuracy results for males versus females by weapon type. The overall accuracy declined and the percent difference increased as the weight of the weapon system increased. The M4 was the lightest weapon and yielded the most

[10] Hits are shots that impact the E-silhouette target. Near misses are shots that are detected within 1-meter of the outside edge of the E-silhouette target, which may have some suppressive effect on the enemy.

accurate results, while the M16A4/M203 was the heaviest weapon and yielded the least accurate results.

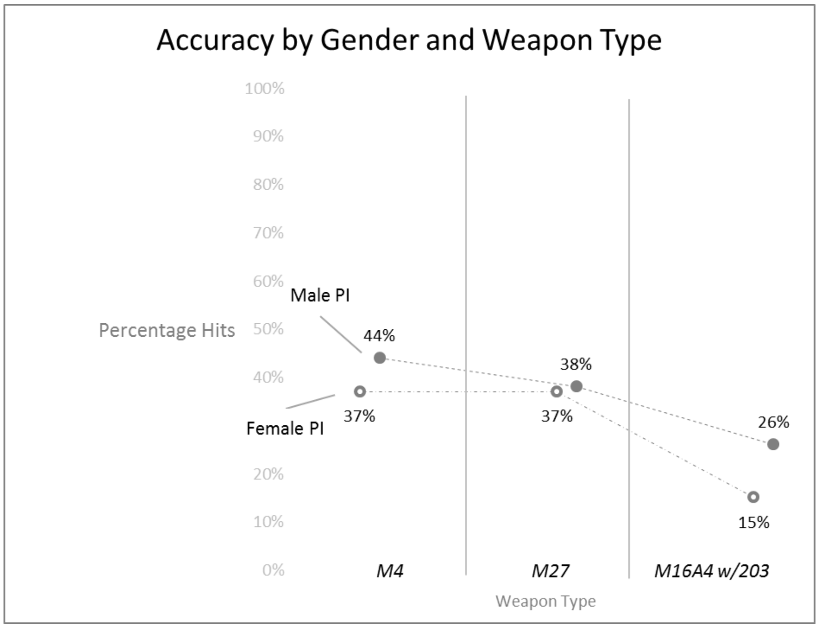


Figure 16. Accuracy Comparison by Gender and Weapon Type for Hits

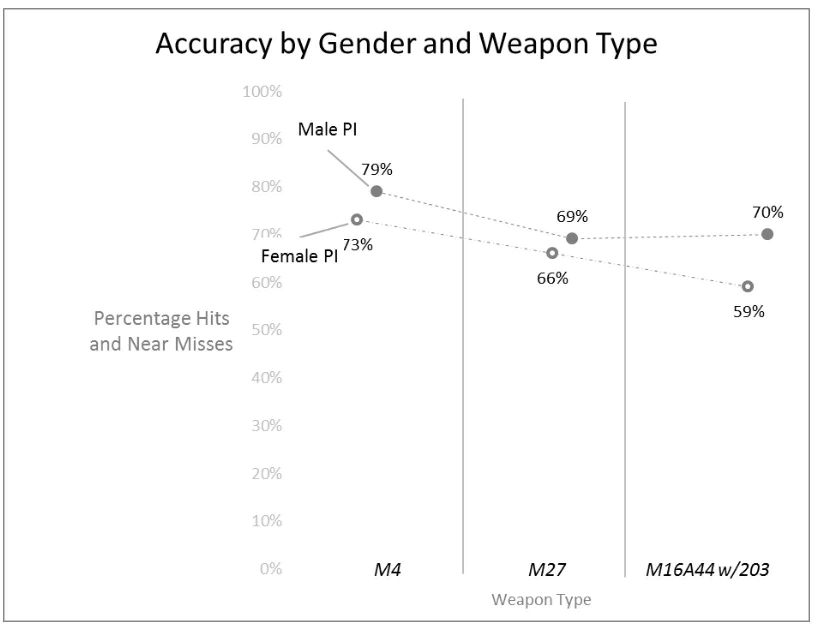


Figure 17. Accuracy Comparison by Gender and Weapon Type for Hits and Near Misses

9.2.1.2.4 Movement to Limit of Advance

(FOUO) The ability to close with the objective after conducting a live-fire attack is crucial to maintaining momentum during offensive operations. At the conclusion of the live-fire attack, the PI squad moved 125 meters to a limit of advance. The PIMG squads also conducted a displacement to a limit of advance, but the distance was 300 meters. The low-density integrated PI and all-male PI squads' performance was similar in the movement to the limit of advance. The high-density integrated PI and integrated PIMG squads took significantly longer to complete their movement to the limit of advance than their all-male counterparts. The late arrival reduced the time to prepare for, or respond to, the counterattack launched by an enemy force.

9.2.1.2.5 Casualty Evacuation

(FOUO) Casualties are an inevitable part of conducting combat operations. Units train to become proficient in the triage, handling, and transport of a casualty. After a Marine is injured, it is essential to move the casualty to the appropriate level of care as quickly as possible. Each PIMG squad was assigned one casualty to evacuate upon conclusion of the live-fire attack and counterattack. PI rifle squads had 12 Marines in a squad as compared to 3 for the PIMG squads. Because the rifle squads had 4 times the number of Marines, each PI rifle squad had 3 casualties to evacuate (i.e., one casualty for every 3 to 4 Marines).

(FOUO) The casualty was a dummy[11] wearing combat equipment that had to be moved a distance of 100 meters to a casualty collection point along with the weapons and gear of the Marines performing the transport. The Marines could use a variety of techniques for transport, but had to carry the dummy off the ground (not drag any part) and could not drop the casualty.

(FOUO) The all-male PI fireteams and PIMG squads were significantly faster than their comparable integrated fireteams and squads. Figure 18 shows the differences observed in casualty evacuations.

[11] The dummy was a full-scale anthropomorphic test device that simulates the dimensions, weight proportions and articulations of the human body. The dummy's weight was equivalent to an average Marine (174 lb). When combat equipment (Marine Corps Combat Utility Uniform, boots, Kevlar helmet, plate carrier vest with small arms protective insert plates, and M16A4) was added to the dummy the total weight increased to 220 lb.

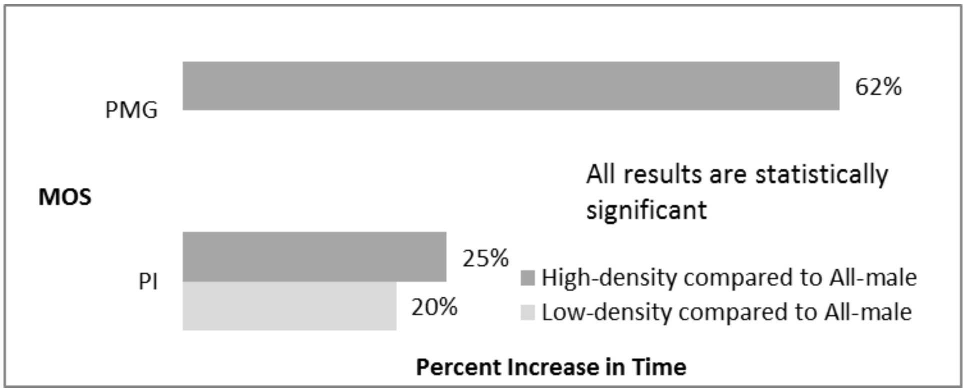


Figure 18. Differences in Integrated Casualty Evacuation Times Compared to All-male Times

9.2.1.2.6 PI and PIMG Day One Fatigue and Workload

9.2.1.2.6.1 Comparing Male Responses to Female Responses

(FOUO) Two self-report surveys (fatigue and workload) were administered to all volunteers at the end of trial activities on day one. The surveys were intended to provide insights into Marines' perceptions of their own level of fatigue and workload during the course of offensive trials.

(FOUO) Overall, males tended to report lower levels of fatigue and workload when compared with females in the PI and PIMG squads.

9.2.1.2.6.2 Comparing Male Responses in Integrated Squads to All-male Squads

(FOUO) In addition to looking at how fatigue and workload self-reports compared between males and females, we also examined whether males responses remained consistent at different integration levels. This comparison was omitted for PIMG squads because of insufficient data for valid comparisons.

(FOUO) The reported fatigue and workload levels of males do not appear to change depending on integration of the squad.

9.2.1.3 Trial Cycle Day Two

(FOUO) The second day of the provisional infantry and provisional machine gun assessment was focused on an approach march for all squads, defensive actions of the rifle squad, and heavy weapons employment of the machine gun squads, as shown in Figure 19. At the conclusion of the defense and weapons employment, the Marines operated under the guidance of their Company leadership, performing tasks that required minimal physical demand and then bivouacking in the field, sleeping in 2-man tents.

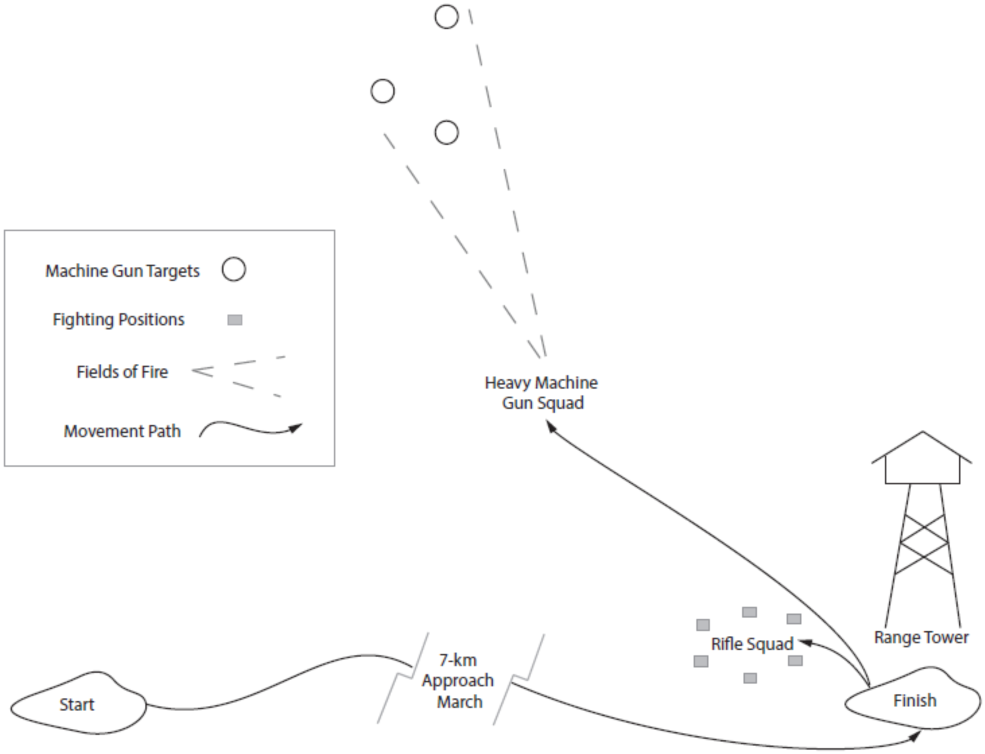


Figure 19. Provisional Infantry and Provisional Machine Guns Day-Two Scheme of Maneuver

9.2.1.3.1 Approach March

~~(FOUO)~~ The beginning of day two started with a march carrying the approach load (55 lb)[12]. Moving under a load is one of the most fundamental tasks of an infantry unit; it is both physically and mentally demanding. Units train by conducting tactical marches with an approach load at increasing distances. The Infantry Training &Readiness Manual states: "the approach march load will be such that the average infantry Marine will be able to conduct a 20 mile hike in 8 hours with the reasonable expectation of maintaining 90% combat effectiveness." Each squad moved a distance of 7.20 km while carrying the approach load and personal weapons. In addition to the approach load, the following additional equipment was carried by PIMG squads:

- ~~(FOUO)~~ PIMGs: The M240B machine gun, tripod, spare barrel, and ammunition (4 cans of 7.62 mm) divided among the three members of the squad (additional 28 – 40 lb per person).

[12] The approach load is in addition to the fighting load (35 lb) and personal weapon (6.4 lb, 7.9 lb, or 10.5 lb depending upon weapon type assigned).

(FOUO) The all-male PI and PIMG squads were significantly faster than the integrated squads. The differences observed between all-male and integrated squads are more pronounced in non-0311 squads, which corresponds to an increase in carried weight of crew-served weapons and ammunition carried in addition to the approach load. Figure 20 shows the percent increase in average times when comparing integrated groups to the all-male groups.

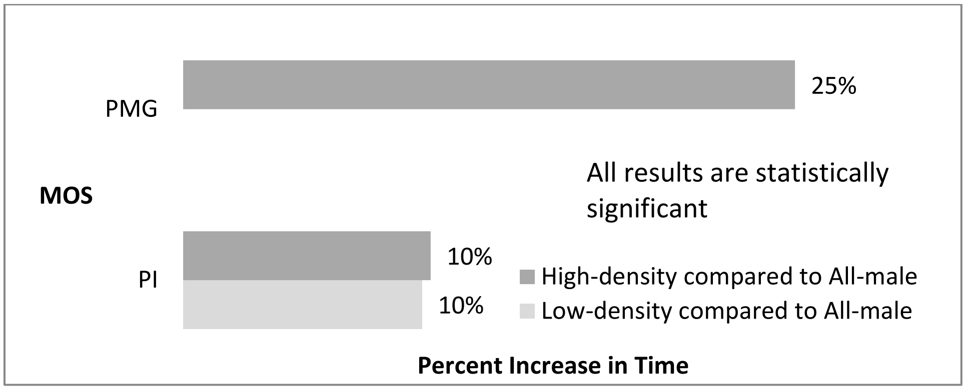


Figure 20. Percent Increase in Approach March Times for Integrated Squads Compared to All-male Squads.

(FOUO) The differences illustrated in Figure 20 follow the same pattern observed in the approach march for 03XX squads. The march with approach load has an applicable standard that can be used to provide a second frame of reference. Similar to the 03XX squads, the movement over 20 miles in 8 hours equates to a pace of 4 kilometers per hour (km/h). This standard is applicable to PI and PIMG squads because Marines from non-03XX MOSs assigned to Infantry units are required to keep pace with the 03XX Marines.

(FOUO) Figure 21 shows a conversion of the average squad times over the 7.2-km distance to pace. This conversion enables a comparison to the 4-km/h required to meet the standard. As shown in the figure, the integrated PI squads and both all-male squads achieved the standard; the integrated PIMG squads were not able to meet the standard. Figure 21 also illustrates the same corresponding decrease in pace as weight of the carried load increases that was seen in the 03XX Marines.

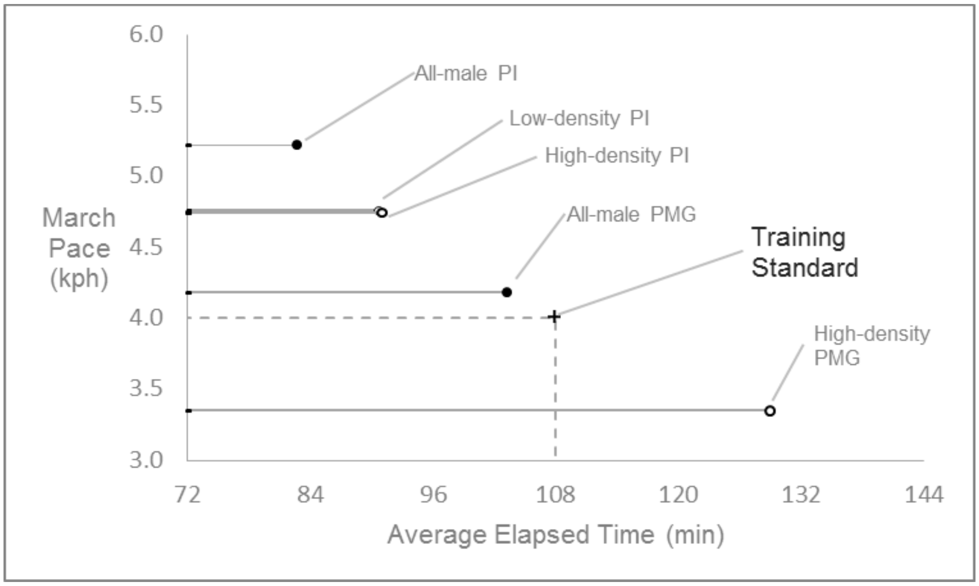


Figure 21. Conversion from Average Elapsed Time for the 7.2 km March to Pace for the Provisional's Approach March

9.2.1.3.2 Heavy Weapons Employment

(~~FOUO~~) Upon completion of the approach march and after a brief rest period, the infantry moved into the defense. The PI Marines moved into the preparation of defensive fighting positions while the PIMG Marines moved into heavy weapons employment.

9.2.1.3.2.1 PI Fireteams Preparing Defensive Fighting Positions

(~~FOUO~~) Protection is vital to infantry units, especially in a defensive posture. Infantry units commonly construct fighting positions to conceal positions and minimize exposure to enemy fire. There are many characteristics to constructing a doctrinal fighting position, but the most physically demanding aspect is digging out the fighting hole. The terrain where the fighting holes were dug consisted of hard, compact sand and rocks. Each PI fireteam dug two 2-man fighting holes in a time limit of 2 hours, maintaining 50% security, meaning two Marines dug while two provided security in the prone position. The Marines swapped positions every 15 minutes. During the course of this task, the earth was scooped into buckets and weighed.

(~~FOUO~~) The integrated PI fireteams average weight of earth moved was not significantly different than all-male groups.

9.2.1.3.2.2 PIMG Squad Employment of M2 Machine Gun

(~~FOUO~~) Providing defensive fires with the M2 heavy machine gun entails moving the system to a position of advantage, engaging the enemy, and conducting a rapid displacement. After completing the approach march, the PIMG squads dismounted the

M2 heavy machine gun from a vehicle platform, moved and emplaced the weapon, engaged targets, displaced, and mounted the weapon onto a vehicle platform.

(FOUO) To dismount the weapon system, the squad worked to manually disassemble and lower each component to the ground from a HMMWV. To mount the weapon system on the HMMWV, the process was reversed. This task determined the time for a squad of three Marines to fully dismount and mount the M2 from a tactical vehicle and required the strength to lift, manipulate, and lower heavy components.

(FOUO) The integrated PIMG squads' average time was significantly faster than all-male squads' times for the aggregate mounting and dismounting of the M2 heavy machine gun. When broken down into individual tasks, the all-male group took significantly longer to mount the M2. The time to dismount for both types of squads was the same.

(FOUO) During the assessment, each machine gun squad emplaced the M2 on a tripod at a specified firing location. The squad engaged targets as they began to be exposed while hits on, and around, the target were recorded by sensors. The assessment began with targets exposed and engaged by the squad.

(FOUO) The integrated PIMG squads had significantly more hits on target than the all-male squads. Figure 10 (see page 25) shows the percentage of hits and near misses[13] on the machine gun targets. The finding that integrated groups have a higher percentage of hits on target is somewhat unusual given the findings from PI and 0311 hit percentages. To see if there is a consistent finding, the results of provisional machine gunners were compared to the 0331 squads. As shown in Figure 10, the trend of integrated groups outperforming all-male groups is repeated and consistent.

(FOUO) After an engagement, a machine gun squad must be able to breakdown their weapon system and move to a position of cover as rapidly as possible. Immediately after completing target engagement, the PIMG squad took the gun out of action and displaced 100 meters from the firing line to a designated location, moved a heavy load a short distance, and manipulated a weapon while fatigued.

(FOUO) The integrated PIMG squads took significantly longer to displace from the firing line compared with all-male squads.

[13] For experimental purposes a hit or near miss for a machine gun target is defined as any detectable round within a 3-meter radius from the center of the target.

9.2.1.3.3 PI and PIMG Day Two Fatigue and Workload

9.2.1.3.3.1 Comparing Male Responses to Female Responses

(FOUO) Two self-report surveys (fatigue and workload) were administered to all volunteers at the end of trial activities on day two. The surveys were intended to provide insights into Marines' perceptions of their own level of fatigue and workload level during the course of defensive trials.

(FOUO) Overall males tended to report lower levels of fatigue and workload when compared to females in the PI and PIMG squads.

9.2.1.3.3.2 Comparing Male Responses in Integrated Squads to All-male Squads

(FOUO) In addition to looking at how fatigue and workload self-reports compared between males and females, we also examined whether male responses remained consistent at different integration levels. This comparison was omitted for PIMG squads because of insufficient data for valid comparisons.

(FOUO) PI males report lower levels of fatigue and workload in high-density integrated groups when compared with all-male and low-density integrated groups. This finding is consistent with the 03XX results, where integrated squads, especially high-density squads, moved at a slower (easier) pace on the 7-km approach march, resulting in lower workloads on males in those squads.

9.2.2 Engineers

(FOUO) The combat engineer assessment was a 2-day scheme of maneuver intended to replicate offensive and defensive tasks. All Marines, regardless of task, wore the fighting load (35 lb) and carried a personal weapon (6.4 lb, 7.9 lb, or 10.5 lb, depending upon weapon type).

9.2.2.1 Trial Cycle Day One

(FOUO) The first day of the combat engineer assessment was a breaching mission in support of an infantry squad attack followed immediately by a demolitions raid. The ability for an infantry unit to quickly overcome unexpected or lightly defended obstacles to maintain the momentum of the attack is critical to mission success. The reinforced infantry squad was supported by two engineer fireteams, each responsible for conducting a hasty breach using Bangalore torpedoes to overcome a concertina wire obstacle once the advancing infantry squad halted.

(FOUO) At the conclusion of the squad attack and successful breach, the two combat engineer fireteams formed an engineer squad and moved to a likely avenue of enemy approach to conduct a demolitions raid and emplace a countermobility obstacle to further disrupt an enemy's mobility and prevent the loss of recently acquired terrain.

(FOUO) Day one was a fluid scenario that allowed the Marines to flow from beginning to end without interruption or breaks. Figure 22 is an illustration of the combat engineer attack. At the conclusion of the demolitions raid, the Marines operated under the guidance of their platoon leadership, performing tasks that required minimal physical demand and then bivouacking in the field, sleeping in 2-man tents.

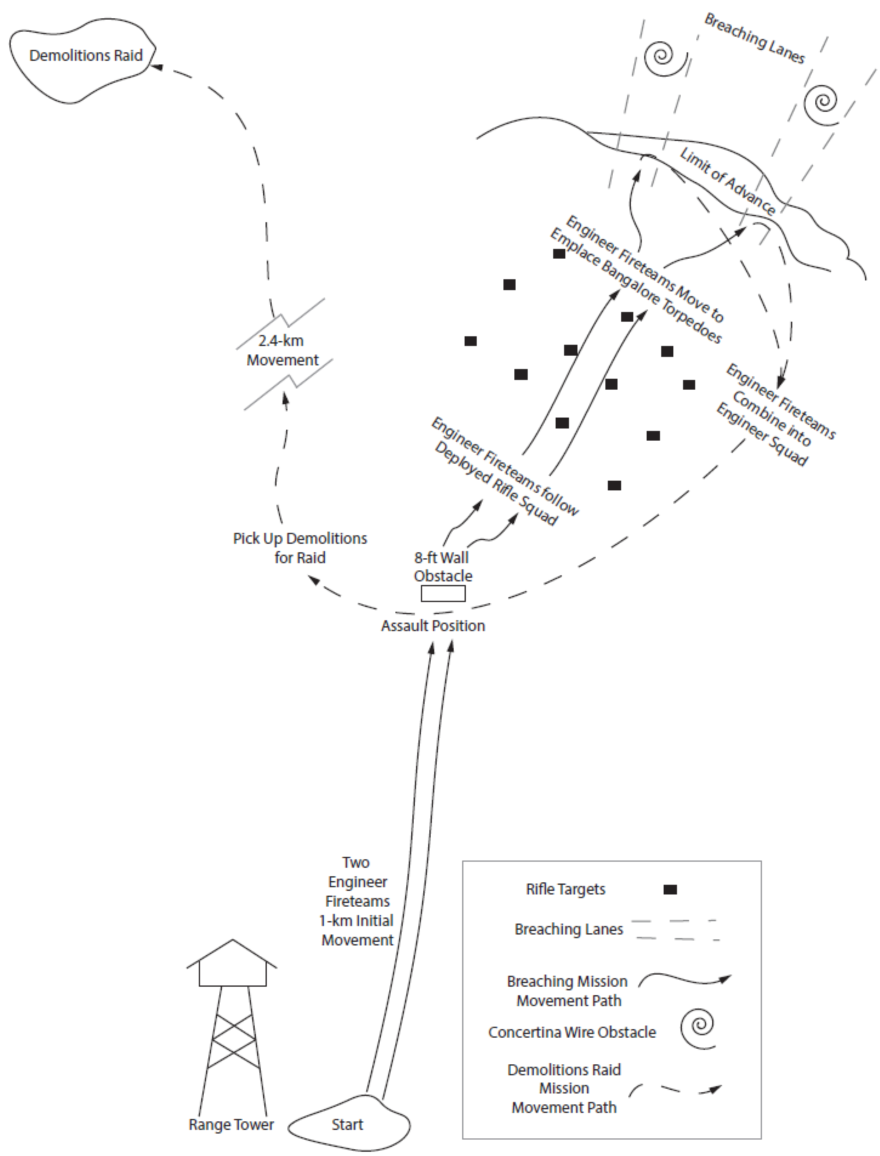


Figure 22. Combat Engineers Day-One Scheme of Maneuver

9.2.2.1.1 Initial Movement

(FOUO) The employment of forces on the battlefield requires maneuver in combination with fire, or fire potential, to achieve a position of advantage in respect to the enemy in order to accomplish the mission. During the experiment, the first combat engineer task was a foot-movement following in trace of an infantry squad for a distance of approximately 1 km to an assault position. Each fireteam moved this distance as quickly as possible while carrying an assault load (20 lb). In addition to the assault load, every individual Marine carried either two or three Bangalore components, weighting 15 lb each, depending on their individual billet.

(FOUO) The all-male fireteams were significantly faster than the integrated fireteams.

9.2.2.1.2 Negotiate Obstacles

(FOUO) Upon arrival at the assault position, the combat engineer Marines were required to negotiate an 8-ft wall by getting all Marines and equipment over as quickly as possible. The fireteams were allowed to negotiate the obstacle any way they could, as long as they maintained the appropriate tactical posture.

(FOUO) The all-male fireteams were significantly faster than the integrated fireteams.

9.2.2.1.3 Breaching

(FOUO) After negotiating the obstacle, the engineer fireteams conducted a 450-meter movement in trace of the infantry squad to the Limit of Advance. Upon arrival at the Limit of Advance, the fireteams held in place until the infantry squad completed the attack.

(FOUO) At the conclusion of the infantry squad's live-fire attack and casualty evacuation, the combat engineer fireteams conducted a hasty breach using Bangalore torpedoes to breach a concertina-wire obstacle out in front of the halted infantry squad. Once the fireteam had all 10 Bangalore sections connected, the explosives were dually primed and the engineers initiated the breach via Modern Demolition Initiators (MDI). Breaching operations enables forces to neutralize, reduce, or overcome obstacles/impediments as soon as possible to allow unrestricted movement and maintain the initiative. The all-male fireteams were significantly faster than the integrated fireteams.

9.2.2.1.4 Mobility Denial

(FOUO) The most obvious way to seize and maintain the initiative is to strike first and keep striking. Once ground is taken, it requires continuous action to hold and maintain. Engineers apply this fundamental tactic by denying the enemy's ability to influence mobility. This is accomplished by being proactive and, often, by conducting countermobility missions to deny the enemy advancing momentum and by providing protection to friendly maneuvering forces. Upon conclusion of the attack, the combat engineers conducted a demolitions raid by moving a 2.4-km distance to a likely avenue of approach with 40-lb shape and 40-lb crater charges to conduct mobility denial operations.

(FOUO) There was no significant difference in the movement under load between the all-male squads and the low-integration groups; however, once the squad was formed as a higher-integration squad, the movement times were significantly slower.

9.2.2.1.5 Fatigue and Workload

9.2.2.1.5.1 Comparing Male Responses to Female Responses

(FOUO) Two self-report surveys (fatigue and workload) were administered to all volunteers at the end of trial activities on day one. The surveys were intended to provide insights into Marines' perceptions of their own levels of fatigue and workload during the course of offensive trials.

(FOUO) Overall, male engineers tended to report fatigue levels that were similar to fatigue levels reported by female engineers, but male engineers reported lower workload levels than female engineers.

9.2.2.1.5.2 Comparing Male Responses in Integrated Squads to All-male Squads

(FOUO) In addition to looking at how fatigue and workload self-reports compared between males and females, we also examined whether male responses remained consistent at different integration levels.

(FOUO) No differences were seen in the reported fatigue and workload levels of males depending on integration of the squad.

9.2.2.2 Trial Cycle Day Two

(FOUO) The second day of the combat engineer assessment focused on defensive actions of the engineer squad, and included a 7.2-km forced march, dismounted route clearance operations, and a cache reduction evolution, as shown in Figure 23.

(FOUO) At the conclusion of the defensive day, the Marines operated under the guidance of their platoon leadership, performing tasks that required minimal physical demand and then bivouacking in the field, sleeping in 2-man tents.

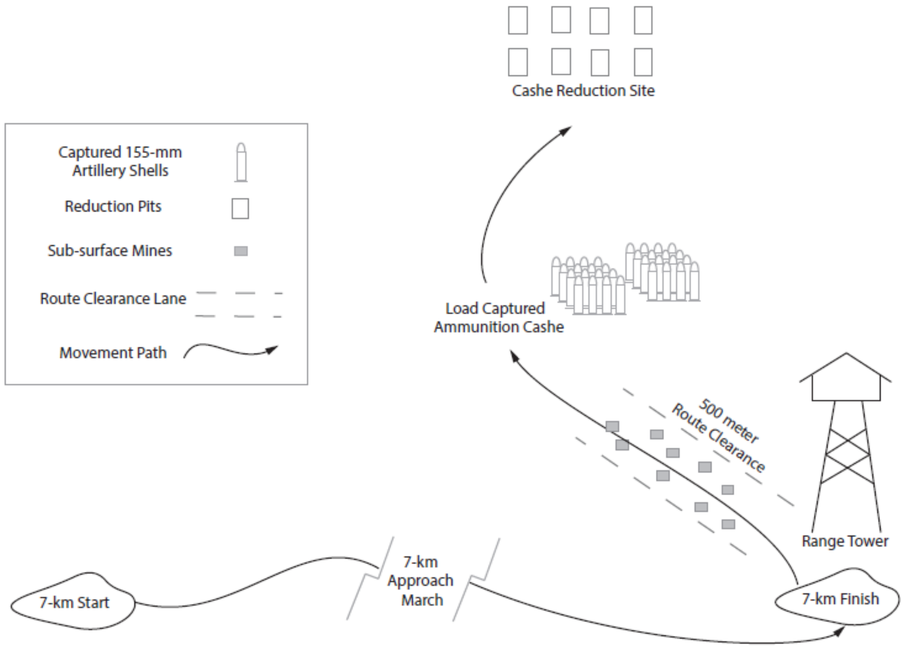


Figure 23. Combat Engineers Day-Two Scheme of Maneuver

9.2.2.2.1 Approach March

(~~FOUO~~) The beginning of day two started with a march carrying the approach load (40 lb)[14]. Combat Engineers serving in the closed Combat Engineer Battalions work in direct support of, or are often attached to, Infantry units at various levels. As such, the Combat Engineers must be able to move through all sorts of terrain by foot. Moving under a load is one of the most fundamental tasks of a combat engineer in support of an infantry unit; it is both physically and mentally demanding. Each squad moved a distance of 7.20 km while carrying the approach march load and personal weapons.

(~~FOUO~~) While the all-male squads were significantly faster than the integrated squads, all squads met the training standard. Figure 24 shows a conversion of the average squad times over the 7.2-km distance to pace. This conversion enables a comparison to the 4-km/h required to meet the standard.

[14] The approach load is in addition to the fighting load (35 lb) and personal weapon (6.4 lb, 7.9 lb, or 10.5 lb depending upon weapon type assigned).

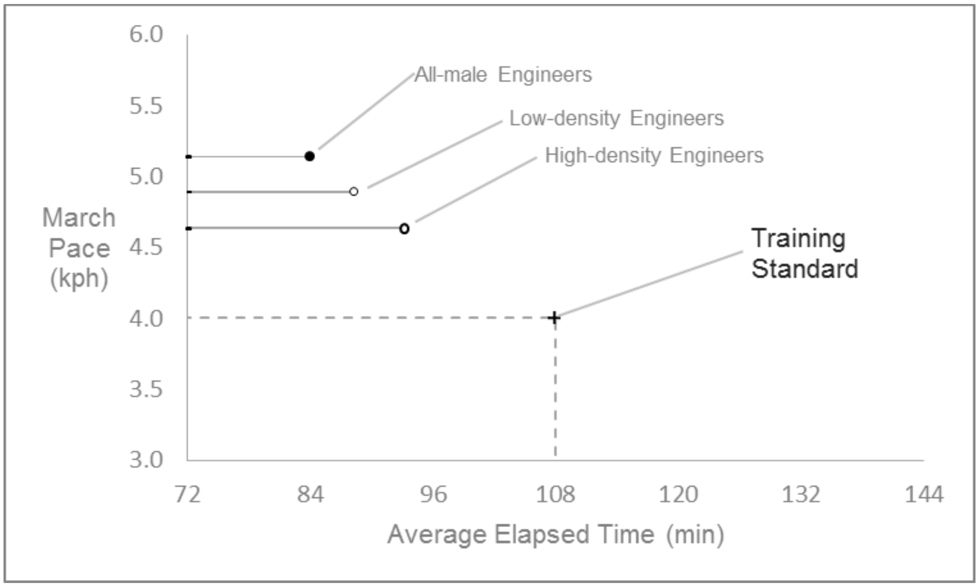


Figure 24. Conversion from Average Elapsed Time for the 7.2-km March to Pace for the Combat Engineers Approach March

(~~FOUO~~) Similar to the 03XX squads, the movement over 20 miles in 8 hours equates to a pace of 4 kilometers per hour (km/h). This standard is applicable to Engineer squads because they are frequently assigned to Infantry units and are required to keep pace with the 03XX Marines.

9.2.2.2.2 Route Sweep

(~~FOUO~~) Advancing foot mobile forces may require units to conduct route clearance missions to ensure that the lines of communication enable safe passage of combat, combat-support, and combat-service-support organizations. The combat engineers are required to perform these route clearance tasks in support of ground forces as a way to reduce casualties and equipment losses on the battlefield. Upon completion of the 7.2-km forced march, the combat engineer squad was required to conduct deliberate route clearance using mine detectors and sweeping a 500-meter lane. Time was not a component of this task; the squads were assessed based on their detection rate.

(~~FOUO~~) There was no significant difference in detection rates between all-male squads and integrated squads.

9.2.2.2.3 Destroy Captured Munitions

~~(FOUO)~~ At the conclusion of the route clearance mission, the engineer squads unearthed a large ammunition cache consisting of (32) 155-mm artillery shells, each shell weighing 95 lb. To deny the enemy access to the munitions, as well as to ensure the safety of the surrounding population, the engineers were required to move artillery shells to a safe location and destroy them with demolitions. During this task, the squad loaded the ammunition onto the bed of a 7-ton truck, transported it to a safe demolition site, and then unloaded, buried, and subsequently destroys all 32 artillery shells.

(FOUO) There were no significant differences in times for loading, digging, unloading, or rigging for detonation between the all-male squads and the integrated squads. One reason for the lack of observed time difference was noted based on the division of labor. In the most demanding portion of the task, loading the ammunition onto the bed of the 7-ton truck, the female Marines were positioned on the bed of the vehicle to aid in the loading and stowing of the ammunition for transport rather than at the base of the vehicle. This resulted in the male Marines lifting the 95-lb artillery shells up from the ground at least 62 inches or higher to the bed of the 7-ton truck.

9.2.2.2.4 Fatigue and Workload

9.2.2.2.4.1 Comparing Male Responses to Female Responses

(FOUO) Two self-report surveys (fatigue and workload) were administered to all volunteers at the end of trial activities on day two. The surveys were intended to provide insights into Marines' perceptions of their own levels of fatigue and workload during the course of defensive trials.

(FOUO) Overall, male engineers tended to report fatigue levels that were similar to female engineers, but male engineers reported lower workload levels than female engineers.

9.2.2.2.4.2 Comparing Male Responses in Integrated Squads to All-male Squads

(FOUO) In addition to looking at how fatigue and workload self-reports compared between males and females, we also examined whether male responses remained consistent at different integration levels.

(FOUO) The reported fatigue level of engineer males does not appear to be different based on squad integration level. Workload level of male engineers does appear to depend upon integration level, but it does appear to depend on whether they are in high-density integrated squads. This finding is consistent with the 03XX and PI results where integrated squads, especially high-density squads, move at a slower (easier) pace on the 7-km approach march, which results in lower workloads on males in those squads.

9.2.2.3 Mountain Warfare (Open MOSs)

(FOUO) Similar to the infantry, the mountaineering assessment replicated a logistical resupply of a forward-staged squad while moving in a mountainous environment. Each 12-Marine squad departed an assembly area located at a lower base camp where they hiked 4.6 km with a 75-lb pack and personal weapon (M4) to an objective rally point.

(FOUO) The route for this movement was on an unimproved, hilly surface; the terrain was hard, rocky, and uphill, with an approximate gain in elevation of 175 meters, as shown in Figure 25.

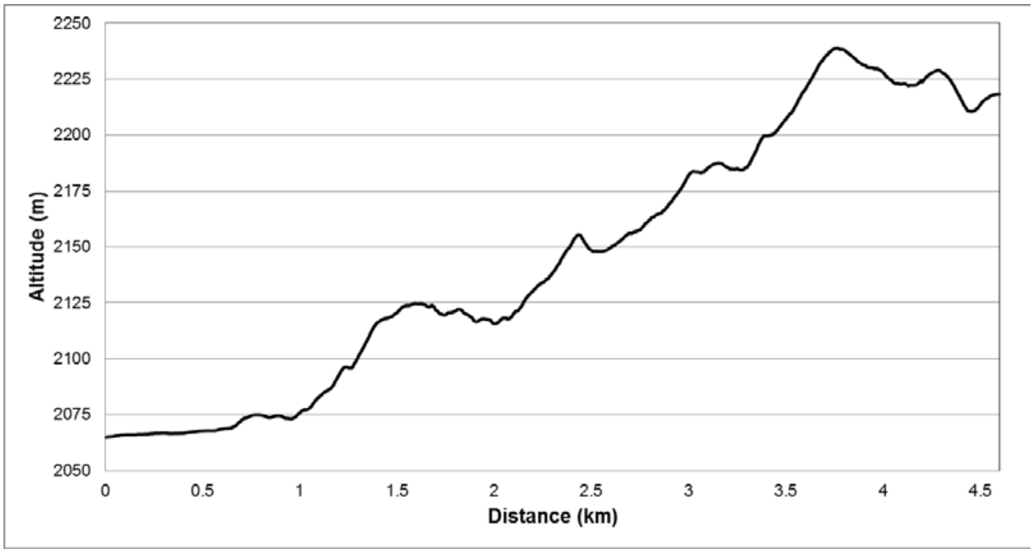


Figure 25. Elevation Change in Meters over 4.6-km Hike Distance

(FOUO) From the objective rally point, the squad crossed a 200-ft gorge via two single-rope bridges. Once across, the squad pulled their packs across and made a short movement to a location requiring a cliff ascent. Each Marine executed a 40-ft cliff ascent using two climbing lanes. The final action was to conduct a return foot-march to lower base camp. The return march was 5 km while wearing the 75-lb pack and personal weapon across terrain similar to the ascent, but downhill. The entire resupply mission took approximately 5 hours to complete.

(FOUO) The mountaineering tasks were non-MOS specific. Due to the non-MOS-specific nature of the tasks, all Marines from provisional infantry, provisional machine guns, and engineers were combined for the purposes of forming 12-Marine squads. Two types of squads were formed each day, all-male and high-density integration squads[15].

(FOUO) Every Marine received 1 day of recovery after each execution day over the course of 15 days. At the end of a day and during off-days, they lived in squad bays at lower base camp and operated under Company leadership, performing tasks that required minimal physical demand.

(FOUO) The all-male squads were faster than integrated squads on hikes, the gorge crossings, and cliff ascents.

[15] High-density integrated squads had six females per squad.

(FOUO) Fatigue and workload self-report surveys were also administered at the end of mountain warfare activities, in particular after the ascent hike and upon return to lower base camp. When comparing males to females, males reported lower fatigue and workload results. Males also reported lower fatigue, but similar workload levels, when in integrated squads compared to when they were in all-male squads.

9.2.3 Open MOS Summary

Figure 26 presents a summary of the differences observed when comparing all-male squads to integrated squads. The horizontal bars depict the percentage change observed when comparing the average all-male squad result to an average integrated squad result. The numerical percentage presented adjacent to the bar is indicative of statistical significance. Statistical significance is related to the size of the difference, variation, and the number of trials for that task. This explains why some larger bars show no significance.

When a bar shifts to the right of the centerline the all-male group's average is better than the integrated group's. When a bar shifts to the left the integrated group's average is better than the all-male's. Most of the horizontal bars are shifted to the right for each MOS indicating that all-male groups typically outperform integrated groups.

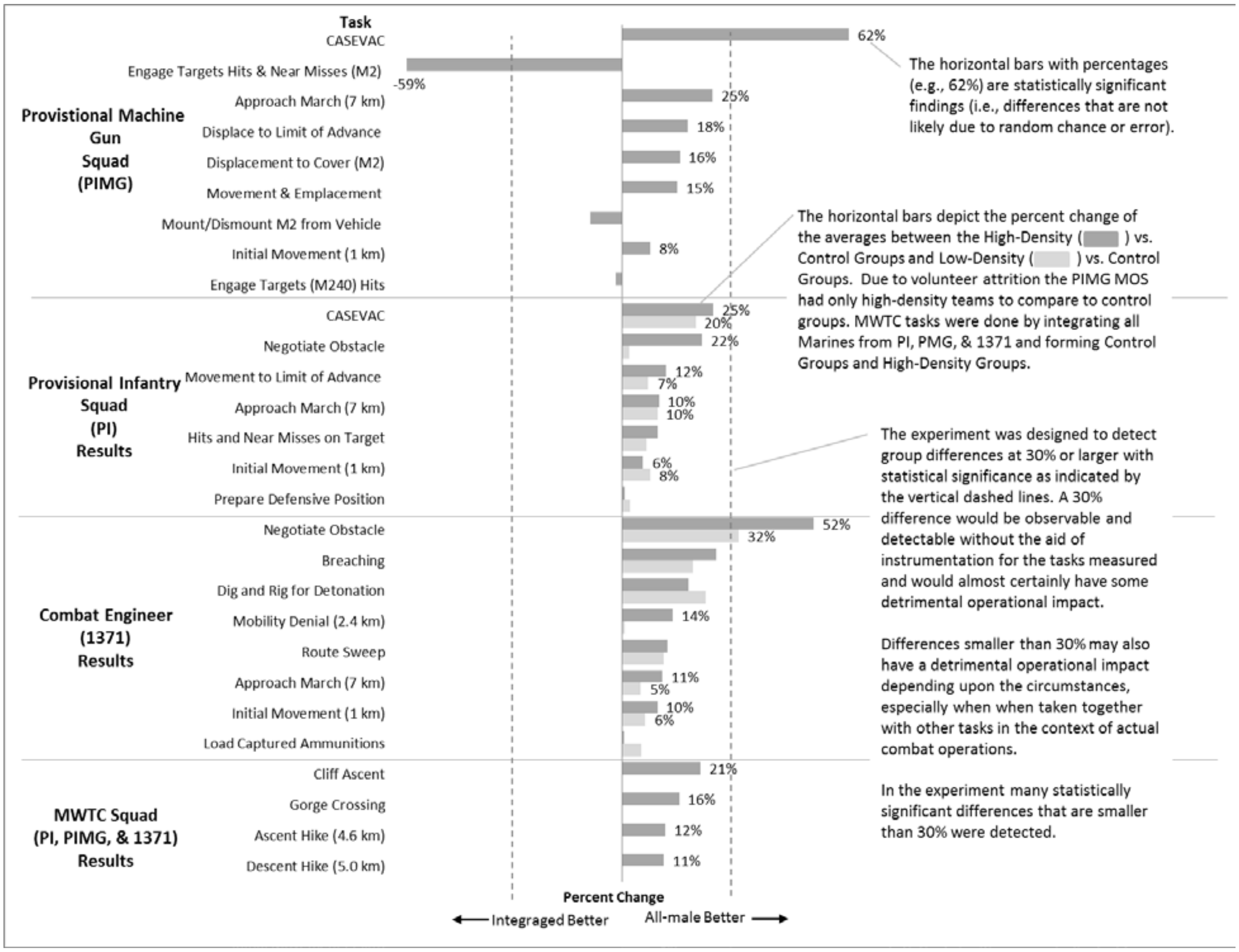


Figure 26. Open MOS Summary Comparison of Percentage Change of All Tasks

9.3 Objectives 1 and 2: Correlating Individual Physical Characteristics to Collective Mission Effects

(FOUO) The previous sections within this objective discussed results only as they pertain to differences due to integration level and critical billets. The goal of statistical modeling, as applied here, is to estimate, simultaneously, the effect of gender-integration levels and other relevant individual variables on squad and team performance.

(FOUO) For selected tasks described in the annexes and where possible, we examined the following individual, physical characteristics by Closed and Open MOSs:

- (FOUO) Age
- (FOUO) Height
- (FOUO) Weight
- (FOUO) Armed Forces Qualification Test score

- (~~FOUO~~) General Technical score
- (~~FOUO~~) Combat Fitness Test Movement to Contact time
- (~~FOUO~~) Combat Fitness Test Maneuver Under Fire time
- (~~FOUO~~) Physical Fitness Test crunches
- (~~FOUO~~) Physical Fitness Test 3-mile run time
- (~~FOUO~~) Rifle score

(~~FOUO~~) What we found is that there was no individual, physical characteristic that was statistically significant and that had a practical impact on the model. Each time an individual, physical characteristic was statistically significant in a model, the effect was practically negligible, conflicting, and/or incomplete for the squad or team. That is, there were almost no tasks for which a variable is significant for all, or even most, members of a unit.

(~~FOUO~~) Integration level, however, consistently appears as statistically significant for modeled tasks, and its effect is clear, causal, and practical. Therefore, integration level is the best variable to describe performance for each task.

9.4 Objectives 1 and 2: Task Cohesion

(~~FOUO~~) Task cohesion in the experiment refers to an individual's beliefs about team closeness, similarity, and bonding around the group's task. The task cohesion survey was an excerpt from a previously designed survey instrument, and was administered upon conclusion of each MOS's cycle. The task cohesion survey was intended to provide insights into Marines' perceptions of how cohesive their unit was over the course of the run cycle. The cohesion survey was broken into five parts with an overall score given that sums the responses from each of the five questions.

(~~FOUO~~) Overall, the Marines reported high levels of cohesion across all MOSs and genders. The MOS with the lowest cohesion result was the 0311s, but these results were still positive with respect to cohesion. We further examined the cohesion results in two different ways. The first comparison examined male responses vs. female responses, by MOS, to determine if there were any significant differences in responses. Figure 27 displays the results of the MOSs broken out into three rows. The MOSs in the top row are those where males report significantly higher cohesion scores than females. The MOSs in the middle row indicates no significant difference based on gender. The bottom row indicates where female cohesion results are reported as significantly higher than males.

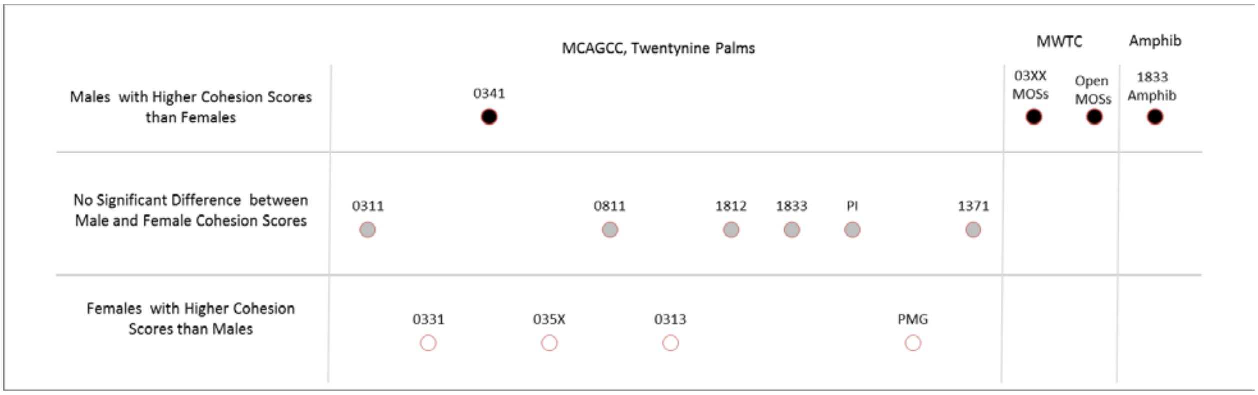


Figure 27. Comparison of Male vs. Female Cohesion Survey Responses

(FOUO) During the first 2 months of the experiment in Twentynine Palms, just over half of the comparisons (6 of 11) showed no significant difference in cohesion based on gender comparisons. Females reported higher cohesion than males in 4 of 11 MOSs with only one MOS (0341) having males reporting higher cohesion results then females. The last month of the experiment at Mountain Warfare Training Center (MWTC) and Camp Pendleton (Amphib) showed a change in the cohesion results.

(FOUO) At MWTC, all 03XX Marines were combined into one group, while PI, PIMG, and 1371 were combined into a second group. What we saw was a change in opinion, where males reported higher cohesion scores than females, although given all the factors that changed going from Twentynine Palms to Bridgeport, we cannot be certain what caused the change in results. The Amphib portion showed a similar change in cohesion results, where the results showed males leaning more favorably than females by the end of the experiment. Again, we cannot be certain of what caused the change.

(FOUO) The second method of examining cohesion results considered the results of only male participants, by MOS and integration level of the squad or crew. We performed this analysis to determine whether males' cohesion responses remained consistent at different integration levels.

(FOUO) What we found is shown in Figure 28 and indicates that almost universally males indicate higher levels of cohesion when working in all-male groups compared to integrated groups, regardless of MOS or location. The one exception was the Marines in the Open MOSs. In the Open MOSs at MWTC, we do see a small, but visible, increase in the percentage of males who rate all-male squads higher than integrated squads - though this is not enough to be statistically significantly. Provisional Machine Guns was omitted from the second analysis because there was insufficient data to draw any meaningful conclusions.

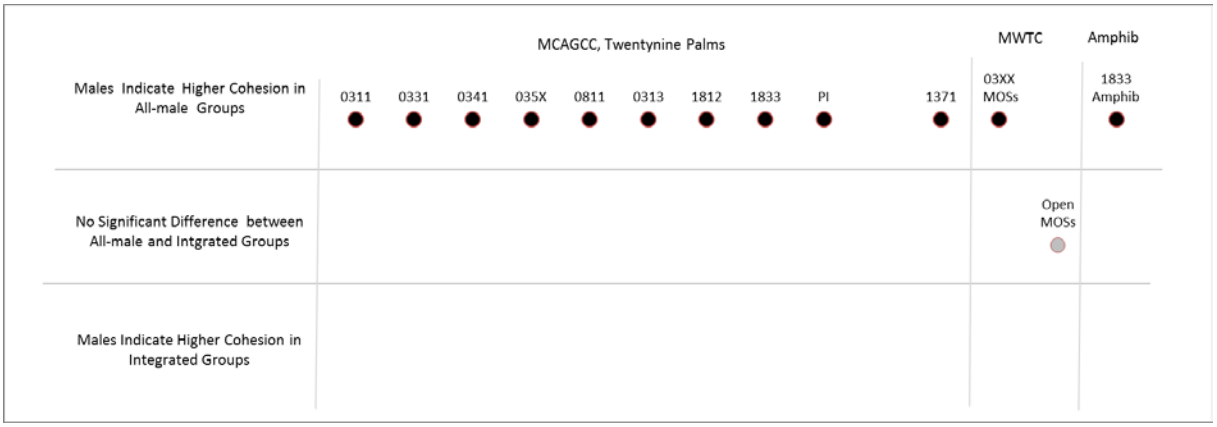


Figure 28. Comparison of Male Results in All-male vs. Integrated Groups, by MOS

9.5 Objective 3

9.5.1 Readiness

(FOUO) Readiness, at its most basic level, is a measure of the combat power available to respond to mission demands. Individual Readiness is a measure of a person's availability to perform required military tasks on a given day. An individual can be unavailable for a variety of reasons, to include misconduct, and administrative or medical conditions, just to name a few. For the purposes of addressing readiness of the volunteers at the GCEITF, we restricted our analysis to medical readiness.

(FOUO) The readiness dataset contains observations for each volunteer's medical readiness at an individual level. The observed time began when the Marine joined the GCEITF and ended at the completion of the experiment. The completion of the experiment for an individual could either be the actual end of the experiment, when a Marine DORed, or when a Marine was dropped (due to medical reasons) from the experiment by GCEITF leadership.

(FOUO) When comparing males to females, males were available 98.4% of the time while females were available 96.8% of the time. The cause of lower availability for females was their likelihood of injury. Females were more likely to be injured than males. However, once an injury occurs, the recovery time for males and females are similar. The breakout of injury sources reveals the greatest source of lost days for females was caused by occupational injuries, whereas males lost more days due to non-occupational causes. Table 3 displays a summary breakout, by cause and gender for unavailable days.

Table 3. Percentage Breakout of Unavailable Days by Gender and Causality

	Occupational	Non-Occupational

Male	40%	60%
Female	80%	20%

(FOUO) Our comparison of male and female injuries considers the fact that Marines are capable of incurring one or more injuries/illnesses during a time period. A descriptive comparison of the number of injuries and illnesses is displayed in Figure 29.

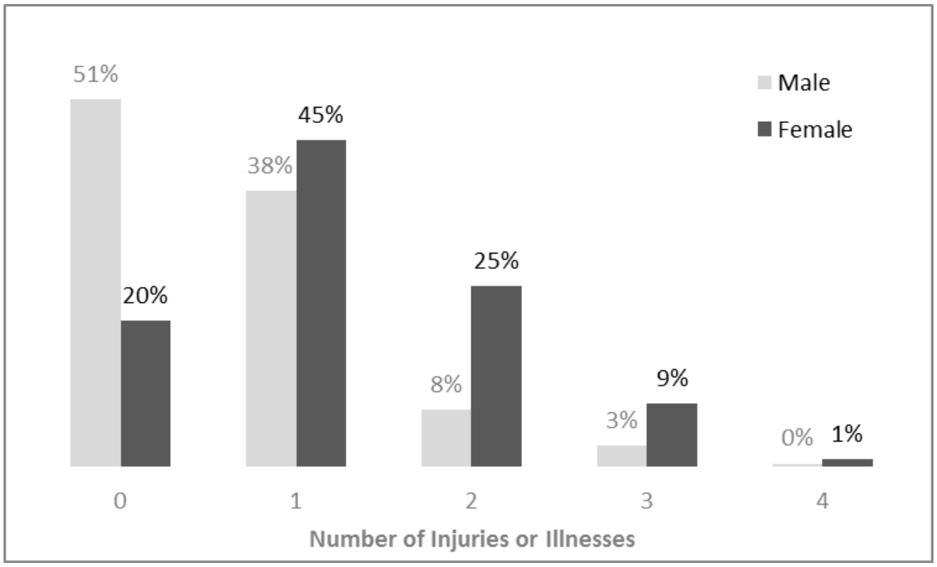


Figure 29. Percentage of Volunteers, by Gender, with One or More Injuries or Illnesses

(FOUO) We analyzed the readiness data by gender, taking into consideration explanatory factors, and considered the fact that Marines often incur more than one injury or illness during a specified time period.

(FOUO) We found that the only personnel variables that consistently stood out as predictors of readiness were gender and MOS type. In particular, with the exception of non-occupational injuries, outcomes were always worse for female volunteers. Additionally, Marines in vehicle MOSs tended to have lower injury rates than those in hiking or Artillery MOSs. Finally, the CFT MANUF time appears several times as a predictor of readiness, so we recommend further investigating the relationship of this measure with injury rates.

9.5.2 Proficiency and Conduct

(FOUO) No valid results were available to report; see paragraph 5.2.

10. Conclusions

10.1 Objectives 1 & 2

(FOUO) The female Marines integrated into the closed MOS units demonstrated that they are capable of performing the physically demanding tasks, but not necessarily at the same level as their male counterparts in terms of performance, fatigue, workload, or cohesion.

(FOUO) Integrated units, compared with all-male units, showed degradations in the time to complete tasks, move under load, and achieve timely effects on target. The size of the differences observed between units and tasks varied widely. The more telling aspect of the comparisons is the cumulative impacts. The pace, timing, and accuracy of any singular task is not necessarily important, but taken together, and in the context of actual combat operations, the cumulative differences can lead to substantial effects on the unit, and the unit's ability to accomplish the mission.

10.2 Objective 3

(FOUO) Gender and MOS type are the best predictors of occupational injuries. In particular, we found that females are more likely to incur occupational injuries, resulting in reduced readiness compared to their male counterparts. Males, on the other hand, are more likely to incur non-occupational injuries. Additionally, Marines in vehicle MOSs tended to have lower injury rates than those in MOSs that march (i.e., foot mobile) or Artillery MOSs.

(FOUO) No clear conclusions can be drawn from the Proficiency and Conduct ratings of the GCEITF volunteers.

11. References

a) MCOTEA. Ground Combat Element Integrated Task Force Experimental Event Plan. January 2015

b) MCOTEA. Ground Combat Element Integrated Task Force Experimental Data Report. May 2015

c) MCOTEA. Ground Combat Element Integrated Task Force Experimental Assessment Plan. May 2014

Annex A. Infantry Rifleman (MOS: 0311)

Annex B. LAV Crewman (MOS: 0313)

Annex A.
Infantry Rifleman (MOS 0311)

This annex details the Infantry Rifleman (MOS 0311) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed 2 March–26 April 2015 at Range 107 and Range 110 aboard the Marine Corps Air-Ground Combat Center (MCAGCC), Twentynine Palms, CA. The sections outline the Infantry Rifleman Scheme of Maneuver (SOM), Limitations, Deviations, Data Set Description, Descriptive and Basic Inferential Statistics, and Modeling Results.

A.1 Scheme of Maneuver

A.1.1 Experimental Cycle Overview

The Infantry Rifleman (MOS 0311) assessment of the GCEITF took place in a field environment aboard MCAGCC, Twentynine Palms, CA. The assessment consisted of 21 trial cycles, each of which was a 2-day test cycle, conducted over the course of 55 days. After every 4 days of trials, the Marines received 1 recovery day spent at Camp Wilson. Each squad consisted of 12 volunteers and a direct assignment (nonvolunteer) squad leader. Each member of the squad was trained to fill each billet within the fireteam: fireteam leader, grenadier, automatic rifleman, and rifleman. The rifle squad contained three fireteams. The assessment was executed under the supervision of MCOTEA functional test managers and a range officer-in-charge (OIC)/range safety officer (RSO) from the GCEITF.

A.1.2 Experimental Details

The 2-day 0311 assessment was modeled to replicate offensive and defensive tasks. The 0311s began each cycle on the Day 1/Offensive task. Two 0311 squads executed each trial cycle: a control (C) nonintegrated squad, and a high-density (HD) integrated squad with 2 females.

Day 1 of the trial cycle was executed on Range 107 and consisted of a squad-reinforced attack. Each squad moved approximately 1 km to an assault position wearing an assault load and carrying personal weapons. As a squad, they moved all personnel and gear over an 8-ft obstacle/wall. After negotiating the obstacle, the squad staged their assault packs, deployed on-line, and conducted a 425-m live-fire assault to a limit of advance (LOA) wearing the fighting load. They immediately repelled an enemy counterattack by fire for 90 seconds. Finally, each fireteam conducted a 100-m casualty evacuation (CASEVAC) of a 220-lb dummy.

Day 2 of the experimental cycle was executed on Range 110 and consisted of defensive actions. The day started with the squads making a 7-km march from Range

107 to Range 110 wearing an approach load and carrying personal weapons. After a 10-minute operational pause, each fireteam spent 2 hours digging 2-man fighting positions wearing a fighting load. Two Marines from each fireteam provided security while the other two dug, rotating every 15 minutes. At the conclusion of Day 2, the composition of the squad was dissolved and a new sample of males and females were randomly assigned for the following cycle.

A.1.3 Additional Context

Throughout the duration of the assessment, Marines bivouacked at Range 107, sleeping in 2-man tents. Prior to the Day 1 trial, each Marine zeroed his or her weapon system to maximize accuracy. During trial execution, Marines wore/carried prescribed loads for each task. Packs were weighed each day prior to the 7-km march to ensure consistency. After each trial day, the Marines operated under the guidance of their company leadership, performing minimal physically demanding tasks. Not all Marines were selected for each trial cycle. Those Marines who were not part of an assessed squad conducted the same experimental subtasks to ensure equity in physical loading between individuals participating in a trial cycle and those not chosen via random selection for that particular cycle. (A detailed discussion of these loading tasks are in the loading section below.)

Fatigue surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. It allows for outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) to be accounted for in the analysis of the performance data. Workload surveys collected the volunteers' perceived average and maximal level of exertion during the specified performance tasks. Cohesion surveys provided a method of collecting subjective data relating to each squad's ability to work as a team and its members' overall perspective on the cohesiveness of the squad.

A.1.4 Experimental Tasks

A.1.4.1 1-km Movement

Assaulting an enemy position never starts from a static position: First, a movement must be conducted to the assault position (AP). The distance from the line of departure to the AP is dependent upon myriad factors. Based on time and space constraints during the GCEITF assessment, this distance was a little less than 1 km. Each rifle squad moved this distance as quickly as possible while carrying an assault load and individual weapon. The Infantry Training and Readiness (T&R) Manual states that "the assault load is the load that is needed during the actual conduct of the assault." This task put the Marines under moderate fatigue prior to commencing the attack.

A.1.4.2 Negotiate an Obstacle

The conduct of an attack often involves reducing or negotiating an obstacle. It is common in an urban environment to make entry through a window or over a wall (obstacle). One of the more difficult tasks is climbing over a wall with a fighting load. Each squad negotiated an 8-ft wall by getting all Marines and equipment over as quickly as possible. Although the technique was not dictated, the 0311 squads used a single launch-point to lift all Marines up onto the wall.

A.1.4.3 Squad Attack and Movement to LOA

The mission of the rifle squad is "to locate, close with, and destroy the enemy by fire and maneuver, and repel the enemy's assault by fire and close combat." Conducting a live fire-and-movement task is fundamental to the Infantry community and operationally relevant. Having moved to an AP, each rifle squad deployed into an on-line formation to maximize firepower in the direction of the enemy and then began movement-to-contact with weapons at the ready. Upon contact with the enemy (targets being presented), the Marines laid down a heavy volume of fire to gain fire superiority. They then began buddy rushes to close with the enemy objective. Each squad conducted approximately 300 meters of buddy rushes, engaging a total of 12 GCEITF targeting systems (GTS) with direct-fire weapons and three machinegun bunkers with the M16A4 and M203 grenade-launcher attachment. After all targets were destroyed (no targets remaining), the squad moved another 125 meters to an LOA and prepared for an enemy counterattack. This task is physically demanding and involves a combination of accuracy, tempo, and squad cohesion.

A.1.4.4 Repel Counterattack

At the conclusion of every assault, the attacking unit must be prepared for the enemy to regroup and organize a counterattack. Upon consolidation at the LOA, the squad oriented its weapons in the direction of the enemy's retreat. Upon target presentation, the counterattack commenced for 90 seconds. Six GTSs were used during the course of the counterattack. The primary purpose of this task was to determine the squad's accuracy engaging targets approximately 300 meters away after conducting a squad attack.

A.1.4.5 CASEVAC

Casualties are an inevitable part of conducting combat operations. Units train to become proficient in triaging, handling, and transporting a casualty. After a Marine is injured, it is essential to move the casualty to the appropriate level of care as quickly as possible. During the GCEITF assessment, each fireteam was assigned a casualty at the conclusion of the live-fire attack and counterattack. Each fireteam had to move a 220-lb casualty (dummy and equipment) a distance of 100 meters to a casualty collection point (CCP). They could use a variety of techniques for transport, but had to carry the dummy off the ground (not drag any part). The primary purpose of this task was to determine the fireteam's proficiency in moving a simulated casualty to a CCP. After the CASEVAC, each Marine took a fatigue and workload survey to assess overall fatigue and workload of the entire offensive task (see GCEITF Experimental Assessment Plan, Annex D).

A.1.4.6 7-km Hike

Moving under a load is one of the most fundamental tasks of an Infantry unit; it is both physically and mentally demanding. Units train by conducting tactical marches with an approach load at increasing distances. The Infantry T&R Manual states that "the approach march load will be such that the average Infantry Marine will be able to conduct a 20 mile hike in 8-hours with the reasonable expectation of maintaining 90% combat effectiveness." During the GCEITF assessment, each rifle squad moved a distance of 7.20 km from Range 107 to Range 110. This route was flat (minimal elevation change) and conducted on an unimproved surface with varying degrees of conditions (compact dirt and loose sand). The squads moved as fast as the slowest person, carrying an approach load and an individual weapon. The individual weapons included the M-4, M-16A4 with M203 attachment, and the M27, resulting in a cumulative load of 96–101lbs per Marine. The primary purpose of this task was to determine the squad's pace over a 7.20-km route while carrying the approach load. Each Marine took a fatigue and workload survey after completion of the 7.20 km hike (see GCEITF EAP, Annex D).

A.1.4.7 Digging Fighting Holes

Protection is vital to Infantry units, especially in a defensive posture. Infantry units commonly construct fighting positions to conceal positions and minimize exposure to enemy fire. There are many characteristics to constructing a doctrinal fighting position, but the most physically demanding aspect is digging out the fighting hole. The terrain at Twentynine Palms consists of hard, compact sand and rocks. Each fireteam dug 2 2-man fighting holes in a time limit of 2 hours, maintaining 50% security, meaning 2 Marines dug while 2 provided security in the prone position. The Marines swapped positions every 15 minutes. The purpose of this task was to determine the fireteam's rate of work while digging fighting positions. After 2 hours, each Marine took a fatigue and workload survey (see GCEITF EAP, Annex D).

At the completion of the 2-day cycle, Marines also took a cohesion survey to record cohesion during the execution of the 2-day trial cycle (see GCEITF EAP, Annex M).

A.1.5 Loading Plan

Due to the number of volunteers, a handful of Marines were not part of an assessed squad for each 2-day cycle. The primary purpose of the loading plan was to ensure, to the greatest extent possible, equity of physical activity among all volunteers throughout the duration of the experimental assessment. Collaboration with the company leadership determined that the best method of loading nonassessed Marines was to have them perform the same tasks as an assessed squad to experience the same conditions and physical strain. Every trial and task was conducted in the same manner and sequence to ensure consistency.

A.1.5.1 Variations

Several factors, including safety and the number of loading Marines, introduced variations to the loading plan.

- In some instances, the loading Marines formed a quasi-squad and conducted the trial after all assessed squads were done for the day. At other times, the quasi-squad was too small, so the loading Marines were attached to an assessed squad, in which case they operated on the flanks and were not given ammunition. At no point in time did a loading Marine aid or interfere with an assessed Marine/squad.
- The loading Marines did not always dig for 2 hours. To gain efficiencies, the loading Marines would rotate between digging, assisting in weighing buckets of sand, and refilling the holes of the assessed squads. Any such variation was carefully calculated to ensure loading Marines were doing an equivalent amount of work as the assessed Marines.

A.1.6 Scheme of Maneuver Summary

The 0311 experiment consisted of a 2-day trial cycle comprising an offensive and defensive day. The offensive day involved five subtasks based around a squad live-fire attack: 1-km movement, negotiating an obstacle, fire and movement, counterattack, and CASEVAC. The defense day involved two subtasks: a 7-km march and digging a fighting position. During the course of the experiment, the 0311/Provisional Infantry (PI) squad executed 2 pilot trial cycles and 21 record trial cycles. During trial execution, Marines rotated through every billet within the rifle squad, carrying the respective weapon system.

A.2 Limitations

A.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a similar manner to those in an operational environment. However, under certain situations, artificial limitations or interruptions were introduced that changed or altered the way a task would normally be performed. While these limitations represent a degree of artificiality, they do not detract significantly from our ability to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for the 0311 assessment.

A.2.2 Relative Difficulty of Record Test

The Infantry GCEITF assessment was designed to gather data associated with some of the most physically demanding tasks within the 0311 MOS. However, these tasks in isolation did not fully replicate life experienced by a Marine during a typical field exercise (FEX) or a combat environment and did not obtain the maximum cumulative load that could be placed on an Infantry Marine. With limited time available, only selected 0311 tasks were assessed. Other tasks/duties outside of the assessment were minimized due to specific experimental constraints and human factors. During a typical FEX, it is common for Marines to conduct 24-hour operations that include day and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. It took 0311 squads approximately 1 hour to complete the offensive day and approximately 4 hours to complete the defensive day. Outside the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

Another concern in designing the 0311 assessment was to ensure that it was achievable and sustainable for a 60-day period. The 7-km march distance was selected based on the training time available prior to the assessment. However, many of the loads carried were decreased prior to the commencement of the 21 cycles used to form the dataset. Once over the wall obstacle, the 0311s staged their assault packs without security rather than wearing them during the fire-and-movement. The 0311s also did not carry additional ammunition, radios, batteries, or other equipment often required when operating in a tactical environment. The Marines were authorized 1 day off after every 4 days of training; this artificial recovery period is not achievable when conducting training or combat operations.

A final factor affecting the relative difficulty of the record test pertains to the intangible physiological impact of the volunteers' ability to drop on request (DOR) at any point during a trial. Any time a volunteer dropped during a trial cycle, that squad/team performed the following subtasks with fewer personnel. This factor affected the cohesion of each squad and influenced its performance.

A.2.3 Digging a Fighting Position

Several artificialities were present as the volunteers dug their fighting holes. Preparing a defensive fighting position involves many continuing actions by those not actually digging the hole, such as clearing fields of fire, creating sectors of fire, etc. However, no continuing defensive actions were conducted other than lying in the prone position and providing notional security. The Marines digging would not normally fill up a 5-gallon bucket for measurement; under normal conditions they would either fill sandbags or build up a parapet. These artificialities may have influenced their performance and morale.

A.2.4 Number of Volunteer Participants

For the 0311 experiment, 36 male and 9 female volunteers began the experiment, but by the end 28 males and 6 females completed the assessment. The results presented in this annex are based on the performance of 34–45 Marines.

A.2.5 Limitations Summary

The 0311 assessment was designed to replicate realistic training in a field environment. The end-state was to create an experiment in which the volunteers felt they were conducting realistic and operationally relevant tasks, but unavoidable limitations to the assessed tasks and non-assessed operating environment introduced a level of artificiality not normally present in a field training or combat environment.

A.3 Deviations

A.3.1 Concentration for Rifle Squad

Marines participated in this experiment voluntarily and could leave the experiment at any time, for any reason, making the total number of subjects available highly volatile. Trials were not executed if there were not enough participants to carry out the assigned mission and if attrition affected the integration levels for the 0311 experiment. The EAP stated that the Rifle Squad would have three levels of integration: C group, low-density (LD) group, and high-density (HD) group, with zero females, one female, and two females, respectively. Prior to the beginning of the first record trial cycle, the population could no longer support an LD group. The MCOTEA Research Team eliminated the LD group and began the Record Trials on 7 March 2015 with the control and HD groups only.

A.4 Data Set Description

A.4.1 Data Set Overview

The 0311 portion of the experiment consisted of 2 pilot trial cycles and 21 record trial cycles. The pilot trial cycles were conducted from 2 March to 6 March 2015. Pilot trial cycle data are not used in analysis due to variations in the conduct of the experiment. We based all analysis on the 21 record trial cycles executed from 7 March to 26 April 2015.

A.4.2 Record Test Volunteer Participants

At the beginning of the first record trial cycle, there were 36 male 0311 volunteers and 9 female volunteers. There were several Marines who voluntarily withdrew, or were involuntarily withdrawn, during the execution of the experiment. The final number of volunteers was 28 males and 6 females.

A.4.3 Planned, Executed, and Analyzed Trial Cycles

Table A-1 displays number of trial cycles planned, executed, and analyzed by task. The planned number of trials for the 0311 MOS per section 7.5.3 of GCEITF EAP was 120 trial cycles, or 40 trial cycles per planned integration level (C, LD, and HD). The original plan called for 6 squads per day (2 per integration level) over the 20 trial cycles. However, due to the number of Marines who voluntarily withdrew or were involuntarily withdrawn from the experiment prior to the execution of the first record trial cycle, only one squad of the C and HD integration levels remained. We chose to keep the HD integration level to employ more Marines in the experiment. The planned number of trial cycles in Table A-1 reflects 21 planned trial cycles for each integration level.

There are several occurrences of missing data for the 7-km hike by individual kilometer. The individual kilometer times were derived from GPS data. Early in the experiment, the Garmin GPS's were set to record a volunteer's position every second. Due to the storage space on the GPS and length of the trial, when volunteers executed the 7-km hike and then 2 hours of preparing a fighting position, the GPS could not hold all the data. Therefore, it overwrote the hike data. Once the problem was found, the GPSs were corrected to record location every 2 seconds.

Table A-1. 0311 Planned, Executed, and Analyzed Trial Cycles

Task and Metric Description	Integration Level	Number of Planned Trials	Number of trials conducted	Number of trials used in analysis	Notes
7-km Hike	C	21	21	20	Remove Mar 28 due to TIR[a]
	HD	21	21	21	
Movement to LOA	C	21	20	19	Did not execute Mar 19 due to targets catching on fire. Mar 17 outlier.
	HD	21	21	20	Mar 17 outlier
Attack & C-Atk Percent Hits	C	21	20	20	Did not execute Mar 19 due to targets catching on fire
	HD	21	21	21	
CASEVAC by FT	C	84	82	82	Did not execute Mar 19 due to targets catching on fire
	HD	42	41	41	
Prepare Fighting Position by FT	C	84	84	84	
	HD	42	41	41	Mar 10 FT 3 attrited
1-km Hike	C	21	20	20	Did not execute Mar 19 due to targets catching on fire
	HD	21	21	20	Mar 17 outlier
Negotiate Obstacle	C	21	20	19	Did not execute Mar 19 due to targets catching on fire. Mar 7 outlier.
	HD	21	21	21	
7-km Hike; 1km Time	C	21	21	17	No data: Mar 8, Mar 13, Mar 15, Mar 18.
	HD	21	21	18	No data: Mar 8, Mar 15, Mar 18.
7-km Hike; 2km Time	C	21	21	16	No data: Mar 8, Mar 10, Mar 13, Mar 15, Mar 18.
	HD	21	21	18	No data: Mar 8, Mar 15, Mar 18.
7-km Hike; 3km Time	C	21	21	16	No data: Mar 8, Mar 10, Mar 13, Mar 15, Mar 18.
	HD	21	21	19	No data: Mar 8, Mar 18
7-km Hike; 4km Time	C	21	21	16	No data: Mar 8, Mar 10, Mar 13, Mar 15, Mar 18.
	HD	21	21	19	No data: Mar 8, Mar 18
7-km Hike; 5km Time	C	21	21	16	No data: Mar 8, Mar 10, Mar 13, Mar 15, Mar 18.
	HD	21	21	19	No data: Mar 8, Mar 18
7-km Hike; 6km Time	C	21	21	16	No data: Mar 8, Mar 10, Mar 13, Mar 15, Mar 18.
	HD	21	21	19	No data: Mar 8, Mar 18
7-km Hike; 7km Time	C	21	21	16	No data: Mar 8, Mar 10, Mar 13, Mar 15, Mar 18.
	HD	21	21	19	No data: Mar 8, Mar 18

a. A TIR in this table refers to a Test Incident Report, which is a report the test team or direct assignment leaders completed when an incident occurred that affected the natural execution of a trial. If a data point is removed due to a TIR, it is because the TIR affected the data in such a way that it is not comparable to the rest of the dataset.

A.5 Descriptive and Basic Inferential Statistics

A.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common Infantry tasks and are indicative of unit-level proficiency during either field exercises or combat operations.

This section presents results for 7 out of 16 tasks. The 0311 Appendix contains the descriptive statistics for the remainder of the 0311 tasks. The words “metric” and “task” are used interchangeably throughout this Annex; both refer to the experimental task.

Each fireteam consisted of four volunteer Marines: the fireteam leader, automatic rifleman, grenadier, and rifleman. Each squad consisted of 12 volunteer Marines (three fireteams) with a direct assignment (nonvolunteer) squad leader. There were two integration levels for all tasks. For squad-level tasks, a C group was non–gender integrated, and a HD group was gender integrated with two female Marines. For fireteam-level tasks, a C group was non-gender integrated and an HD group contained at least one, but not more than two, females.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA), and scatterplots. The subsequent sections will cover each task in detail. Finally, contextual comments, additional insights, and subjective comments (as applicable) tying back to each experimental task are incorporated.

Use caution when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks such as distances, techniques, leadership, load carried, group size, and group composition.

A.5.2 0311 Selected Tasks Descriptive Statistics Results

The two tables below display the results for the 7 selected 0311 metrics. Table A-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations. Table A-3 displays ANOVA results, including metrics and integration levels, p-values suggesting statistical significance, integration level elapsed-time differences, and percentage differences between integration levels. For each task, an ANOVA and t-test were conducted to compare the two groups. If p-values are less than the a priori–determined significance level of 0.10, we conclude that there is statistical evidence that the response for the HD group is different from that in the C group.

Table A-2. 0311 Selected Task Results (Descriptive Statistics)

Metric Description	Integration Level	Sample Size	Mean	SD
7-km Hike (minutes)[a]	C	20	79.76	3.12
	HD	21	82.40	3.90
Movement to LOA (minutes)[a]	C	19	2.65	0.21
	HD	20	2.89	0.25
Attack & C-Atk Percent Hits (%)[a]	C	20	0.40	0.05
	HD	21	0.36	0.07
CASEVAC by FT	C	82	1.46	0.25

Metric Description	Integration Level	Sample Size	Mean	SD
(minutes)[a]	HD	41	1.57	0.47
Prepare Fighting Positions by FT (lbs.)	C	84	4284.86	1005.43
	HD	40	4041.81	1017.44
1-km Hike (minutes)[a]	C	20	8.07	0.53
	HD	20	8.49	0.39
Negotiate Obstacle (minutes)	C	19	2.77	0.34
	HD	21	2.92	0.48

[a] Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA.

Table A-3. 0311 Selected Task ANOVA and Welch's T-test Results

Metric	F statistic (df)	F Test P-value	Comparison	Difference	% Difference	2-sided P-Value	1-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
7-km Hike (minutes)*	5.73 (1, 39)	0.02[a]	HD-C	2.64	3.32%	0.02[a]	0.01[a]	1.21	4.08	0.79	4.50
Movement to LOA (minutes)*	10.76 (1, 37)	< 0.01[a]	HD-C	0.24	9.06%	< 0.01[a]	< 0.01[a]	0.15	0.34	0.12	0.36
Attack & C-Atk (% hit)*	4.76 (1, 39)	0.04[a]	HD-C	-0.04	-10.40%	0.03[a]	0.02[a]	-0.07	-0.02	-0.07	-0.01
CASEVAC (minutes)*	2.96 (1, 121)	0.09[a]	HD-C	0.11	7.63%	0.16	0.08[a]	0.01	0.21	-0.02	0.24
Prepare Fighting Positions (lbs)	1.57 (1, 122)	0.21	HD-C	-243.05	-5.67%	0.21	0.11	-494.78	8.67	-567.28	81.18
1-km Hike (minutes)*	8.04 (1, 38)	0.01[a]	HD-C	0.42	5.18%	0.01[a]	< 0.01[a]	0.23	0.61	0.17	0.67
Negotiate Obstacle (minutes)	1.24 (1, 38)	0.27	HD-C	0.15	5.33%	0.27	0.14	-0.02	0.32	-0.07	0.37

[a] Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA.

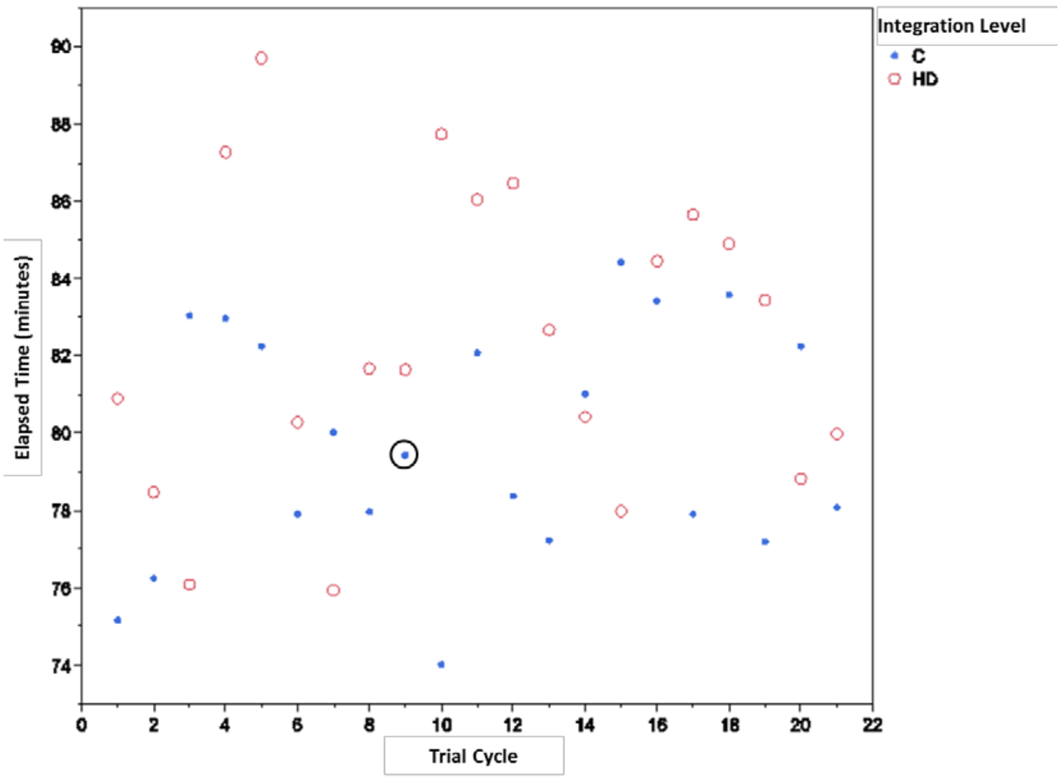
A.5.2.1 7-km Hike Results

A.5.2.1.1 7-km Hike Overview

This experimental task assessed a squad of 12 Marines moving 7.20 km while each carried an approach load and an individual weapon. The individual weapons included the M-4, M-16A4 with M203 attachment, and the M27, resulting in a cumulative load of 96–101 lbs per Marine The recorded time for this task started when the squad departed the Range 107 start point and stopped when the squad arrived at the Range 110 stop point. Each squad moved as fast as the slowest person and could take as many breaks as necessary.

Figure A-1 displays all 0311 Infantry 7-km hike time data. On trial cycle 9, the C data point was removed from the analysis due to a TIR; this data point is circled in black in the figure. With the exception of this data point, all data on the scatterplot are valid for analysis.

Figure A-1. 7-km Hike



The data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.17 for the C group and 0.81 for the HD group.

The C group had a mean time of 79.76 minutes. This time is statistically significantly faster than the HD mean time of 82.40 minutes. This difference results in a 3.32%, or 2.64-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 0.78-minute increase in standard deviation (SD) (3.12 minutes for the C group and 3.90 minutes for the HD group). (See Tables A-2 and A-3 for detailed analytical results.)

A.5.2.1.2 7-km Hike Contextual Comments

A.5.2.1.2.1 USMC Hike Standards

The 7-km hike is a task that the following MOSs completed during the experiment: 0311, 0331, 0341, 0351 and 0352, Provisional Infantry (PI), Provisional Machine Gunners, and Combat Engineers. There are varying standards to which we can compare this result. The following sections define those standards, as well as the one we chose as a comparison.

The Infantry T&R Manual (30 Aug 2013) identifies minimum standards that Marines must be able to perform in combat, including standards for tactical marches. In Chapter 8 of the Infantry T&R Manual in the MOS 0300 Individual Events section, task 0300-COND-1001 (March under an approach load) is applicable to all 03XX MOSs, ranks PVT–LtCol. The condition and standard established by this task is: "Given an assignment as a member of a squad, individual weapon, and an approach load, complete a 20-kilometer march in under 5 hours." The march pace required by this standard is 4 kilometers per hour (km/h). In Chapter 9 of the Infantry T&R Manual in the MOS 0302/0369 Individual Events section, tasks 0302-OPS-2001 (Lead an approach march) and 0369-OPS-2501 (Lead an approach march) are applicable to MOS 0302 and 0369, ranks SSgt–MGySgt and 2ndLt–LtCol. The condition and standard established by these tasks is: "Given a mission, time constraints, an approach march load, organic weapons, and a route, move 24.8 miles (40 km) in a time limit of 8 hours." The march pace required by this standard is 5 km/h. Appendix E of the Infantry T&R Manual (Load Terms and Definitions) states: "The approach march load will be such that the average Infantry Marine will be able to conduct a 20 mile hike in 8-hours with the reasonable expectation of maintaining 90% combat effectiveness." The march pace required by this definition is 4.02 km/h.

Chapter 3 of Fleet Marine Force Reference Publication (FMFRP) 3-02A: *Marine Physical Readiness Training for Combat* (16 Jun 2004) states: "The normal pace is 30 inches. A pace of 30 inches and a cadence of 106 steps per minute result in a speed of 4.8 km/h or 3 mi/h and a rate of 4 km/h or 2.5 mi/h if a 10-minute rest halt per hour is taken." Common Infantry practice today is to hike for 50 minutes and take a 10-minute break, while maintaining an overall pace of 4 km/h (resulting in a hiking pace of 4.8 km/h).

Driven by the need to pick an evaluative reference standard, this report follows the T&R Manual's intent to establish minimum standards and uses the 4 km/h march pace for a 20 km march established by task 0300-COND-1001 and supported by the definition of an approach load. Further, while an established reference standard is required to anchor observed performance to the T&R program, more important information is provided by performance differences observed between gender-integrated and non–gender integrated units.

A.5.2.1.2.2 0311 7-km Hike Pace

The difference in 7-km hike times is relevant to both the training and combat environment as it will take integrated squads more time to conduct foot marches. Per the tactical march standards noted above, the Marine Corps standard of hiking is 4.0 km/h. Both the C and HD groups surpassed this standard. The average C group's

7.20-km pace was 5.42 km/h and the average HD group's pace was 5.24 km/h, meaning the HD group was 0.17 km/h slower. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), the HD group would finish 7.33 minutes behind the C group.

The 0311 Appendix contains the descriptive statistics for each individual km of the 7-km hike, and these data show that the C and HD groups had an increasing trend in 1-km split-times over the course of the 7.20-km march. Both groups' march pace decreased over the distance marched. The C group's speed decreased at a trend rate of 0.20 km/h per km, while the HD group's speed decreased at 0.24 km/h per km. Assuming this speed decrement continued for every kilometer in a 20-km march (a pessimistic assumption that ignores attainment of a sustainable pace), then the HD group would have been 57 minutes behind the C group at the finish of a 20-km march. More realistically—and allowing for the attainment of a sustainable pace—if we assumed each group followed its speed decrement trend down to the first kilometer, where its speed was just under the 4 km/h standard and then finished the march at that predicted pace, then the HD group would have been 14.4 minutes behind the C group at the finish of a 20-km march. On any given day (under the same environmental conditions), the C group was faster than the HD group 76.2% of the time (16 of 21 trial cycles).

A.5.2.1.3 7-km Hike Additional Insights

While the difference in the 7-km hike time and speed is statistically significant, the size of the difference is small. A purely objective evaluation of 0.17 km/h slower hike speed is elusive, but Marine Corps Doctrinal Publications (MCDP) consistently emphasize the importance of speed. MCDP 1-3: *Tactics* devotes an entire chapter to "Being Faster," stating: "Physical speed, moving more miles per hour, is a powerful weapon in itself." MCDP-6: *Command and Control* also speaks to speed relative to the enemy, stating: "The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results." Nonetheless, the difference observed between the C and HD groups in the 0311 Infantry 7-km hike task was a nominal 2.64 minutes. For an indication of performance when the load is increased, see the 0331, 0341, and 035X 7-km hike data and contextual comments.

A.5.2.1.4 7-km Hike Subjective Comments

For subjective comments relating to this task, see the 0311 Appendix. The subjective comments reveal a fairly equal spread of males and females who were falling behind and slowing the overall movement.

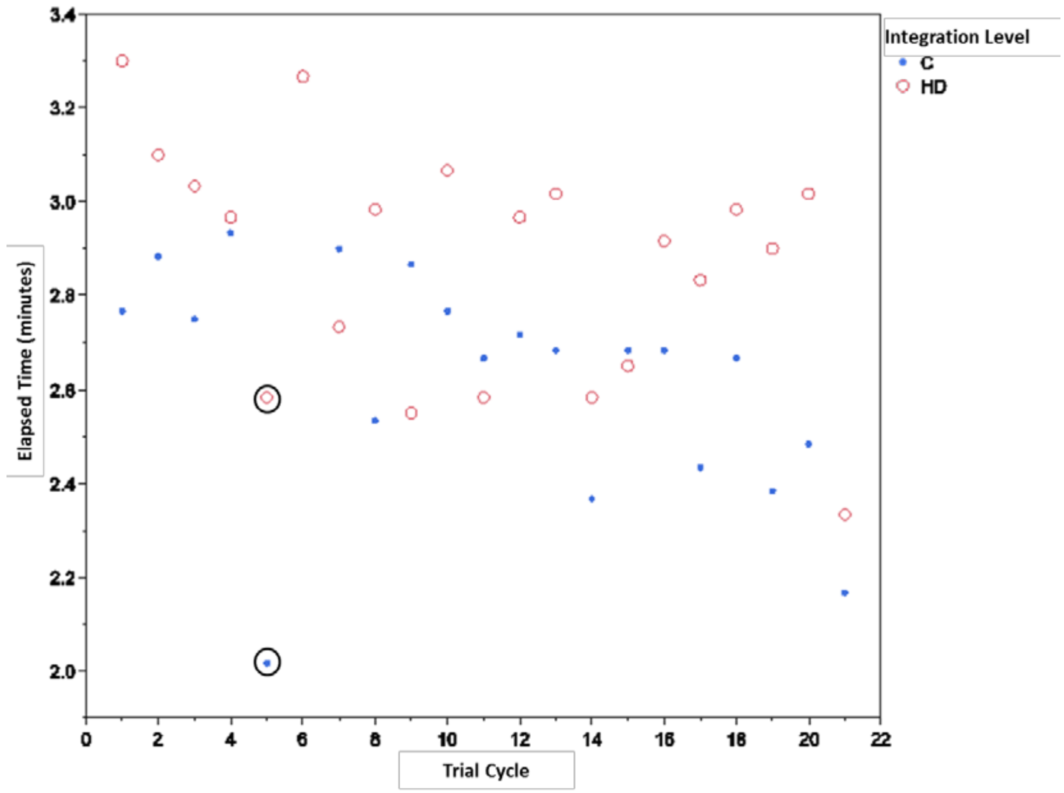
A.5.2.2 Movement to LOA

A.5.2.2.1 Movement to LOA Overview

This experimental task assessed a squad moving approximately 125 meters to an LOA immediately after conducting a fire-and-movement assault. The recorded time started when the last target went down after the assault and stopped when the last member of the squad was prone at the LOA.

Figure A-2 is a scatterplot showing all 0311 Movement to LOA time data. On Trial Cycle 5, the C and HD group data points were removed from the analysis due to data collection errors; data omitted from the analysis are circled in black. With the exception of the Trial Cycle 5 data points, all data on the scatterplot is valid for analysis.

Figure A-2. Movement to LOA



The data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.22 for the C group and 0.31 for the HD group.

The C group has a mean of 2.65 minutes. This time is statistically significantly faster than the HD mean time of 2.89 minutes. This difference results in a 9.06%, or 0.24-minute, degradation in movement time between the groups. The HD group has greater variability, as shown by the 0.04-minute increase in SD (0.21 minutes for the C group and 0.25 minutes for the HD group). See Table A-2 and Table A-3 for detailed analytical results.

A.5.2.2.2 Movement to LOA Contextual Comments

The ability to close with the objective after conducting a live-fire attack is a crucial aspect to maintaining momentum during offensive operations. In this case, the integrated squads were exposed to the enemy 14.4 seconds (9.06%) longer and had that much less time to prepare for a counterattack or conduct follow-on actions. On any given day (under the same environmental conditions), the C group was faster than the HD group 80% of the time (16 of 20 trial cycles).

A.5.2.2.3 Movement to LOA Additional Insights

A purely objective evaluation of 14.4 seconds is elusive, but it may possess some practical significance on the battlefield that would reduce the survivability of an integrated squad. Considering the 12–15 rounds-per-minute sustained rate of fire for the M4 and AK-47 rifles, a single enemy fighter would have the opportunity to take two to three more well-aimed shots on Marines in an integrated squad than they would on a non-integrated squad moving to the LOA. Similarly, the delay degrades Marines' pursuit of the enemy by fire, denying the slowest Marines 2 to 3 well-aimed engagements on the enemy from a fixed position at the LOA. The resulting trade in casualty exchange could be significant.

A.5.2.2.4 Movement to LOA Subjective Comments

For subjective comments relating to this task, see the 0311 Appendix.

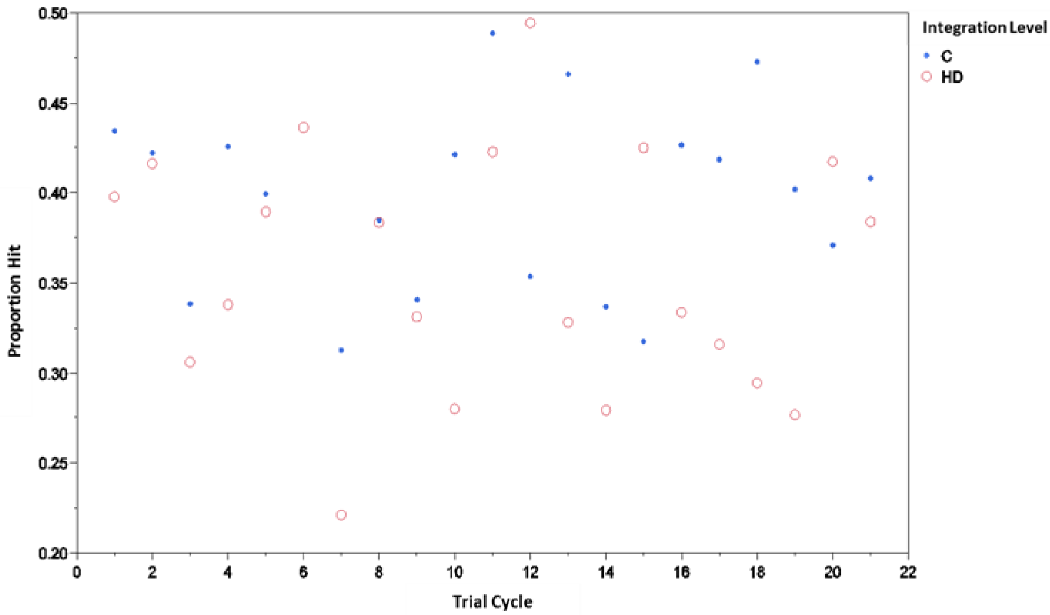
A.5.2.3 Attack and Counterattack (C-Atk) Percent Hits Results

A.5.2.3.1 Attack and C-Atk Percent Hits Overview

This experimental task assessed the accuracy of a rifle squad engaging 12 GTSs during a 300-meter fire-and-movement assault. Each GTS captured the precise location of a round that impacted a target silhouette using a location-of-hit-and-miss (LOMAH) sensor. The GTSs were equally spread over the downrange area at varying distances and were exposed according to predetermined parameters consistent for every squad attack. The accuracy was determined by dividing the number of hits on each target by the total ammunition expended by each squad.

Figure A-3 displays all Attack & C-Atk Percent Hit data. All data on the scatterplot are valid for analysis.

Figure A-3. Attack and C-Atk Percent Hits



The data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.47 for the C group and 0.73 for the HD group.

The C group had an average percent hit of 40%. This proportion is statistically significantly higher than the HD average of 36%. This difference results in a 10.40%, or 4-percentage point, degradation in accuracy. In addition, the HD group has greater variability, as shown by the 2-percentage point increase in SD (5% for the C group and 7% for the HD group). See Table A-2 and Table A-3 for detailed analytical results.

A.5.2.3.2 Attack and C-Atk Percent Hits Contextual Comments

A.5.2.3.2.1 Attack and C-Atk Percent Hits Contextual Comments Overview

In combat operations, accuracy is highly desirable when destroying or effectively suppressing an enemy position. In the execution of this task, the HD group sustained an average 10.4% decrement in the percentage of hits for the number of shots taken. Operationally, this decrease equates to a degradation of lethality and an increase in ammunition expenditure. On any given day (under the same environmental conditions), the C group was more accurate than the HD group 85% of the time (17 of 20 trial cycles).

A.5.2.3.2.2 Analysis by Gender and Weapon System

The use of a Weapons Player Pack (WPP) on each weapon system allowed data to be captured on each shot taken by a Marine during the conduct of the attack and C-atk. When synchronized with the data obtained from the GTS, a shooter-to-shot correlation was possible. For the analysis of shot accuracy, accuracy percentages by gender and weapon were analyzed that measured percent hits and percent hits and near misses, where a hit indicates that a round hit the target silhouette and a near miss indicates that the LOMAH sensor detected a round within a 1.5 meter detection arc.

Since the accuracy results by gender were collected, a t-test was used for this analysis. Table A-4 displays the shot accuracy results by gender and weapon system.

Table A-4. Shot Accuracy by Gender and Weapon System

	Probability of Hit				Probability of Hit & Near Miss			
Weapon	**F**	**M**	**% Difference**	**2-sided p-value**	**F**	**M**	**% Difference**	**2-sided p-value**
M4 0311[b]	.28	.42	33%	<0.01	.73	.75	3%	0.32
M27 0311[c]	.25	.43	42%	<0.01	.58	.69	16%	.0144
M16A4/M203 0311[c]	.15	.28	46%	<0.01	.50	.67	25%	<0.01
M4 PI[c]	.37	.44	16%	.0189	.73	.79	8%	0.0341
M27 PI	.37	.38	3%	.7571	.66	.69	4%	.3663
M16A4/M203 PI[c]	.15	.26	42%	<0.01	.59	.70	16%	0.0453

[a.] The M16A4/M203 shot accuracy data is only for 5.56mm ammunition shot from the weapon and does not include the 40mm practice round accuracy.

[b.] Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the percent hits values for the gender.

[c.] Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the percent hits and percent hits and near miss values for the gender.

The shot accuracy by gender and by weapon reveals that there is a statistically significant difference between genders for every weapon system within the 0311 squads, except for the probability of hit/near miss with the M4. The overall accuracy declined and the percent difference increased as the weight of the weapon system increased. The M4 was the lightest weapon and yielded the most accurate results, while the M16A4/M203 was the heaviest weapon and yielded the least accurate results.

One might think that the experience level of the Marines (male vs. female) influenced the results. However, the PI results disprove this conjecture, as the male and female PI Marines came to the unit with the same experience level and their performance mirrors the results of the 0311 Marines. With the exception of the M27 accuracy, which was not statistically significant, the PI accuracy and percent difference declined as the weapon got heavier. Furthermore, a close look at the unit-training plan, methodology, leadership, and ammunition expended shows that the Marines in the GCEITF were very well prepared for this assessment, minimizing the impact of past experience. One could conclude that given the same type and amount of training, female Marines will be less lethal on the battlefield than male Marines.

A.5.2.3.3 Attack and C-Atk Percent Hits Additional Insights

To explore the operational effect of a 4-percentage point difference in percentage of hits, a Monte Carlo simulation was conducted using the Lanchester Square Law as a model of tactical-level engagement. We chose the Lanchester Square Law for rudimentary analysis using a well-known combat model to explore the potential effects of the differences discovered between integration levels. (For more information on the Lanchester Square Law, see the Methodology Annex, Annex R.)

There are several important assumptions we make for this Lanchester model in terms of parameter selection. We chose parameters that model a worst-case scenario for a squad engagement in which the squad faces an equivalent adversary in terms of capability. The first assumption is that the force sizes are equal, which is a 12-on-12 fight. The second assumption is the relative rate of fire between forces; we assume this ratio of friendly-to-enemy rate of fire to be equal to 1. Finally, with respect to accuracy, the enemy was assumed to be as good as the best experimental group, which was the C group. Thus the enemy probability of hit was set at 0.40.

Two hundred thousand simulated engagements were run for the C and HD groups. A force won an engagement when its opposition was eliminated. By construct, the C group (with a probability of hit equal to the enemy's, at 0.40) won 50% of the engagements, with an average of three Marines remaining. This even-fight case is useful only to assess the effect of the HD group's decrement in probability of hit. In another two hundred thousand simulated engagements, the HD group (with a probability of hit equal to 0.36) won 46.9% of the engagements, with three Marines remaining. The difference in these results represents a 6.2%, or 3.1 percentage point, degradation in tactical-level engagement wins.

A.5.2.3.4 Attack and C-Atk Percent Hits Subjective Comments

For subjective comments relating to this task, see the 0311 Appendix.

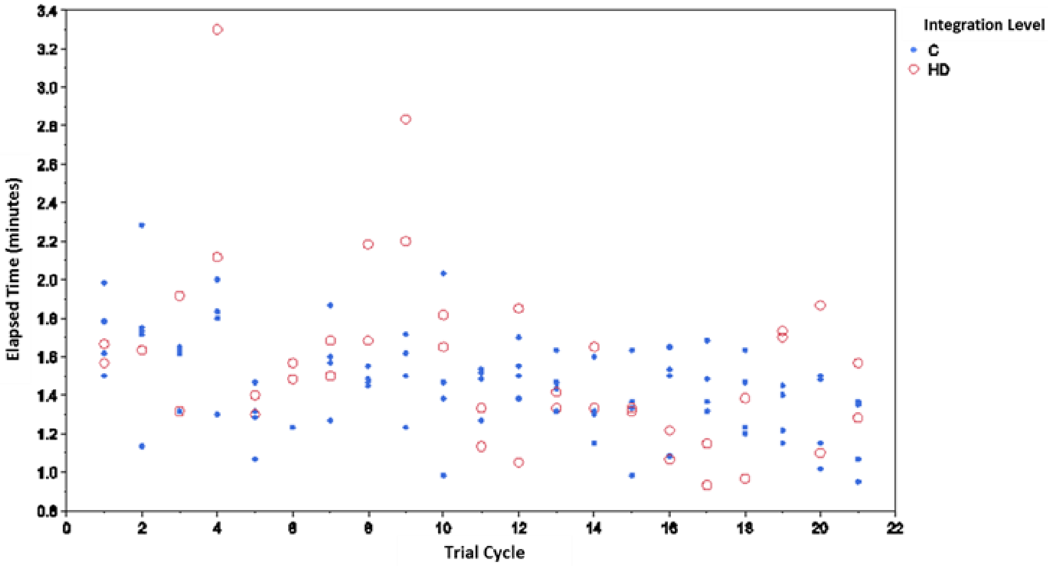
A.5.2.4 CASEVAC by FT Results

A.5.2.4.1 CASEVAC by FT Overview

This experimental task assessed a fireteam's ability to move a 220-lb dummy a distance of 100 meters to a CCP while wearing a fighting load and individual weapon. Marines conducted this task at the conclusion of the squad attack/counterattack. At the discretion of each fireteam, Marines used a 2-Marine, 3-Marine, or 4-Marine carry to move the casualty. The recorded time started when a member of the fireteam touched the dummy, and it stopped when the dummy and all members of the fireteam arrived at the CCP.

Figure A-4 displays all CASEVAC by FT data. All data on the scatterplot are valid for analysis.

Figure A-4. CASEVAC by FT



For this task, sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

The C group had a mean of 1.46 minutes. This time is statistically significantly faster than the HD mean time of 1.57 minutes. This difference results in a 7.63%, or 6.6-second, degradation in evacuation times between the groups. In addition, the HD group has greater variability, as shown by the 0.22 increase in SD (0.25 minutes for the C group and 0.47 minutes for the HD group). See Table A-2 and Table A-3 for detailed analytical results.

A.5.2.4.2 CASEVAC by FT Contextual Comments

The implications of this task contain relevance to the training and combat environments, as survival is dependent upon expeditious movement of the casualty to higher levels of care. Based on the standard deviations, the variation in performance of the HD group is nearly twice as much as the variation in performance of the C group. This inconsistency in the performance of the HD group leads to greater uncertainty and less confidence in its future performance from the mean.

A.5.2.4.3 CASEVAC by FT Additional Insights

While the "Golden Hour" is a common medical battlefield care construct for C2 and logistical support, medical literature supports a "Platinum Ten" construct of first response. Noting that the majority of casualties die within ten minutes of the trauma, the "Platinum Hour" construct holds that a patient needs to be correctly triaged and moved to medical care as fast as possible; any time degradation will reduce the probability of survival.

A.5.2.4.4 CASEVAC by FT Subjective Comments

For subjective comments relating to this task, see the 0311 Appendix.

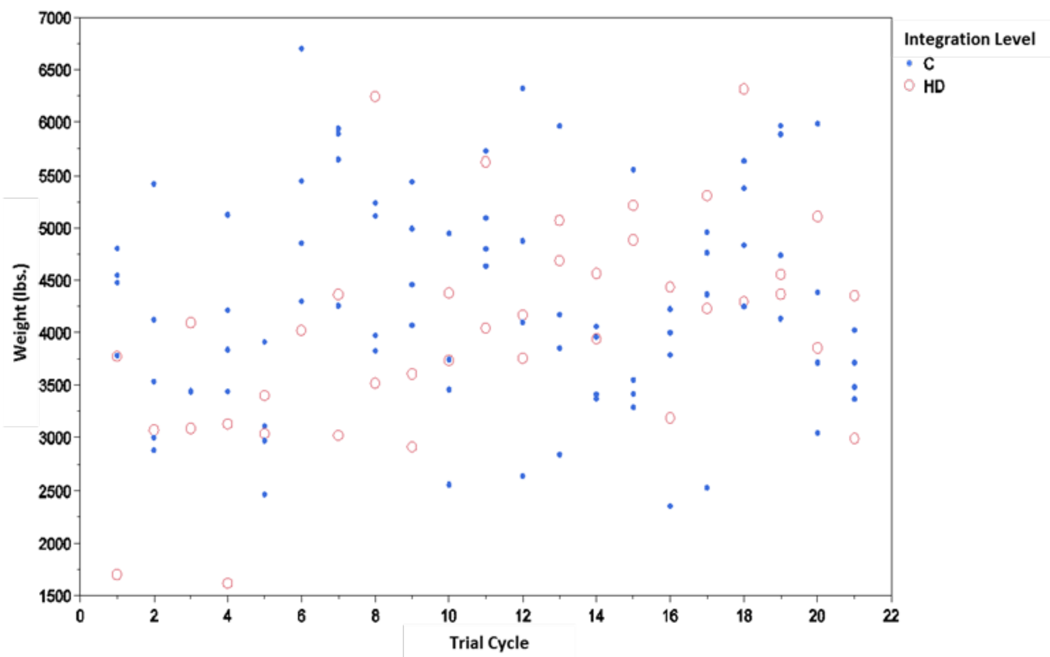
A.5.2.5 Prepare Fighting Positions by FT Results

A.5.2.5.1 Prepare Fighting Positions by FT Overview

This experimental task assessed the amount of earth moved by a fireteam in a 2-hr period. Each fireteam was required to dig 2 2-man fighting holes while wearing a fighting load and maintaining 50% security (only two Marines were digging at any given time). All the earth was scooped into buckets and weighed by a nonassessed Marine. The recorded time began 10 minutes after finishing the 7-km hike, and time stopped 2 hours after beginning to dig.

Figure A-5 displays all 0311 Prepare Fighting Positions data. All data on the scatterplot are valid for analysis.

Figure A-5. Prepare Fighting Positions by FT



For this task, sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

The C group moved a mean of 4,284.86 lb. of earth. This amount is greater (but not statistically significantly) than the HD group mean of 4,041.81 lb. There is a 5.67%, or 243.05 lb, degradation between groups. In addition, the HD group has slightly greater variability, as shown by the 12.01-lb increase in SD (1005.43 lb. for the C group and 1017.44 lb. for the HD group). See Table A-2 and Table A-3 for detailed analytical results.

A.5.2.5.2 Prepare Fighting Positions by FT Contextual Comments

Marine Corps Warfighting Publication (MCWP) 3-11.2: *Marine Rifle Squad* specifies the dimensions of a 2-person fighting hole: 6-ft long x 2-ft wide x 5-ft deep (assuming occupants of height 5 ft. 6 in.–6 ft). The volume of this fighting hole is 60 cu ft. Two such holes are required for a fireteam, with a combined volume of 120 cu ft. In Engineering Manual 1110-1-1905: *Engineering and Design: Bearing Capacity of Soils*, the Army Corps of Engineers estimates that collapsible soil (of the type found at MCAGCC 29 Palms) has an average density of 85 lb/cu ft. Thus a total of 10,200 lbs of dirt would have to be moved to construct two fighting positions.

The C group moved earth at a rate of 2,142.43 lb./h. At this rate, the C group would have completed digging their fighting holes in 4 hr, 46 min. The HD group moved earth at a rate of 2,020.91 lb/h. At this rate, the HD group would have completed digging their fighting holes in 5 hr, 3 min. Thus, the HD group would have completed digging their fighting holes 17 min later than the C group.

A.5.2.5.3 Prepare Fighting Positions by FT Additional Insights

Motivation and perspective provided by the platoon leadership was a contributing factor to the overall morale and work ethic of each Marine. Leadership and inspiration were not explicitly measured in this tactical task, but are thought to have had a larger impact than anticipated. The 0311 Platoon leadership was regularly observed sitting as a group approximately 50 meters away from their Marines with little interaction during the assessment. While the platoon leadership fulfilled their duties during the assessment, they did not attempt to maximize the output of their Marines.

A.5.2.5.4 Prepare Fighting Positions by FT Subjective Comments

For subjective comments relating to this task, see the 0311 Appendix.

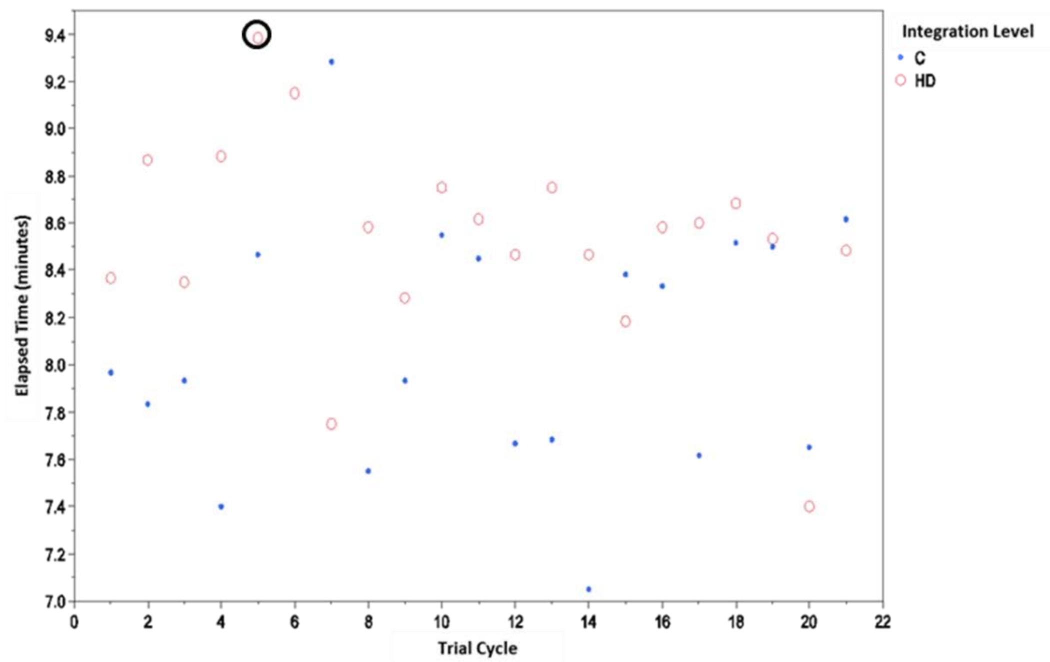
A.5.2.6 1-km Hike Results

A.5.2.6.1 1-km Hike Overview

This experimental task assessed a squad of 12 Marines moving approximately 1 km while each Marine carried an assault load and individual weapon. Each squad moved as quickly as possible to reinforce a notional friendly squad pinned down by enemy fire. The recorded time started when the squad departed the assembly area on Range 107, and ended upon reaching a designated attack position just prior to the wall obstacle.

Figure A-6 displays all 0311 1-km hike data. On Trial Cycle 5, the HD group data point was removed from analysis as it represents a data outlier. With the exception of this point, all data on the scatterplot are valid for analysis.

Figure A-6. 1-km Hike



The data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.49 for the C group and 0.04 for the HD group.

The C group had a mean of 8.07 minutes. This time is statistically significantly faster in a one-sided t-test than the HD mean time of 8.49 minutes. This difference results in a 5.18%, or 25-second, degradation in hike time. The C group has greater variability, as shown by the 0.14-minute increase in SD (0.53 minutes for the C group and 0.39 minutes for the HD group). See Table A-2 and Table A-3 for detailed analytical results.

A.5.2.6.2 1-km Hike Contextual Comments

The Infantry T&R Manual states that "the maximum assault load weight will be such that an average Infantry Marine will be able to conduct combat operations indefinitely with minimal degradation in combat effectiveness." While moving 1 km with the assault load, the HD group moved 5.18% slower, making them 25 seconds less responsive to reinforcing a unit in contact with the enemy. On any given day (under the same environmental conditions), the C group was faster than the HD group 80% of the time (16 of 20 trial cycles).

A.5.2.6.3 1-km Hike Additional Insights

A purely objective evaluation of 25 seconds is elusive but may possess some practical significance on the battlefield that would reduce the survivability of a unit waiting on an integrated squad. Considering the 12–15 rounds-per-minute sustained rate of fire for the M4 and AK-47 rifles, a single enemy fighter would have the opportunity to take 3 to

4 more well-aimed shots on Marines while waiting for reinforcements from an integrated squad. A fireteam of enemy fighters would have the time to call in indirect 82-mm mortar fire or maneuver during this time delay. The resulting trade in casualty exchange could be significant.

A.5.2.6.4 1-km Hike Subjective Comments

For subjective comments relating to this task, see the 0311 Appendix.

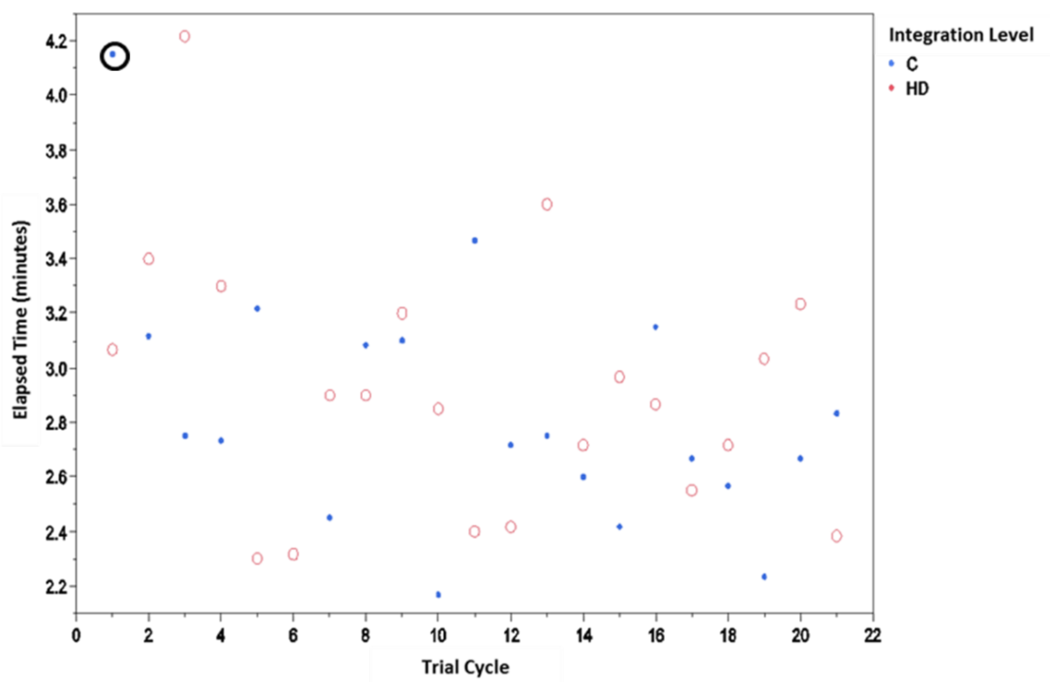
A.5.2.7 Negotiate Obstacle Results

A.5.2.7.1 Negotiate Obstacle Overview

This experimental task assessed the time it took to get a squad of 12 Marines and all equipment over an 8-ft ISO Container (wall obstacle). The squad began this task as soon as they finished the 1-km movement. The recorded time started when the first Marine touched the wall, and time stopped when the last Marine touched both feet down on the other side.

Figure A-7 displays all 0311 Negotiate Obstacle data. On Trial Cycle 1, the C group data point was removed from analysis, as it represents a data outlier due to trial execution. With the exception of this point, all data on the scatterplot is valid for analysis.

Figure A-7. Negotiate Obstacle



The data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.75 for the C group and 0.17 for the HD group.

The C group had a mean of 2.77 minutes. This time is faster (but not statistically significantly) than the HD mean time of 2.92 minutes. There is a 5.33%, or 9-second, degradation in time. In addition, the HD group had a greater variability, as shown by the 0.14-minute increase in SD (0.34 minutes for the C group and 0.48 minutes for the HD group). See Table A-2 and Table A-3 for detailed analytical results.

A.5.2.7.2 Negotiate Obstacle Contextual Comments

According to Sun Tzu: "Speed is the essence of war. Take advantage of the enemy's unpreparedness; travel by unexpected routes and strike him where he has taken no precautions" (MCDP-1, p. 69). Speed and surprise are crucial to success. The enemy is reliably expected to cover obstacles with fires. When possible, obstacles are to be avoided. When this is not possible, the obstacles must be negotiated quickly. While no purely objective standard can be set for the negotiation of the obstacle presented in this task, any decrement in speed translates into increased exposure to enemy fires and greater risk for friendly casualties. On any given day (under the same environmental conditions), the C group was faster than the HD group 63.2% of the time (12 of 19 trial cycles).

A.5.2.7.3 Negotiate Obstacle Additional Insights

Of note, a large discrepancy may be observed when comparing the 0311 obstacle time with the Provisional Infantry (PI) obstacle time. This difference can be explained by the different techniques used by each group. The 0311 squads provided their own security and only used one launch-point up the wall, while the PI squads did not establish security and used three launch-points to get over the wall. Each group (0311 and PI) was consistent within itself, but the two groups should not be compared with each other.

A.5.2.7.4 Negotiate Obstacle Subjective Comments

One subtask involved in negotiating the wall included each Marine throwing his or her 30-lb assault pack up onto the wall. According to leadership observations, females were observed needing help throwing packs onto the wall during the majority of trials. The record of subjective comments reveals 8 occasions in which females required assistance getting their packs onto the wall, compared with zero documented instances of males requiring assistance. For additional subjective comments relating to this task, see the 0311 Appendix.

A.6 Statistical Modeling Results

A.6.1 Statistical Modeling Results Overview

The previous section discussed results only as they pertain to differences due to integration level alone. The goal of statistical modeling, as applied here, is to estimate simultaneously the effect of gender integration levels and other relevant variables on squad performance. (Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.)

For the same seven selected tasks described in the previous section, this section presents an overview of the analysis and results, and then presents the modeling results for each of the tasks.

For each task, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates that an increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time but not a desired outcome for the percent hits or pounds of earth moved outcome. The results indicate where certain patch numbers are significant for a given variable. The experiment tracked Marines within the rifle squad by a patch number that associated their random position within the squad to a specific billet. Table A-5 displays the patch numbers and associated billet titles for the rifle squad.

Table A-5. Patch Numbers and Billet Titles for the Rifle Squad

Patch Number	Billet Title
1	FT 1 Fireteam Leader
2	FT 1 Automatic Rifleman
3	FT 1 Grenadier
4	FT 1 Rifleman
5	FT 2 Fireteam Leader
6	FT 2 Automatic Rifleman
7	FT 2 Grenadier
8	FT 2 Rifleman
9	FT 3 Fireteam Leader
10	FT 3 Automatic Rifleman
11	FT 3 Grenadier
12	FT 3 Rifleman

A.6.2 Mean Value Imputation

Because the personnel data were not available for one volunteer, we impute this Marine's information from all other male volunteers who completed the experiment. A mean value imputation method was used for each measurement using only male volunteers who completed the entire Twentynine Palms portion of the experiment.

A.6.3 0311 Selected Tasks Method of Analysis

Because there are more volunteers than trials for any given 0311 task, a mixed-effects model does not work for the 0311 dataset. Therefore, we model the 0311 selected tasks using ordinary least squares regression.

For the majority of each of the primary metrics for the rifle squad, there are only 41 observations for each result. Because there are 12 Marines in a rifle squad and so few results, the regression model does not have a sufficient number of degrees of freedom to create a model using all types of personnel data for each squad member for each result. Thus each variable combined with integration level is modeled separately. For example, age for each member of the rifle squad (12 variables) and integration level are modeled with the result (response time, percentage hits, or pounds of earth moved) as the response variable. Where possible, a backward stepwise regression using AIC determined which variables are optimal in the model. If there were missing values, backward stepwise could not run, and significant variables are reported based on p-values from the overall model.

A.6.4 0311 Selected Tasks Overall Modeling Results

There are no personnel data variables that are both statistically significant and have a practical impact to the model. Each time personnel data variables are statistically significant in a model, their effects are practically negligible, conflicting, and/or incomplete for the squad (i.e., there are no tasks for which a variable is significant for all, or even most members of the rifle squad).

Integration level, however, consistently appears as statistically significant in each task, and its effect is clear, causal, and practical. Therefore, integration level is the best variable to describe performance for each task. Refer to the section A.5 for the ANOVA summary for each 0311 task mentioned below.

A.6.4.1 7-km Hike

We model elapsed time for the 7-km hike as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- GT score
- PFT crunches
- PFT 3-mile run.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Age
- Height
- Weight
- AFQT score
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 7-km hike time:

- CFT MTC of patches 2, 4, and 9
- CFT MANUF of patches 9 and 12
- Rifle score of patch 11.

The following personnel variables are significant in their respective models and are negatively correlated with the 7-km hike time:

- Age of patches 1 and 6
- Height of patches 10 and 11
- AFQT score of patch 7
- CFT MTC of patches 5 and 10
- Rifle score of patch 5.

The following personnel variables have no significant variables in their respective models:

- None.

There are no patterns for any personnel variables for the 7-km hike. See section A.5.2.1 for the ANOVA summary of this task.

A.6.4.2 Movement to LOA

We model elapsed time for the movement to LOA as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- GT score
- PFT 3-mile run.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Age
- Height
- Weight
- AFQT score
- CFT MANUF
- PFT crunches
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the movement to LOA time:

- Age of patch 2
- Height of patch 5
- Weight of patches 4, 5, and 10
- CFT MTC of patch 8
- CFT MANUF of patches 1 and 8.

The following personnel variables are significant in their respective models and are negatively correlated with the movement to LOA time:

- Age of patches 7, 9, and 11
- Height of patches 9 and 12
- Weight of patches 3, 11, and 12
- AFQT score of patch 9
- CFT MTC of patches 4, 9, and 12
- CFT MANUF of patches 3 and 10.

The following personnel variables have no significant variables in their respective models:

- None.

There are no patterns for any personnel variables for the movement to LOA time. See section A.5.2.2 A.5.2.1for the ANOVA summary of this task.

A.6.4.3 Attack & C-Atk Percent Hits

We model the attack & c-atk percent hits as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- GT score
- PFT 3-mile run.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Age
- Height
- Weight
- AFQT score
- CFT MTC
- CFT MANUF
- Rifle Score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the attack and c-atk percent hits:

- Age of patches 6 and 12
- Height of patch 10
- AFQT score of patches 3, 7, and 10
- CFT MTC of patches 2 and 6

- CFT MANUF of patches 1 and 2
- Rifle score of patches 4 and 10.

The following personnel variables are significant in their respective models and are negatively correlated with the attack and c-atk percent hits:

- Age of patches 4 and 5
- Height of patch 8
- Weight of patches 4, 8 and 12
- AFQT score of patches 4 and 6
- CFT MTC of patch 9
- CFT MANUF of patch 9.

The following personnel variables have no significant variables in their respective models:

- PFT crunches.

There are no patterns for any personnel variables for the attack and c-atk percent hits. See section A.5.2.3 for the ANOVA summary of this task.

A.6.4.4 CASEVAC by FT

We model elapsed time for the CASEVAC as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Age
- Height
- Weight
- AFQT score
- PFT crunches
- PFT three-mile run
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the CASEVAC time:

- Age of patch 2
- GT score of patch 1
- CFT MANUF of patches 3 and 4
- PFT crunches of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the CASEVAC time:

- Age of patch 4
- PFT crunches of patch 1
- Rifle score of patch 3.

The following personnel variables have no significant variables in their respective models:

- CFT MTC.

There are no patterns for any personnel variables for the CASEVAC time. See section A.5.2.4 for the ANOVA summary of this task.

A.6.4.5 Prepare Fighting Positions by FT

We model the amount of earth moved to prepare a fighting position as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report any statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- Weight
- PFT crunches
- PFT 3-mile run.

The following personnel variables are significant in their respective models and are positively correlated with the amount of earth moved for the defensive position:

- Age of patch 4
- AFQT score of patch 4
- GT score of patches 1 and 3
- Rifle score of patch 1.

The following personnel variables are significant in their respective models and are negatively correlated with the amount of earth moved for the defensive position:

- CFT MTC of patches 2 and 4
- CFT MANUF of patches 3 and 4.

The following personnel variables have no significant variables in their respective models:

- Height.

There are no patterns for any personnel variables for the amount of earth moved to prepare a fighting position. See section A.5.2.5 for the ANOVA summary of this task.

A.6.4.6 1-km Hike

We model elapsed time for 1-km hike a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report any statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- GT score
- PFT three-mile run.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Age
- Height

- Weight
- AFQT score
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 1-km hike time:

- Age of patch 2
- Height of patches 5, 8, 9, and 10
- Weight of patch 7
- AFQT score of patches 1, 5, and 11
- CFT MTC of patch 6
- Rifle score of patch 12.

The following personnel variables are significant in their respective models and are negatively correlated with the 1-km hike time:

- Age of patches 6, 8, and 9
- Height of patches 3 and 6
- AFQT score of patches 3, 7, and 10
- Rifle score of patches 3, 8, and 10.

The following personnel variables have no significant variables in their respective models:

- CFT MANUF
- PFT crunches.

There are no patterns for any personnel variables for the 1-km hike time. See section A.5.2.6 for the ANOVA summary of this task.

A.6.4.7 Negotiate Obstacle

We model elapsed time for the negotiate obstacle of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- GT score
- PFT 3-mile run.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Age
- Height
- Weight
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the negotiate obstacle time:

- Age of patch 6
- Height of patches 4 and 11,
- Weight of patches 2, 4, and 5
- AFQT score of patch 5
- CFT MTC of patches 3, 4, 8, and 11
- CFT MANUF of patches 6 and 11
- Rifle score of patches 1, 5, 10, and 12.

The following personnel variables are significant in their respective models and are negatively correlated with the negotiate obstacle time:

- Height of patches 7 and 9
- Weight of patch 9
- CFT MTC of patches 5, 9, 10, and 12
- CFT MANUF of patch 2
- Rifle score of patches 2 and 11.

The following personnel variables have no significant variables in their respective models:

- PFT crunches.

There are no patterns for any personnel variables for the negotiate obstacle time. See section A.5.2.7 for the ANOVA summary of this task.

Appendix to Annex A
0311 Supplemental Information

This appendix provides supplemental information for the 0311 portion of the GCEITF experiment. It provides information regarding the GCEITF leadership subjective comments and additional descriptive and basic inferential statistics not described in Annex A.

Section 1. GCEITF Leadership Subjective Comments

The GCEITF leadership provided comments on their observations of the experiment throughout it execution. Table A A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table A A – Summary of GCEITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement			Requesting extra breaks			Requires extra assistance			Needs no assistance			Compensating for another Marine			Gear pass off			Other			No category			Total
		C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	
7-km Hike	M	9	7	16	5	2	7	1	0	1	6	0	6	2	0	2	0	0	0	1	1	2	1	0	1	35
	F	0	10	10	0	2	2	0	2	2	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	16
Movement to LOA	M	16	5	21	0	0	0	3	5	8	4	1	5	0	0	0	0	0	0	0	1	1	0	0	0	35
	F	0	8	8	0	0	0	0	11	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19
CASEVAC to CCP	M	0	0	0	0	0	0	0	1	1	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	4
	F	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Prepare Fighting Positions	M	1	1	2	0	0	0	0	0	0	4	0	4	0	0	0	0	0	0	0	0	0	0	0	0	6
	F	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1-km Hike	M	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	F	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Negotiate Obstacle	M	0	0	0	0	0	0	0	0	0	11	1	12	0	2	2	0	0	0	0	1	1	0	0	0	15
	F	0	0	0	0	0	0	0	5	5	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	6

Section 2. Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for nine additional 0311 tasks. Annex A contains the descriptive statistics for the remainder of the 0311 tasks. The words “metric” and “task” are used interchangeably throughout this Appendix. They both refer to the experimental task.

The two tables below display the results for nine additional 0311 metrics. Table A B displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels.

Table A C displays ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare the two groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the HD group is different from that in the C group. We present basic inferential statistics for two tasks.

Table A B. 0311 Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
7-km Hike; first km	C	17	10.08	0.41	0.43%
	HD	18	10.13	0.59	
7-km Hike; second km	C	16	10.21	0.43	0.21%
	HD	18	10.24	0.40	
7-km Hike; third km	C	16	9.74	0.35	1.51%
	HD	19	9.89	0.64	
7-km Hike; fourth km	C	16	10.26	0.45	1.85%
	HD	19	10.45	0.63	
7-km Hike; fifth km	C	16	11.14	0.52	3.84%
	HD	19	11.57	0.51	
7-km Hike; sixth km	C	16	12.38	1.70	8.00%
	HD	19	13.37	2.05	
7-km Hike; seventh km	C	16	11.90	0.63	3.37%
	HD	19	12.30	0.65	
Attack (Hits on Target)	C	20	546.40	79.3	-8.95%
	HD	21	497.48	101.40	
C-Atk (Hits on Target)	C	20	74.15	31.06	-18.25%
	HD	21	60.62	25.07	

Table A C. 0311 ANOVA Results and Welch's T-Test Results

Metric	F statistic (df)	F Test P-value	Comparison	2-sided P-Value	1-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
Attack (hits on target)*	2.94 (1, 39)	0.09*	HD-C	0.09*	0.05*	-85.91	-11.94	-96.74	-1.11
C-Atk (Hits on Target)*	2.37 (1, 39)	0.13	HD-C	0.13	0.07*	-25.07	-1.99	-28.45	1.39

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA.

Additional Task Results:

7-km Hike by km. The general trend for the 7-km hike was that the difference between the HD and C groups increased over the course of the hike.

Attack (hits on target). The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.47 for the C group and 0.84 for the HD group.

The C group had a mean of 546.40 hits on targets. This number is statistically significantly higher than the HD group mean of 497.48 hits. The HD group produced 8.95% fewer hits.

- **Contextual Comments**. The ability to effectively engage enemy targets on the battlefield is the hallmark of the Marine rifleman. The number of hits scored by the C group squads was consistently greater than those scored by the HD groups. The operational effect of this is obvious: more hits equates to greater enemy casualties, which means less chance of being hit by the enemy. On any day, under the same conditions, the C group scored more hits than the HD group 80% (16 of 20 trial cycles) of the time.

Counterattack (hits on target). The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.99 for the C group and 0.21 for the HD group.

The C group had a mean of 74.15 hits on targets. This number is statistically significantly higher than the HD group mean of 60.62 hits. The HD group produced 18.25% fewer hits.

- **Contextual Comments**. This event produced extremely variable results for both squad types. The C groups produced a high of approximately 135 hits and a low of about 15. The HD group results ranged between about 25 and 100. This was at least in part a function of the number of rounds each squad had remaining after the “attack” event. Since the number of rounds remaining after the attack was a random variable, not controlled for in any way, the results of this event cannot be used to meaningfully differentiate the performance of the squad types. On any day, under the same conditions, the C group scored more hits than the HD group 55% (11 of 20 trial cycles) of the time.

THIS PAGE INTENTIONALLY LEFT BLANK

Annex B.
Light Armored Reconnaissance (MOS 0313)

This annex details the Light Armored Reconnaissance Vehicle Crewman (MOS 0313) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 3 March – 4 May 2015 at Range 500, aboard the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, CA. The sections outline the Light Armored Vehicle (LAV) Crewman Scheme of Maneuver, Limitations, Deviations, Dataset Description, Descriptive and Basic Inferential Statistics, and Modeling Results.

B.1 Scheme of Maneuver

B.1.1 Experimental Cycle Overview

The LAV Crewmen (MOS 0313) were assessed in a field environment aboard MCAGCC, Twentynine Palms, CA. This timeline included 2 pilot trial cycles, 16 record test cycles, and 3 makeup test cycles. One cycle was equivalent to a 3-day period which included a live-fire day, non-live-fire day and a maintenance day. The record test was conducted as planned from 9 March through 1 May 2015. One makeup test cycle was required. A range maintenance day was conducted every four cycles (12 days) to allow contractor support to conduct range and targetry upkeep, and allowed the Marine volunteers to receive 1 recovery day spent at Camp Wilson. The evaluated subtasks within the cycle were considered the most physically demanding and operationally relevant tasks that a junior 0313 LAV crewman would perform on a recurring basis. An LAV-25 crew is made up of three Marines: a Vehicle Commander (VC), Gunner, and Driver. For each trial, male and female volunteers rotated through the role of gunner and driver on the vehicle; due to required experience, the five VCs remained in their billet throughout the duration of the experiment. The assessment was executed under the supervision of MCOTEA Light Armored Reconnaissance (LAR) Functional Test Manager and a range OIC/RSO from the GCEITF.

B.1.2 Experimental Details

The members of the LAR experiment executed a variety of subtasks over the course of a 3-day trial cycle. The first day consisted of non-live-fire, maintenance-evaluated subtasks divided into three categories: casualty evacuation (CASEVAC) actions, vehicle recovery operations, and maintenance actions. The second day consisted of live-fire evaluated subtasks divided into two categories: prepare for combat actions and main gun target engagements. The third day of the trial cycle consisted of non-evaluated maintenance tasks meant to keep the vehicles and equipment combat ready to continue to execute subsequent trial cycles. Initial Fatigue surveys were given to each volunteer at the beginning of each experimental day, followed by a final Fatigue survey at the completion of that particular day's tasks. Workload surveys were administered immediately following certain specified tasks. Lastly, volunteers completed cohesion surveys at the end of each trial cycle. For survey instruments, see the GCEITF EAP Annexes D and M.

B.1.3 Additional Context

Prior to any trial day with observed subtasks, the Marines bivouacked at a platoon position on their LAVs. On maintenance days, Marines bivouacked in a permanent structure at Range 500. Most subtasks had a maximum time allowed for completion. If a crew exceeded the maximum time allowed, they completed the trial and the time was recorded, but the subtask was flagged as a unit failure. This ensured that all data from the trial and subtask was collected and recorded. All volunteers not randomly selected to participate on a particular day's trials conducted additional tasks outside the assessed events to provide realistic loading. The additional tasks were equivalency tasks to ensure equity between individuals participating in a trial cycle and those not chosen via random selection. These tasks are discussed in detail in the loading section below.

Fatigue surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. It allows for outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) to be accounted for in the analysis of the performance data. Workload surveys collected the volunteers' perceived average and maximal level of exertion during the performance specified tasks. Cohesion surveys provided a method of collecting subjective data relating to each Tank crew's ability to work as a team and their overall perspective on the cohesiveness of the crew.

B.1.4 Scheme of Maneuver Experimental Tasks

B.1.4.1 Conduct Crew / CASEVAC

The CASEVAC portion of the non-live-fire trial day consisted of two separate subtasks: conduct crew evacuation and conduct CASEVAC. Prior to start of the subtasks, each vehicle crewman completed a baseline fatigue survey. Subtasks were conducted as a crew event, and evaluated on the overall time required to evacuate the vehicle and move to a rally point 25 meters away. LAV-25 crews had 9 minutes to complete each task. An initial Fatigue survey was given to each volunteer prior to the beginning of this task.

B.1.4.1.1 Conduct Crew Evacuation

All crewmen (VC, Gunner, and Driver) were required to exit the vehicle safely and move as a crew to the rally point 25 meters away. Marines wore fighting loads and carried individual weapons, as would be expected when evacuating a vehicle in a tactical environment. Although the crew evacuation subtask was not as physically demanding as extracting a casualty from an LAV-25 turret, the ability to evacuate a vehicle quickly is important in the event of a vehicle fire or other life-threatening issue.

B.1.4.1.2 Conduct Casualty Evacuation

This event required the LAV-25 Gunner and Driver to extract a combat-loaded test dummy and its associated protective equipment, weighing 204 lb, from the vehicle and move it to a rally point 25 meters away. This weight was derived from the average weight of a male Marine wearing a standard, vehicle fighting load. The casualty simulated an incapacitated VC, leaving the two remaining crewmen to conduct the evacuation. The crewmen evacuated the dummy from the VC's station by extracting it from the turret and placing it on the front of the vehicle, then moving the simulated casualty to the rally point. The subtask also simulated the arrival and aid of another LAV-25 crew or scouts from a supplementary position to assist in the CASEVAC. This was accomplished by the VC assisting in the evacuation after 6 minutes. Evacuating the dummy from the turret is the most difficult form of evacuation a crew faces in a tactical environment should the vehicle suffer damage that prevents normal turret operations.

B.1.4.2 Vehicle Recovery and Tow Operations

The vehicle recovery portion of the non-live-fire trial day consisted of two separate subtasks: rig for recovery and rig for tow. All tasks were conducted as a crew event. Crews had varying times to finish each subtask and wore personal protective equipment.

B.1.4.2.1 Rig for Recovery

This subtask required the LAV-25 crew to use the winch mounted on the vehicle's front and the side-mounted snatch blocks of the LAV-25 and the LAV Logistics Variant (LAV-L). With the vehicles spaced 10 m apart, the crewmen paid out the winch cable and used the two snatch blocks, each weighing 75 lb, to create a three-point recovery system with the shackles and tow-points of the vehicles. An LAV mechanic inspected the equipment to ensure it was properly constructed and that the crew followed all necessary safety requirements. Upon verification, the crew then detached and stowed all materials. The subtask was complete when all tools and equipment were returned to their proper locations. The crew had 25 minutes to accomplish this task. The ability to quickly recover a mired vehicle is an important skill that all crewmen must possess. A three-point recovery system is used often, as it only requires two vehicles to establish and can be accomplished fairly quickly. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.2.2 Rig for Tow

This subtask required the LAV-25 crew to use the tow bar, weighing approximately 175 lb., located on the side of the LAV-L. With the two vehicles spaced 10 meters apart, the crewmen detached the tow bar from the LAV-L and connected it to the tow points on the LAV-25. They then ground guided the LAV-L and, with one crewman holding the tow bar, attached the eye of the tow bar to the tow pintle of the LAV-L. An LAV mechanic inspected the equipment to ensure it was properly constructed and that the crew followed all necessary safety requirements. Upon verification, the crew then detached and stowed all materials. The subtask was complete when all tools and equipment were returned to their proper locations. The crew had 35 minutes to accomplish this task. Marines completed workload surveys at completion. Vehicles will often break down, either due to mechanical issues or enemy action, and the ability to rig a vehicle for towing, quickly, directly affects combat effectiveness. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.3 Conduct Maintenance Actions

The maintenance actions portion of the non-live-fire trial day consisted of three separate subtasks: remove scout hatch armor, remove side panel armor, and change a tire. All tasks were conducted as a crew event. Prior to the start of each maintenance subtask, Marines completed a mid-trial fatigue survey. Crews had varying times to finish each subtask, and wore personal protective equipment throughout. Each task was evaluated on the time to reach the mid-point of the task and the overall time to completion.

B.1.4.3.1 Mount/Remove Scout Hatch Armor

The vehicle crew used the tools and equipment (SL-3) associated with the LAV-25 to remove each armor panel mounted on the back hatches. Each panel is a solid piece of composite armor weighing approximately 125 lb. The crew dropped and remounted one hatch panel, followed by the second. The crew had 30 minutes to accomplish this task. Marines completed workload surveys at completion. Removing armor panels, while not conducted as regularly as other maintenance actions, is required when the panels are damaged or during wash-downs and cleaning after extended operations. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.3.2 Mount/Remove Side-Panel Armor

The vehicle crew used the SL-3 associated with the LAV-25 to remove the front-most armor panels, mounted on the left and right sides. Each panel is a solid piece of composite armor weighing approximately 60 lb. The crew dropped and remounted one panel, followed by the second. The crew had 20 minutes to accomplish this task. Marines completed workload surveys at completion. Removing armor panels, while not conducted as regularly as other maintenance actions, is required when the panels are damaged or during wash-downs and cleaning after extended operations. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.3.3 Mount/Remove Spare Tire

The vehicle crew used the SL-3 associated with the LAV-25 to remove a tire, weighing approximately 200 lb, and mount the spare. For the first half of the subtask, the crew jacked up the vehicle, removed a vehicle tire, mounted the vehicle's spare tire, and tightened the lug nuts. The crew then removed the spare tire, mounted the original tire, tightened the lug nuts, and returned the spare tire and all additional equipment to the proper location. The crew manually lifted the spare tire from the ground to its mounting bracket on the side of the LAV-25. The crew had 40 minutes to accomplish this task. After completion, Marines completed a post-trial fatigue and a post-trial workload survey. The ability to rapidly change a tire has critical implications to ensuring mission accomplishment. Depending on the terrain, it would not be unusual for a vehicle crew to change multiple tires throughout the course of an operation. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer. Additionally, a final Fatigue survey was administered at the end of this task.

B.1.4.4 Prepare LAV for Combat

The preparation for combat actions portion of the live-fire trial day consisted of six separate subtasks, as discussed below. Prior to the start of the subtasks, each vehicle crewman completed a baseline fatigue survey. After completion of each subtask, Marines completed a mid-trial workload survey. Many of the tasks were conducted as an individual Gunner event that afforded assistance, if required, while some were crew events. Crews had varying times to finish each subtask, and wore personal protective equipment throughout. With the exception of load and stow ammunition, each task was evaluated on the time to reach the mid-point of the task and the overall time to completion.

B.1.4.4.1 Prepare the M240 Coax

This subtask required the LAV-25 Gunner, without assistance, to remove, disassemble, reassemble, and install the M240 coaxially mounted machine gun. Disassembly was achieved when the Gunner removed the bolt assembly from the weapon system. The Gunner had 15 minutes to accomplish this task. The ability to rapidly troubleshoot and correct issues within the turret is a critical skill for LAV Gunners. As the very basis of their role on a vehicle crew, a Gunner must rapidly accomplish any task involving the LAV-25's weapons systems. An initial Fatigue survey was given to each volunteer at the beginning of this task since it was the first task of an experimental day. Additionally, a workload survey was administered to measure the average and maximum workload experienced by each volunteer.

B.1.4.4.2 Manually Traverse and Elevate the LAV-25 Turret

This subtask required the LAV-25 Gunner, without assistance, to manually traverse, elevate, and depress the vehicle turret. The Gunner used the manual traverse and elevation controls within the turret to accomplish the subtask. The Gunner manually traversed the turret 180 degrees from vehicle centerline. The Marine then elevated the barrel until the M240 Coax barrel was even with the barrel of the VC's M240B. The Gunner then depressed the barrel until it touched the rear of the vehicle and returned it to center. Finally, the Gunner traversed the turret 180 degrees back to its original starting location. The Gunner had 5 minutes to accomplish this task. In the event of power loss within the turret, a Gunner must rapidly manipulate the turret to acquire and engage targets. This skill is regularly assessed with manual engagements during table qualification gunnery. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.4.3 Prepare the M242 Main Gun

This subtask required the LAV-25 Gunner, without assistance, to remove, disassemble, reassemble, and install the M242 Bushmaster Cannon, weighing approximately 263 lb. The Driver was permitted to assist the Gunner after 30 minutes had elapsed. Disassembly was achieved when the Gunner disassembled the bolt and track assembly and removed the barrel from the weapon system. The Gunner and Driver had 45 minutes to accomplish this task. In many situations, other crewmembers will be unavailable to assist the Gunner in manipulation and troubleshooting of the main gun. The Gunner must rapidly troubleshoot the primary weapon system of the platform. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.4.4 Load and Stow Additional Ammunition

This subtask required the LAV-25 crew to load and stow a full complement of 25-mm ammunition into the vehicle. An LAV-25 can hold 150 rounds of High Explosive (HE) and 60 rounds of Armor Piercing (AP) ammunition, linked and prepared in the vehicle's HE and AP ready-boxes. Two more uploads of each ammo type is stored within the vehicle in their original cans, for a total of 180 AP and 450 HE rounds. Dummy rounds weighing just over a pound each were used to simulate the linked rounds. These dummy rounds were broken out of their containers and linked prior to the start of the subtask. Ten cans of HE training rounds, each weighing 50 lb, and four cans of AP training rounds each weighing 55 lb were used. The crew had 20 minutes to accomplish this task. Without assistance, a crew must load and stow their ready rounds and additional uploads on their vehicle as part of normal combat preparation. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.4.5 Upload the M242 Main Gun

This subtask required the LAV-25 crew to upload the HE and AP rounds from their respective ready-boxes into the M242 main gun. This was accomplished by ratcheting the rounds from the ready box, through the gun's feed chutes, and into the feeder assembly on the gun. The AP ready-box is located near the Driver, while the HE ready-box is located next to the Gunner. The crew had 20 minutes to accomplish this task. Without assistance, a crew must upload their weapons systems as part of normal combat preparation. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.4.6 Upload the M240 Coax

This subtask required the LAV-25 Gunner, without assistance, to upload a belt of 200 7.62-mm rounds from coax ready box into the M240 coaxially mounted machine gun. This was accomplished by feeding the rounds through the feed chute that extends across the turret and onto the feed tray of the M240 coax. The Gunner had 10 minutes to accomplish this task. Without assistance, a crew must upload their weapons systems as part of normal combat preparation. A workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

B.1.4.5 Engage Main Gun Targets

The main gun engagement portion of the live-fire trial day consisted of five separate subtasks, as discussed below. All tasks were conducted as crew events. Crews had varying times to finish each subtask, and wore their vehicle fighting load throughout.

Live-fire engagements were adapted from table VI gunnery qualification engagements, which all vehicle crews are required to fire to achieve qualified status.

B.1.4.5.1 Gunner's Defensive Engagement 1

The first live-fire engagement was conducted from a defensive battle position. The crews engaged two vehicle targets, one with AP and one with HE ammunition. The time to engage each target and number of bursts required to kill each target was recorded. An LAV Crew Evaluator (LCE) observed and completed an LCE form, which is a standard LAV gunnery template that evaluates a crew's performance based on exposure times, fire commands, correct engagement procedures, and other metrics. Live-fire engagements were adapted from table VI gunnery qualification engagements.

B.1.4.5.2 Gunner's Offensive Engagement 1

The first live-fire offensive engagement was conducted on the move through a maneuver box. The crews engaged one vehicle target with the LAV-25 main gun and a bank of troop targets with the coaxially mounted machine gun. The time to engage each target and number of bursts required to kill the main gun target was recorded. An LCE observed and completed an LCE form. Live-fire engagements were adapted from table VI gunnery qualification engagements.

B.1.4.5.3 Main Gun Remedial Action

This subtask required the LAV-25 Gunner, with assistance from the VC and Driver, to conduct remedial action on the main gun. The crew downloaded the ammunition from the main gun, removed the feeder assembly and ensured it was correctly timed, and then reinstalled and uploaded ammunition. A data collector recorded the time required to disassemble and reassemble the weapon system, and the overall time required to complete the subtask, as well as the nature of the assistance provided by the Driver and VC. Disassembly was achieved when the Gunner removed the feeder assembly from the weapon system. The vehicle crew had 40 minutes to accomplish this task. This task simulates a misfire on the main gun, and the actions that the crew would take to get their vehicle back in the fight.

B.1.4.5.4 Gunner's Defensive Manual Engagement 2

The second live-fire defensive engagement was conducted from a defensive battle position. This engagement required the gunner to use AP ammunition to engage a vehicle target without turret power. The Gunner used the manual traverse and elevation controls within the turret to acquire and destroy the target. The time to engage the target and number of bursts required to kill the target was recorded. An LCE observed and completed an LCE form. Live-fire engagements were adapted from table VI gunnery qualification engagements.

B.1.4.5.5 Gunner's Offensive Engagement 2

The second live-fire offensive engagement was conducted on the move through a maneuver box as a retrograde engagement. With the turret oriented over the rear end of the vehicle, the crews engaged one vehicle target with the LAV-25 main gun and a bank of troop targets with the coaxially mounted machine gun. The time to engage each target and number of bursts required to kill the main gun target was recorded. An LCE observed and completed an LCE form. Live-fire engagements were adapted from table VI gunnery qualification engagements. After completion of this task, Marines completed a post-trial fatigue, a post-trial workload survey, and a crew cohesion survey.

B.1.5 Loading Events

The loading events ensured, to the greatest extent possible, equity of physical activity among all volunteers, and attempted to simulate the intense physical and mental workload present in combat or in a combat-focused training environment. Loading activities included nightly security and observation posts, maintenance actions, and execution of all subtasks to ensure equitable effort. The first enduring task was turret watch, held by all Marines scheduled to participate in the following days' trials and consisted of 2 hours of watch. This watch only occurred on nights prior to evaluated assessed events. On the platoon's maintenance day, for preventative maintenance checks and services (PMCS), Marines conducted maintenance on winch cables and reset any gear stowed incorrectly during evaluations, as well as any major maintenance actions required from vehicle issues. At the end of the established 12-day cycle, each crew conducted lubrication orders on their vehicle prior to movement back to Camp Wilson. The day spent at Camp Wilson was a range maintenance day where contractors performed any repairs or maintenance on all targets used during the 12-day cycle. Weekly maintenance checks and services were conducted every other PMCS day.

B.1.5.1 Loading Challenges

There were multiple challenges enacting a loading plan that simulated the physical and mental difficulties of combat or a combat-training environment. The foremost problem was the mindset of the volunteers. When executing subtasks during loading trials, Marines moved slower or exerted less effort knowing the trial was not part of formal data collection efforts. Concerns for troop welfare and morale support led to decisions that allowed Marines to bivouac in a covered administrative position every maintenance day during a trial rather than maintain a tactical posture on their vehicles over the course of multiple trials. This administrative posture alleviated the cumulative effect that fatigue has on Marines in a tactical environment.

B.1.6 Scheme of Maneuver Summary

The LAR Experiment consisted of a 3-day trial cycle consisting of non-live-fire, live-fire, and maintenance days. The non-live-fire trials had seven subtasks in three categories: CASEVAC, recovery operations, and maintenance actions. The live-fire trials had 11 subtasks in two categories: prepare for combat and live-fire engagements. During the course of the experiment, the LAR platoon executed 2 pilot trial cycles, 16 record trial cycles, and 1 makeup trial cycle. During trial execution, Marines wore a standard, vehicle fighting load while conducting all subtasks, which weighed approximately 35 lb. When not participating in a trial, Marines executed all subtasks to ensure equivalency loading amongst volunteers and wore the standard vehicle fighting load. Knowing their efforts were not being directly observed and recorded did not ensure equivalency of effort. Additional physical loading included watch rotations, vehicle maintenance, and a vehicle bivouac site. A rotation into an administrative posture every maintenance day alleviated the cumulative effect fatigue has on Marines in a tactical environment.

B.2 Limitations

B.2.1 Limitations Overview

The members of the LAR assessment executed a variety of subtasks over the course of a 3-day trial cycle meant to encompass, as accurately as possible, the demands of the LAV Crewman MOS. The evaluated subtasks within the cycle were considered the most physically demanding and operationally relevant tasks that a junior LAV crewman would be expected to perform on a recurring basic. These subtasks were executed as they would normally be performed; however, due to time constraints, data collection requirements, and other factors, limitations and artificialities existed in the assessment. These limitations had little to no effect on data collection, but impacted the cumulative loading effects and fatigue of the volunteers.

B.2.1.1 Number of Volunteer Participants

For the LAV experiment, 14 male and 7 female volunteers began the experiment, but by the end 13 males and 7 females were able to complete the assessment. The results presented in this annex are based on the performance of 21 Marines.

B.2.2 Additional Duties as Dismounted Infantry

During normal LAR operations, vehicle crewmen may be required to operate as 0311 dismounted infantry (known as scouts within the LAR community) at any time. The continuing actions of combat operations, such as vehicle maintenance, turret watch, and patrolling, are shared between the scouts and the crewmen. During defensive operations, the four-man scout team may augment the three-man vehicle crew and share in the collective burden of turret watch and observation throughout the course of a night, if not otherwise engaged. Crewmen will be expected to conduct dismounted operations, such as patrolling or defensive position improvement, to aid the scouts.

The limited scope, duration, and resources available meant crewmen did not participate in any dismounted operations nor were any scouts used during the execution of the assessment. The Infantry Functional Test comprehensively captured most of the basic dismounted actions of scouts during their execution on the 100 series ranges. Crewmen would normally participate in a number of dismounted actions that contribute to the overall loading and fatigue inherent in combat operations and training exercises.

B.2.3 Cycle of Operations Constraints

The LAR assessment attempted to observe and capture the most physically demanding tasks at the core of the LAV crewman billet description, but it could not fully replicate the demands of continuous 24-hour field operations, either in training or combat environment. A vehicle crew typically finished their assessed non-live-fire maintenance tasks within 2 hours on their first trial day and their assessed live-fire tasks within 4 hours on their second trial day. Under normal operating conditions, Marines would not complete their day and responsibilities in such a short time. Limited maintenance and data collection resources, as well as time constraints, limited additional observed loading.

Although many of the tasks required an intense amount of effort for a short time, Marines were exempt from auxiliary tasks and duties that come with sustained field operations. To ensure sustainability over a test that could have spanned 75 days, Marines spent two nights on their vehicles and one night in an improved structure during the 3-day cycle of operations. During normal operations, a platoon would spend nearly all of their time on their vehicles, executing a robust watch and patrolling rotation throughout the course of the night. Human factors and morale concerns prevented a fully tactical environment in which the Marines executed these continuing actions for extended periods. Observation and time constraints made many such actions impractical, as the data collection team would be unable to gather information as the Marines executed such duties. Marines returned from the field every 12 days, an artificiality that may not be present during extended operations.

B.2.4 Missing Trials that Affect Analysis

The Toughbook programming used to record time hack data for Prep M242 Main Gun and Conduct Main Gun Remedial Action incorrectly captured start and stop times for the tasks, leading to lost data. The program captured the disassembly time for the task, but did not record the completion time, leading to lost reassembly time data for 29 trials. Therefore, only 56 trials were available for certain metrics associated with those tasks.

B.2.5 0313 Limitations Summary

The LAR assessment sought to replicate the realities of field conditions and loading Marines experience during field training and combat operations. The assessment team balanced these demands with the necessity of collecting equitable and uniform data throughout the assessment's duration and across many participants. This led to certain artificialities that departed from normal operations but did not affect the validity of data collection or the assessment as a whole.

B.3 Deviations

Deviations to the execution of the LAV Crewman scheme of maneuver were made and annotated in the Experimental Data Report; however, there were no deviations that occurred that affected the analysis methodology outlined in the Experimental Assessment Plan.

B.4 Data Set Description

B.4.1 Data Set Overview

The LAV experiment timeline included 2 pilot trial cycles, 16 record trial cycles, and 3 makeup trial cycles. Pilot data were not analyzed. The record trial cycles were conducted as planned from 9 March through 1 May 2015. One makeup trial cycle was required and conducted from 2-4 May 2015.

B.4.2 Record Test Volunteer Participants

Twenty-one Marine volunteers participated in the Light Armored Vehicle (LAV) Crewman experiment, including 14 males and 7 females. One male volunteer was dropped from the experiment after 2 cycles, leaving a total of 20 volunteers, 13 males and 7 females.

B.4.3 Planned, Executed, and Analyzed Trial Cycles

Sixteen record cycles equates to 80 planned trials for each task. With the inclusion of one makeup trial cycle, volunteers executed 85 trials. Of note, there are several occurrences of missing data. As discussed above, the Toughbook programming used to record time data for Prep M242 Main Gun and Conduct Main Gun Remedial Action did not record the completion time, leading to lost reassembly time data for 29 trials. The data management team identified the problem on 26 March and fixed the program for the subsequent trial cycles. Live-fire Engagement Data is missing on several occasions due to wind delays on the range. Sustained winds above 35 mph have the potential to damage targetry at Range 500, and gunnery went into a check fire status any time these conditions were observed. Wind delays caused the loss of 2 trials on 13 March and 5 trials on 21 April. One trial was not executed on 17 March due to the loss of a vehicle commander mid cycle. Any other lost trial data is due to data collector or Toughbook program error, or equipment malfunction that led to an invalid data point.

Table B-1. 0313 Planned, Executed and Analyzed Trial Cycles[1]

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
Crew Evacuation from Turret	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
CASEVAC; Overall Time	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
CASEVAC; Time to Extract	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
CASEVAC; Time to Rally Point	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
Rig for Recovery; Overall Time	C	27	29	29	
	LD	27	28	27	Data Collector/Toughbook (DC/TB) Error; 9 Mar
	HD	26	28	28	
Rig for Recovery; Time to Rig	C	27	29	28	DC/TB Error; 11 Apr
	LD	27	28	28	
	HD	26	28	28	
Rig for Recovery; Time to Stow	C	27	29	27	DC/TB Error; 1 Apr, 2 May
	LD	27	28	27	DC/TB Error; 7 Apr
	HD	26	28	28	
Rig for Tow; Overall Time	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
Rig for Tow; Time to Rig	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
Rig for Tow; Time to Stow	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
Remove Scout	C	27	29	29	

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
Hatches; Overall Time	LD	27	28	27	DC/TB Error; 19 Mar
	HD	26	28	28	
Remove Scout Hatches; 1st Hatch	C	27	29	29	
	LD	27	28	27	DC/TB Error; 19 Mar
	HD	26	28	28	
Remove Scout Hatches; 2nd Hatch	C	27	29	29	
	LD	27	28	27	DC/TB Error; 19 Mar
	HD	26	28	28	
Remove Side Panels; Overall Time	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
Remove Side Panels; 1st Panel	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
Remove Side Panels; 2nd Panel	C	27	29	29	
	LD	27	28	28	
	HD	26	28	28	
Mount/Remove Spare Tire; Overall Time	C	27	29	26	DC/TB Error; 9 Mar, 1 Apr, 11 Apr
	LD	27	28	27	DC/TB Error; 23 Apr
	HD	26	28	28	
Mount/Remove Spare Tire; Mount Spare	C	27	29	27	DC/TB Error; 9 Mar, 11 Apr
	LD	27	28	27	DC/TB Error; 23 Apr
	HD	26	28	28	
Mount/Remove Spare Tire; Remount Original	C	27	29	27	DC/TB Error; 9 Mar, 1 Apr
	LD	27	28	28	
	HD	26	28	28	
Prep M240 Coax; Overall Time	C	27	29	29	
	LD	27	27	27	
	HD	26	28	27	Equip Failure; 15 Apr
Prep M240 Coax; Disassembly	C	27	29	29	
	LD	27	27	27	
	HD	26	28	27	Equip Failure; 15 Apr
Prep M240 Coax; Reassembly	C	27	29	29	
	LD	27	27	27	
	HD	26	28	27	Equip Failure; 15 Apr
Manually	C	27	29	29	

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
Manipulate Turret; Overall Time	LD	27	27	27	
	HD	26	28	28	
Manually Manipulate Turret; Traverse Time	C	27	29	28	DC/TB Error; 23 Mar
	LD	27	27	27	
	HD	26	28	26	DC/TB Error; 13 Mar, 17 Mar
Manually Manipulate Turret; Elevation Time	C	27	29	28	DC/TB Error; 23 Mar
	LD	27	27	27	
	HD	26	28	26	DC/TB Error; 13 Mar, 17 Mar
Prep M242 Main Gun; Overall Time	C	27	29	18	TB Error; 3-26 Mar
	LD	27	27	18	TB Error; 3-26 Mar
	HD	26	28	18	TB Error; 3-26 Mar
Prep M242 Main Gun; Disassembly	C	27	29	27	DC/TB Error; 12 Apr, Equip Failure; 15 Apr
	LD	27	27	27	
	HD	26	28	28	
Prep M242 Main Gun; Reassembly	C	27	29	17	TB Error; 3-26 Mar, DC/TB Error; 12 Apr
	LD	27	27	18	TB Error; 3-26 Mar
	HD	26	28	18	TB Error; 3-26 Mar
Load and Stow Additional Ammunition	C	27	29	29	
	LD	27	27	27	
	HD	26	28	28	
Upload M240 Coax	C	27	29	29	
	LD	27	27	27	
	HD	26	28	28	
Upload M242 Main Gun	C	27	29	29	
	LD	27	27	25	DC/TB Error; 26 Mar, 30 Apr
	HD	26	28	27	DC/TB Error; 17 Mar
Defensive Engagement #1	C	27	27	27	Wind Delays; 13 Mar, 21 Apr
	LD	27	25	25	Wind Delays; 13 Mar, 21 Apr, VC Drop; 17 Mar
	HD	26	25	25	Wind Delays; 13 Mar, 21 Apr
Offensive Engagement #1	C	27	27	27	Wind Delays; 13 Mar, 21 Apr
	LD	27	25	25	Wind Delays; 13 Mar, 21 Apr, VC Drop; 17 Mar

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
	HD	26	25	25	Wind Delays; 13 Mar, 21 Apr
Conduct Remedial Actions; Overall Time	C	27	27	18	TB Error; 3-26 Mar
	LD	27	27	18	TB Error; 3-26 Mar
	HD	26	27	18	TB Error; 3-26 Mar
Conduct Remedial Actions; Disassembly	C	27	27	27	Wind Delay 13 Mar
	LD	27	27	27	Wind Delay 13 Mar
	HD	26	27	27	Wind Delay 13 Mar
Conduct Remedial Actions; Reassembly	C	27	27	18	TB Error; 3-26 Mar
	LD	27	27	18	TB Error; 3-26 Mar
	HD	26	27	18	TB Error; 3-26 Mar
Defensive Manual Engagement #2	C	27	27	22	Failed Engagements Excluded
	LD	27	25	22	Failed Engagements Excluded
	HD	26	25	21	Failed Engagements Excluded
Offensive Engagement #2	C	27	27	27	Wind Delays; 13 Mar, 21 Apr
	LD	27	25	25	Wind Delays; 13 Mar, 21 Apr, VC Drop; 17 Mar
	HD	26	25	25	Wind Delays; 13 Mar, 21 Apr

1. Data was not captured or captured incorrectly due to human (Data Collector - DC) error or data processing equipment (Toughbook - TB) error. Wind delays also caused a loss of data on some occasions. Some data points were classified as outliers or potential influential points and were excluded from the analysis as described in the methodology section.

B.5 Descriptive and Basic Inferential Statistics

B.5.1 Descriptive Statistics Overview

The evaluated subtasks within the cycle were considered the most physically demanding and operationally relevant tasks that a junior 0313 LAV crewman would perform on a recurring basis. Twelve out of 18 tasks are presented in this section. Appendix B contains the descriptive statistics for the remaining 0313 tasks.

As described in the Scheme of maneuver (see Section B.1), an LAV-25 crew is made up of three Marines: a VC, Gunner, and Driver. For each trial, male and female volunteers rotated through the role of gunner and driver. Due to required experience for the VC billet, Marines selected to serve as VCs were rotated among the five billets for the duration of the experiment. There were three integration levels assessed for this task: a control (C) group of all male Marines, a low-density (LD) group of one female and two male Marines, and a high density (HD) group of two females and one male Marine.

In addition to integration level comparisons, there are certain tasks also analyzed by critical billet. Although most tasks are completed as an LAV Crew where Marines can assist other Marines, certain tasks (such as turret manipulation) require individuals holding the critical billet (such as vehicle gunner) to perform his or her specific duties with little or no assistance. A critical billet was identified for tasks where a physically demanding duty was primarily performed by that billet.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA), Tukey Tests (or non-parametric equivalent as necessary), and scatter plots. The first table titled Descriptive Statistics displays the metric, integration levels, sample sizes, means and standard deviations. The second table shows ANOVA and Tukey test results including, but not limited to, metrics, p-values suggesting statistical significance, integration level elapsed time differences and percentage differences between integration levels. If non-parametric tests were needed, the second table displays these results instead of ANOVA and Tukey test results. Subsequent subsections will cover each task in detail along with scatter plots of the data. If p-values are less than the a-priori significance level of 0.10, we conclude that there is statistical evidence that the mean elapsed time for the experimental groups, LD and HD, are different from the C group.

Contextual comments and additional insights tying back to each experimental task are incorporated. Two separate subsections for each task are presented in order to report analytical results with and without potential influential points. These potential influential points form part of the valid data.

Lastly, special caution should be taken when comparing similar tasks executed by different MOSs across the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks such as distances, techniques, leadership, load carried, group size, and group composition. The words “metric” and “task” are used interchangeably throughout this Annex. They both refer to the experimental task.

B.5.2 Conduct Casualty Evacuation Overview

The ability to perform a fast and efficient casualty or crew evacuation is a vital skill, especially while confined to a vehicle. This event required the LAV-25 Gunner and Driver to extract a combat loaded test dummy weighing 205 lbs. from the vehicle and move it to a rally point 25 meters away. The casualty simulated an incapacitated Vehicle Commander, thus leaving the two remaining crewmen to conduct the evacuation. The extraction required crewmen to physically lift the dummy straight up from the turret and move it to the front of the vehicle; this task also required either a fireman carry or two-man carry technique to move the casualty to the rally point. The subtask also simulated the arrival and aid of another LAV-25 crew or scouts from a supplementary position in order to assist in the CASEVAC. This was accomplished by the VC assisting in the evacuation after 6 minutes. The start time for this task began when all crewman were in their stations. Times were recorded once the casualty was extracted from the turret and placed on the front slope of the vehicle. Time was stopped when the crew reached the rally point 25 meters away. Additionally, another time hack was recorded once a crew reached six minutes and the simulation allowed for another crewman to assist in the CASEVAC. Evacuating the dummy from the turret simulated the most difficult form of evacuation a crew may face in a tactical environment after vehicle damage prevents normal turret operations.

B.5.2.1 Conduct Casualty Evacuation Scatterplots

The full data set is displayed in the scatter plots below. This data set did not contain outliers or potential influential data points. All data points shown in the scatter plots were determined to be valid and used in the analysis and modeling. The first plot shows the total elapsed time to evacuate the casualty from the turret and movement to the rally point. The second plot only displays the elapsed time for the extraction of the casualty from the turret. The third plot shows the times recorded for the movement of the casualty to the rally point after it was placed on the hood of the vehicle.

Figure B-1. Evacuate Casualty from Vehicle by Integration Level

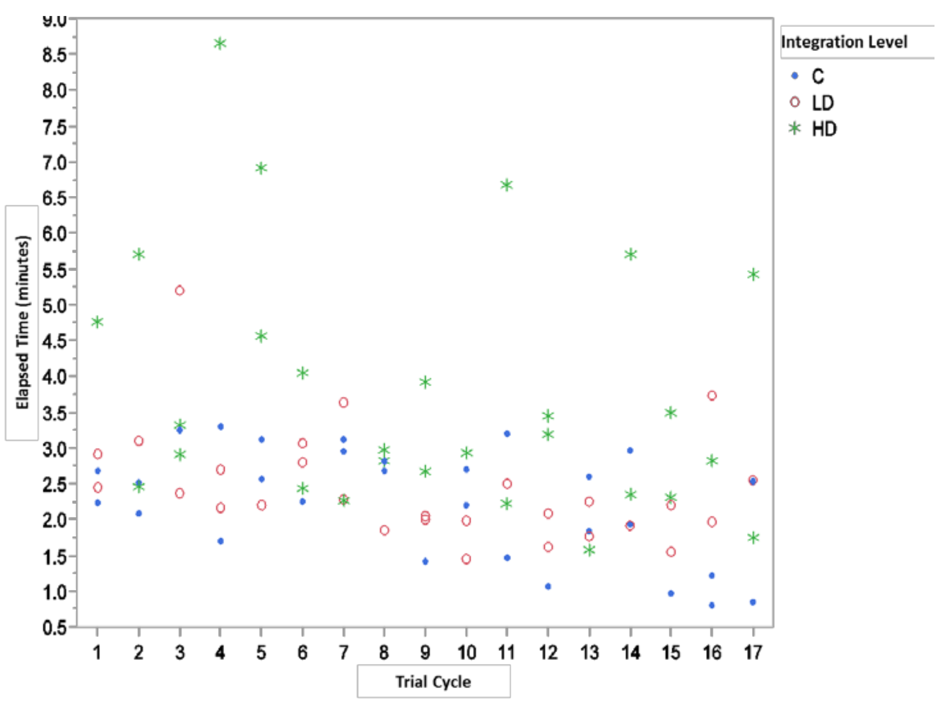


Figure B-2. Evacuate Casualty from Vehicle; Extraction from Turret

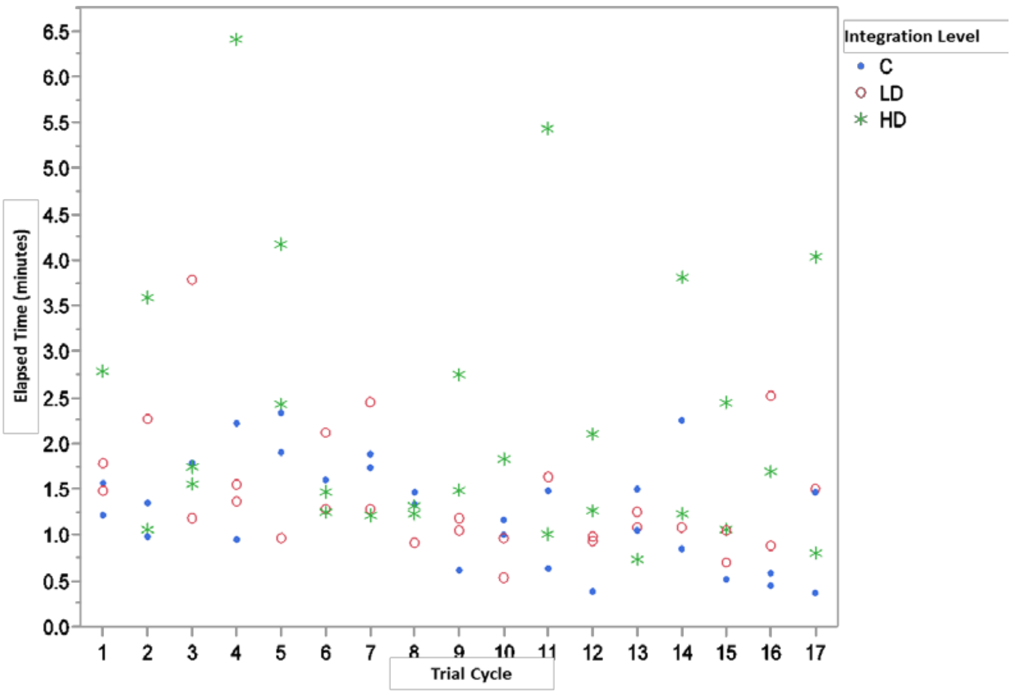
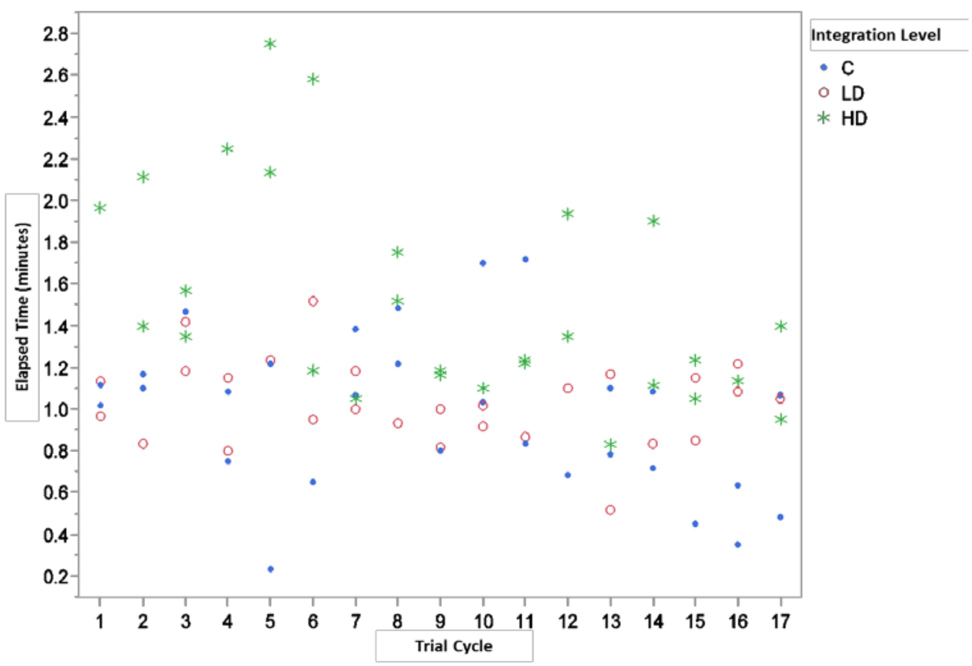


Figure B-3. Evacuate Casualty from Vehicle; Movement to Rally Point



B.5.2.2 Conduct Casualty Evacuation Data Tables and Analysis

The tables below summarize the results of the task, Conduct Casualty Evacuation. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-2. Evacuate Casualty from Vehicle Results

Metric	Integration Level	Sample Size	Mean	SD
Evac Casualty from Vehicle (minutes) *	C	29	2.24	0.78
	LD	28	2.44	0.79
	HD	28	3.73	1.73
Evac Casualty from Vehicle; Extraction (minutes) *	C	29	1.26	0.58
	LD	28	1.42	0.68
	HD	28	2.21	1.44
Evac Casualty from Vehicle; Movement to Rally Point (minutes) *	C	29	0.98	0.38
	LD	28	1.02	0.21
	HD	28	1.51	0.51

*Indicates there is a statistically significant difference in a two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-3. Evacuate Casualty from Vehicle ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Evac Casualty from Vehicle (minutes)	8.52 (2, 51)	< 0.01††	LD-C	0.20	8.88%	0.34	-0.07	0.47	-0.15	0.55
			HD-C	1.48	66.20%	< 0.01††	1.02	1.95	0.88	2.09
			HD-LD	1.29	52.65%	< 0.01††	0.82	1.75	0.68	1.89
Evac Casualty from Vehicle; Extraction (minutes)	5.19 (2, 50)	0.01††	LD-C	0.16	12.48%	0.35	-0.06	0.38	-0.12	0.44
			HD-C	0.95	75.05%	< 0.01††	0.57	1.33	0.45	1.44
			HD-LD	0.79	55.64%	0.01††	0.40	1.18	0.28	1.30
Evac Casualty from Vehicle; Movement to Rally Point (minutes)	12.29 (2, 49)	< 0.01††	LD-C	0.04	4.24%	0.61	-0.06	0.15	-0.09	0.18
			HD-C	0.54	54.78%	< 0.01††	0.38	0.69	0.34	0.73
			HD-LD	0.49	48.48%	< 0.01††	0.36	0.63	0.32	0.67

*Indicates there is a statistically significant difference in a two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

††Indicates statistical significance using a robust ANOVA (accounts for unequal variances). For comparisons utilizing robust ANOVA, Tukey confidence intervals have been replaced by individual Welch's T-tests with p-values compared to a Bonferroni corrected threshold.

For the overall time to evacuate the casualty, the C group data are normally distributed with Shapiro-Wilk Test p-value of 0.05, while the LD and HD groups are not normally distributed as evidenced by p-values of 0.001 and 0.004, respectively. Additionally, since the standard deviation of the HD group is more than twice that of the C and LD groups, we recommend using the robust ANOVA results presented above. We proceed with presenting robust ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.01).

For the extraction portion of the time, the C group data are normally distributed with Shapiro-Wilk Test p-value of 0.36, while the LD and HD groups are not normally distributed as evidenced by p-values less than 0.01. Additionally, since the standard deviation of the HD group is more than twice that of the C and LD groups, we recommend using the robust ANOVA results presented above. We proceed with presenting robust ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.01).

For the movement to rally point portion of the time, the C and LD group data are normally distributed with Shapiro-Wilk Test p-values of 0.70 and 0.90, respectively, while the HD group is not normally distributed as evidenced by a p-value of 0.01. Additionally, since the SD of the HD group is more than twice that of the LD group, we recommend using the robust ANOVA results presented above. We proceed with presenting robust ANOVA results, since they are confirmed by a Kruskal-Wallis Test (p-value < 0.01).

For the overall time to evacuate the casualty the C group had a mean time of 2.24 minutes, which was statistically significantly faster than the LD and HD groups who recorded mean times of 2.44 minutes and 3.73 minutes, respectively. The C group’s 66.20% faster mean time when compared to the HD group was statistically significant. The other two metrics shown in Table B-2, Evacuation Casualty from Vehicle Extraction and Movement to Rally Point, also show statistical significance when comparing the C group and HD group. The percentage differences equate to degradation in evacuation times for LD and HD groups when compared to the C group. Additionally, all HD groups for the three metrics above show greater variability as shown by the SD column. See Table B-1 and Table B-2 for detailed analytical results.

B.5.2.3 Conduct Casualty Evacuation Contextual Comments

Rapidly removing a wounded crewman from the vehicle is the first step to providing lifesaving care. Longer times to extract a casualty could lead to greater exposure to various threats, such as vehicle on fire, both for the casualty and the Marines conducting the evacuation. Casualty extract should proceed as quickly as possible to minimize this exposure. Additionally, moving a casualty to a place of cover and concealment or to a CCP increases survivability.

B.5.2.4 Conduct Casualty Evacuation Additional Insights

Slower movements lead to greater exposure time. A greater reliance on two-man carrying methods removes another Marine from providing local security during the movement. Integrated crews performed noticeably slower during the execution of this subtask.

B.5.3 Prepare Light Armored Vehicle for Combat Overview

MCWP 3-14.1, Light Armored Vehicle-25 Gunnery and Employment, states the following:

> On future battlefields, the tempo will be such that an LAV-25 crew must be prepared to move and to rapidly acquire and engage multiple targets…The LAV-25's speed and mobility, coupled with battle drills, increase the likelihood of opposing and allied forces becoming intermingled during combat operations. Survival in these situations depends on the crew's ability to effectively search for, detect, locate, identify, classify, confirm, and rapidly engage enemy targets. The LAV-25 crews must take advantage of the tactical situation and engage first. Speed and accuracy of an engagement depend on the degree of crew proficiency in target acquisition techniques and gunnery procedures.

This passage effectively illustrates the importance of speed and tempo not only in the conduct of gunnery, but also in the basic skills and preparatory tasks that lend themselves to effective gunnery procedures and correct target acquisition techniques. The same publication also details the conduct of the LAV Gunnery Skills Test (LGST), a basic skills test that drove the inclusion of many of the prep for combat subtasks. By the end of the experiment, most if not all volunteers were able to meet LGST requirements. Important to note, however, is that the LGST is a tool that ensures Gunners "meet minimum standards for training and safety before advancing into live fire." The times stipulated in the LGST should be viewed as safety minimums for training environments, not the expectation for gunnery standards in combat.

Prep for combat tasks were conducted as individual Gunner's events despite being crew level events. In an LAV crew, the Gunner's primary responsibility is the preparation, upkeep, and employment of the vehicle's turret and weapons systems, especially the M240 coaxially mounted machine gun and the M242 Main Gun. Therefore, the Gunner holds the critical billet in most prep for combat and live-fire tasks. The Gunner expends the most effort with little or no physical assistance from the other crewman.

The data collected for this task was analyzed in two ways. The first comparison looked at performance differences between the crews grouped by the gender of the Marine serving as Gunner and actively participating in the task. The second comparison looked for performance differences between the control group, low-density group, and high-density group.

B.5.4 Prepare the M240 Coax Overview

This subtask required the LAV-25 Gunner, without assistance, to remove, disassemble, reassemble and install the M240 coaxially mounted machine gun. Per MCWP 3-14.1, crewman have 15 minutes to complete this task during the LAV Gunnery Skills Test. Disassembly was achieved when the Gunner removed the bolt assembly from the weapon system. The ability to rapidly troubleshoot and correct issues within the turret is an absolutely critical skill for LAV Gunners. As the very basis of their role on a vehicle crew, a Gunner must be able to rapidly accomplish any task involving the LAV-25's weapons systems. When the Vehicle Commander and Driver are engaged in their primary duties, they often will not be able to lend assistance to the Gunner. Gunners were evaluated on the time required to disassemble and to reassemble the weapon system.

The data collected for this task was analyzed by critical billet and by integration level.

B.5.4.1 Prepare the M240 Coax by Critical Billet

See B.5.4 for task description.

B.5.4.1.1 Prepare the M240 Coax by Critical Billet Scatterplots

All data points shown in the scatter plots below were determined to be valid and used in the analysis and modeling. There were no outliers or potential influential points for this data set. The first plot shows the total elapsed time to execute the task of Prep M240 Coax. The second plot displays the elapsed time for disassembly of the M240, while the third plot displays the reassembly of the M240.

Figure B-4. Prep M240 Coax by Critical Billet

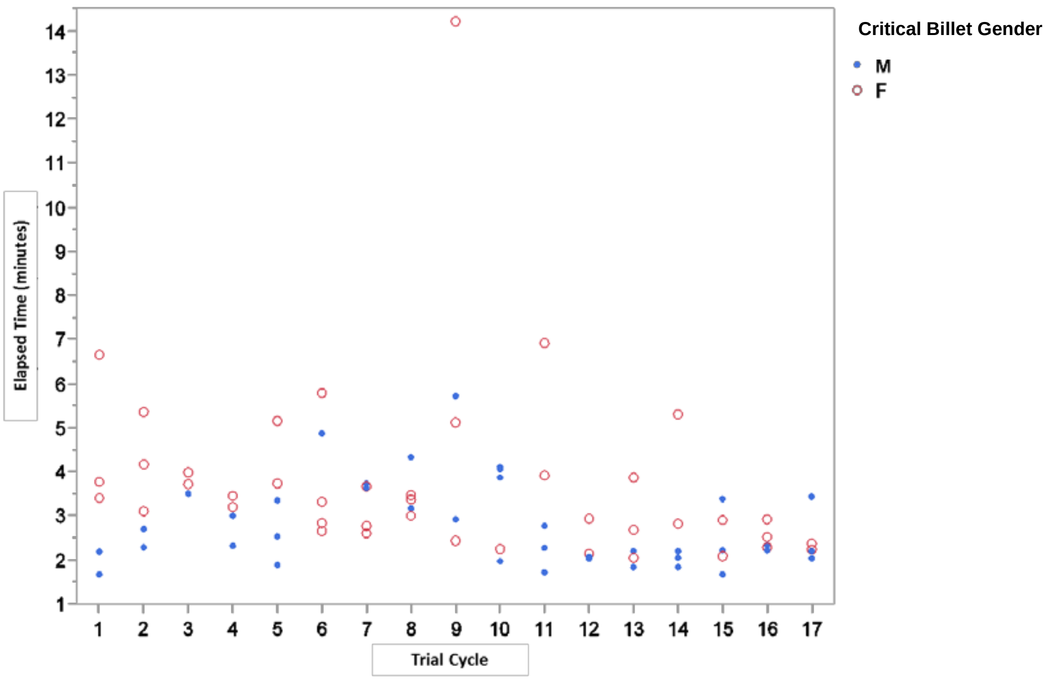


Figure B-5. Prep M240 Coax; Disassembly by Critical Billet

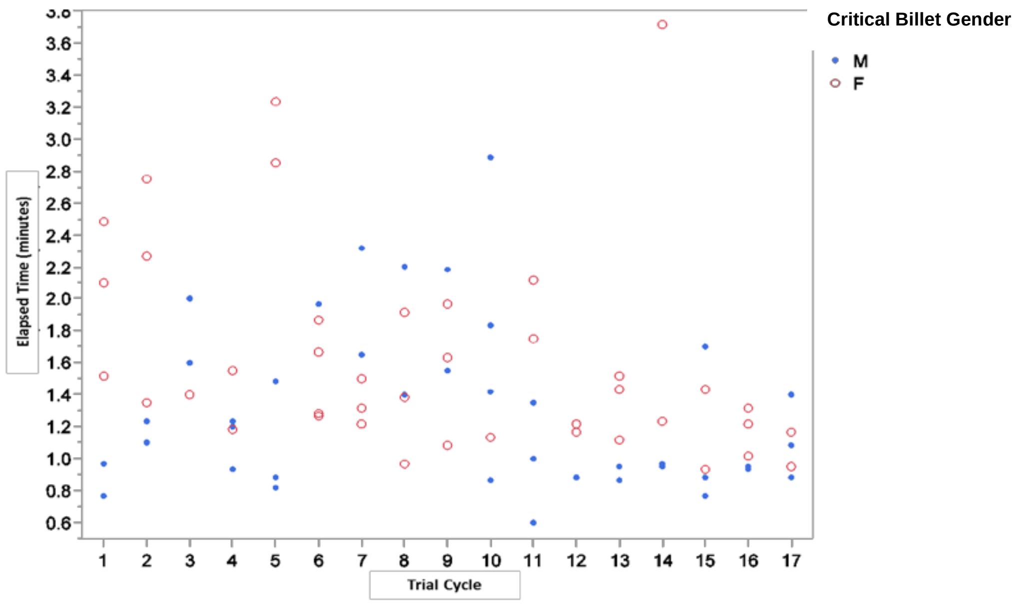
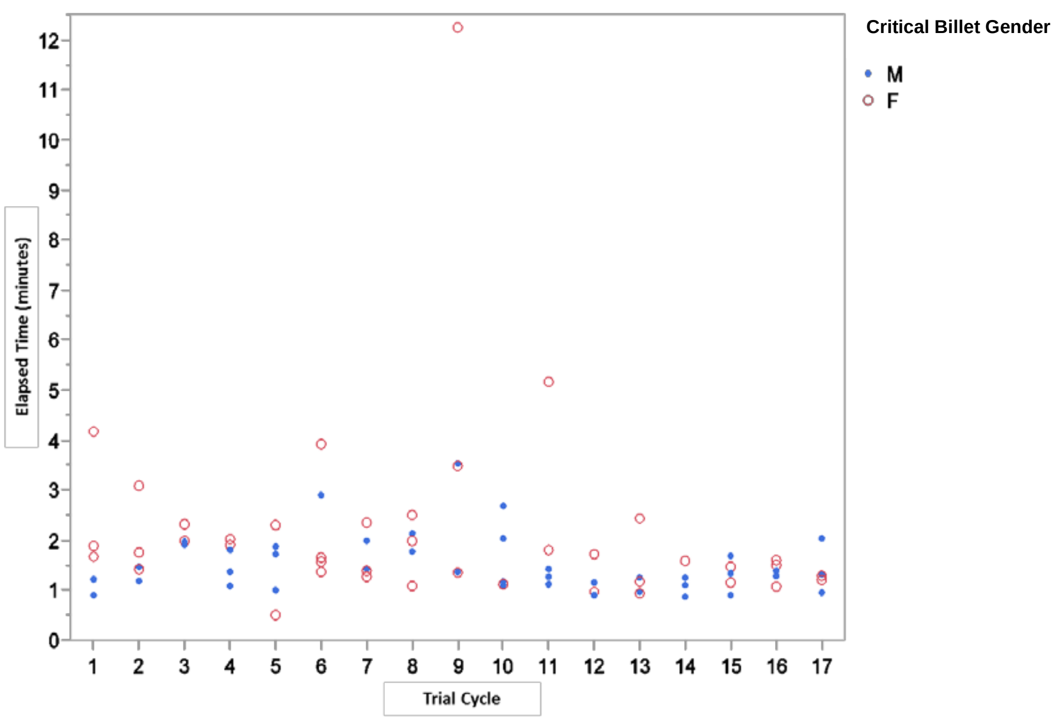


Figure B-6. Prep M240 Coax; Reassembly by Critical Billet



B.5.4.1.2 Prepare the M240 Coax by Critical Billet Data Tables and Analysis

The tables below summarize the results of the task, Prepare the M240 Coax. The data was analyzed by critical billet. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-4. Prep M240 Coax by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Prep M240 Coaxial Machine Gun (Gunner) (minutes) *	M	41	2.79	0.96
	F	42	3.74	2.05
Prep M240 Coaxial Machine Gun; Disassembly (Gunner) (minutes) *	M	41	1.29	0.51
	F	42	1.62	0.63
Prep M240 Coaxial Machine Gun; Re-assembly (Gunner) (minutes) *	M	41	1.51	0.57
	F	42	2.12	1.85

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-5. Prep M240 Coax by Critical Billet ANOVA and Welch's T-Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Prep M240 Coaxial Machine Gun (Gunner) (minutes)	7.33 (1, 59)	0.01††	F-M	0.95	33.92%	2.71 (59)	< 0.01*	< 0.01*	0.49	1.40	0.36	1.53
Prep M240 Coaxial Machine Gun; Disassembly (Gunner) (minutes)	7.12 (1, 81)	< 0.01*	F-M	0.34	26.14%	2.68 (78)	< 0.01*	< 0.01*	0.17	0.50	0.13	0.55
Prep M240 Coaxial Machine Gun; Re-assembly (Gunner) (minutes)	4.17 (1, 49)	0.05††	F-M	0.61	40.56%	2.04 (49)	0.02*	0.05*	0.22	1.00	0.11	1.11

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances).

For all tasks above, the normality assumption of ANOVA is satisfied since all samples are sufficiently large (n > 30). For total time to prep the M240, the standard deviation of the F group is more than twice that of the M group, thus we recommend using the robust ANOVA results presented above. For the disassembly portion of time, the equal variance assumption is satisfied and we proceed with presenting ANOVA results. For the assembly portion of time, the standard deviation of the F group is more than twice that of the M group, thus we recommend using the robust ANOVA results presented above.

For all tasks above, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). For the overall coax preparation time, the Male group had a mean time of 2.79 minutes, which was statistically significantly faster than the Female group that recorded a mean time of 3.74 minutes. The Male group's 33.92% faster mean time when compared to the Female group was statistically significant. The other two metrics shown in Table B-4 above, M240 Disassembly Time and M240 Reassembly Time also show statistical significance when comparing the Male group and Female group. The percentage differences equate to degradation in disassembly and reassembly times for the Female group when compared to the Male group. Additionally, the Female group for the three metrics above shows a greater variability as shown by the SD column. See Tables B-4 and B-5 above for detailed analytical results.

B.5.4.1.3 Prepare the M240 Coax by Critical Billet Contextual Comments

Disassembly and Assembly of the M240 Coax is an individual task solely executed by the Gunner. A LD integration crew could have either a male or a female serving as the Gunner. This has the potential to mask an individual Gunner's performance using integration level analysis alone. Therefore, it was necessary to explore the results of this subtask based on the gender of the Gunner.

B.5.4.1.4 Prepare the M240 Coax by Critical Billet Additional Insights

Integrated crews performed noticeably slower during the execution of this subtask. Per the actuarial tables in MCWP 3-14.1, engagements against dismounted or troop targets most often engaged with the M240 are measured in seconds. Should the coax jam or require remedial action, rapidly troubleshooting and restoring functionality must occur as rapidly as possible to ensure crew and vehicle survivability.

B.5.4.2 Prepare the M240 Coax by Critical Billet (Excluding Potential Influential Results)

See B.5.4 for description of task.

B.5.4.2.1 Prepare the M240 Coax by Critical Billet (Excluding Potential Influential Results) Scatterplots

Influential points are identified by a lightly colored solid black circle in the scatter plots below. These points were excluded for analysis. The first plot shows the total elapsed time to execute the task. The second plot displays the elapsed time only for the disassembly of the M240. The third plot shows the times recorded for the reassembly portion of the task.

Figure B-7. Prep M240 Coax by Critical Billet (Excluding)

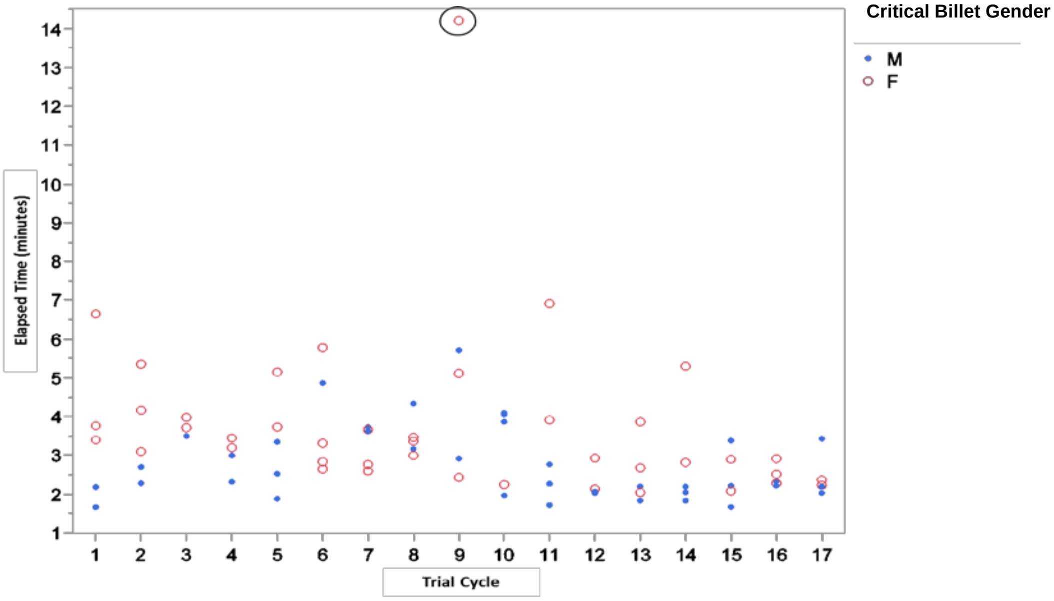


Figure B-8. Prep M240 Coax by Critical Billet; Disassembly (Excluding)

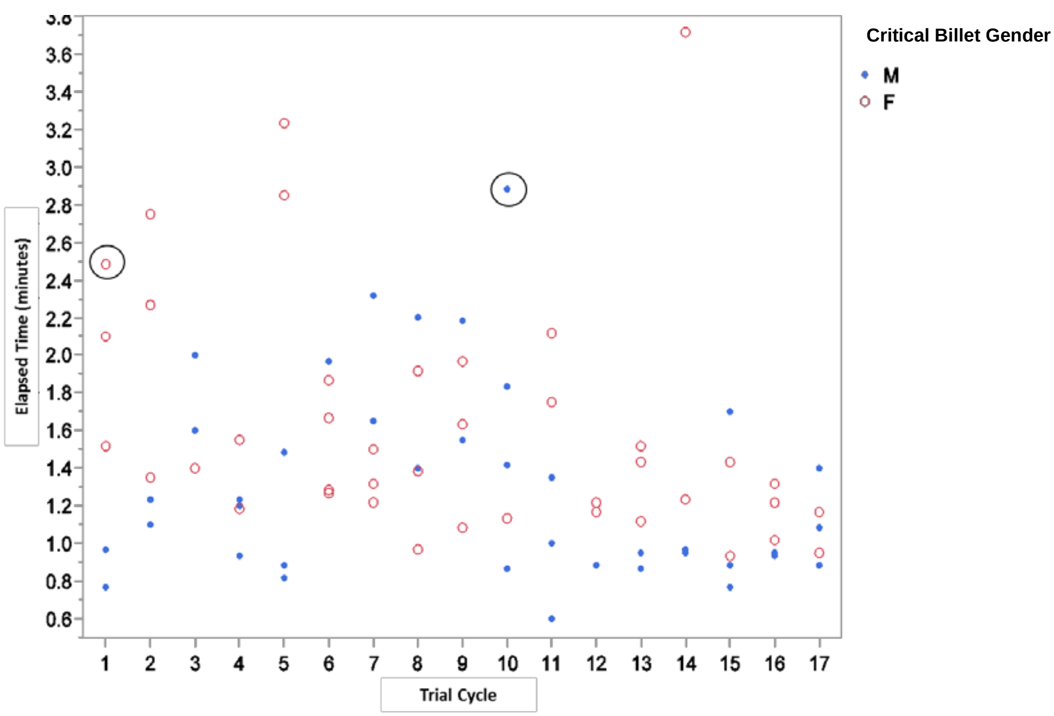
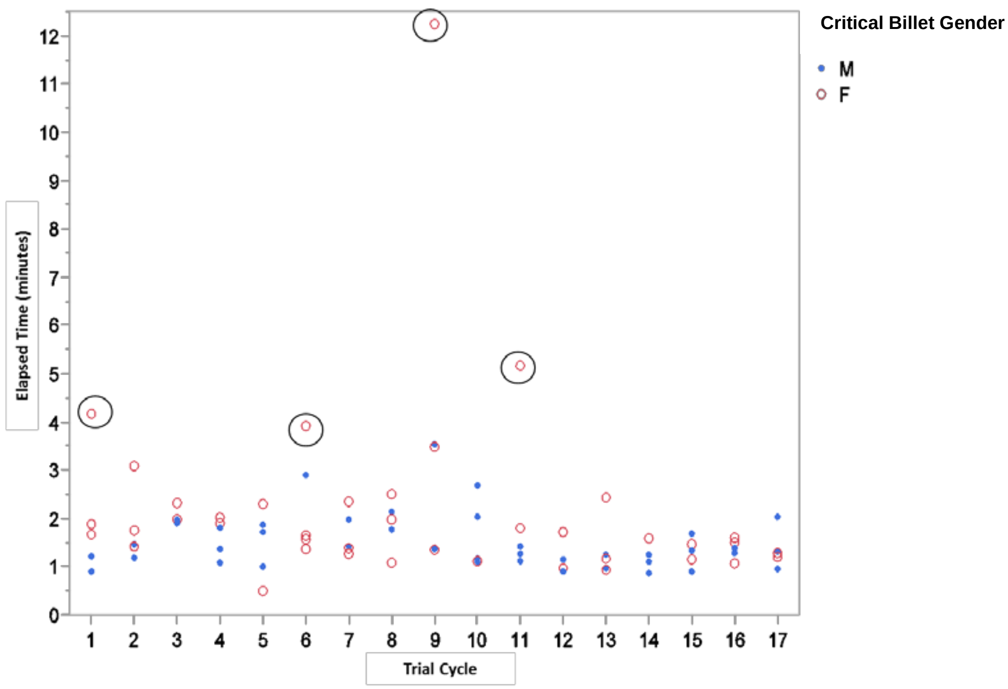


Figure B-9. **Prep M240 Coax by Critical Billet; Reassembly (Excluding)**



B.5.4.2.2 Prepare the M240 Coax by Critical Billet (Excluding Potential Influential Points) Tables and Analysis

The tables below summarize the results of the task, Prepare the M240 Coax. The results are presented by critical billet and exclude potential influential points. The first table compares means across metrics and integration levels that display critical billet by gender. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-6. Prep M240 Coax by Critical Billet (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Prep M240 Coaxial Machine Gun (Gunner) **[excluding potential influential points]** (minutes) *	M	41	2.79	0.96
	F	41	3.48	1.22
Prep M240 Coaxial Machine Gun; Disassembly (Gunner) **[excluding potential influential points]** (minutes) *	M	40	1.25	0.45
	F	41	1.60	0.62
Prep M240 Coaxial Machine Gun; Reassembly (Gunner) **[excluding potential influential points]** (minutes)	M	40	1.51	0.57
	F	37	1.67	0.60

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-7. Prep M240 Coax by Critical Billet (Excluding) ANOVA and Welch's T-Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Prep M240 Coaxial Machine Gun (Gunner) **[excluding potential influential points]** (minutes)	8.14 (1, 80)	< 0.01*	F-M	0.69	24.76%	2.85 (76)	< 0.01*	< 0.01*	0.38	1.00	0.29	1.09
Prep M240 Coaxial Machine Gun; Disassembly (Gunner) **[excluding potential influential points]** (minutes)	8.66 (1, 79)	< 0.01*	F-M	0.36	28.50%	2.95 (73)	< 0.01*	< 0.01*	0.20	0.51	0.15	0.56
Prep M240 Coaxial Machine Gun; Re-assembly **[excluding potential influential points]** (Gunner) (minutes)	1.52 (1, 77)	0.222	F-M	0.16	10.78%	1.23 (76)	0.11	0.22	-0.01	0.33	-0.06	0.38

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

For all tasks above, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). Additionally, for all tasks, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

For all tasks above, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). For the overall coax preparation time, the Male group had a mean time of 2.79 minutes, which was statistically significantly faster than the Female group that recorded a mean time of 3.48 minutes. The Male group's 24.76% faster mean time when compared to the Female group was statistically significant. M240 Disassembly Time shown in Table B-7 above also shows statistical significance when comparing the Male group and Female group. The percentage differences equate to degradation in disassembly time for the Female Group when compared to the Male group. Additionally, the Female group for the three metrics above shows a greater variability as shown by the SD column. See Table B-6 and Table B-7 for detailed analytical results.

B.5.4.2.3 Prepare the M240 Coax by Critical Billet (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.4.3 Prepare the M240 Coax by Integration Level

See B.5.4 for description of task.

B.5.4.3.1 Prepare the M240 Coax by Integration Level Scatterplots

The three scatter plots below display Prep M240 coax results by integration level for the total elapsed time of the task, followed by hack time means for disassembly and reassembly.

Figure B-10. Prep M240 Coax by Integration Level

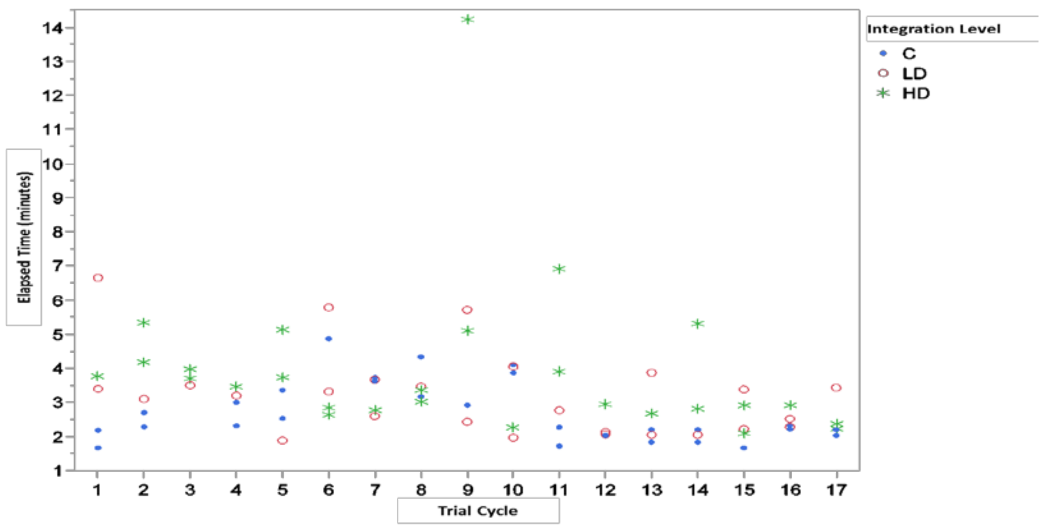


Figure B-11. Prep M240 Coax; Disassembly by Integration Level

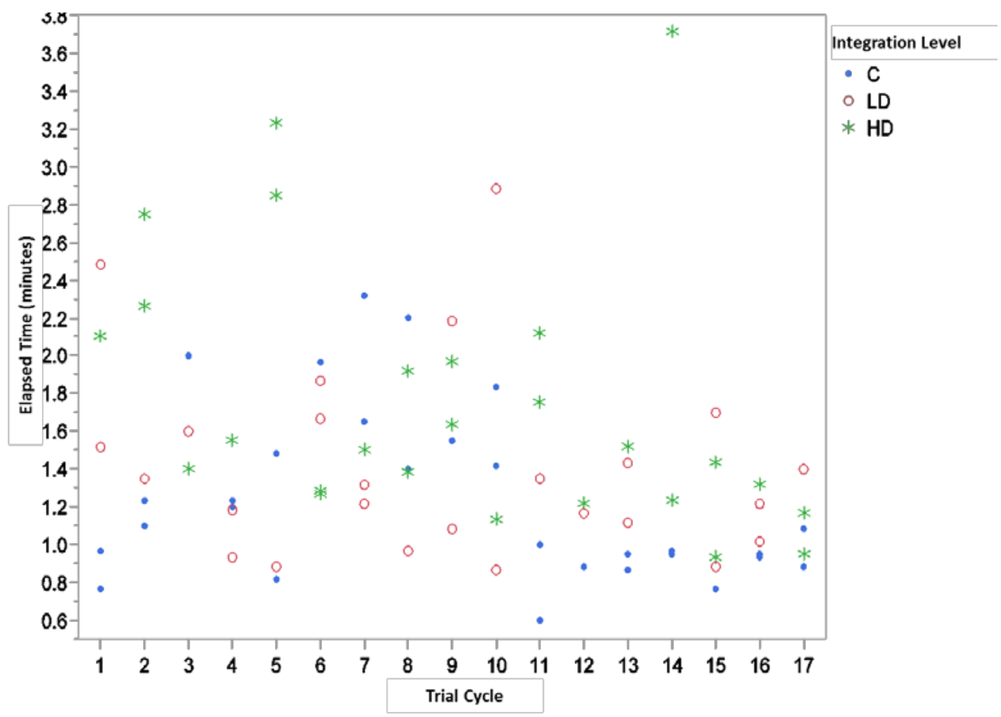
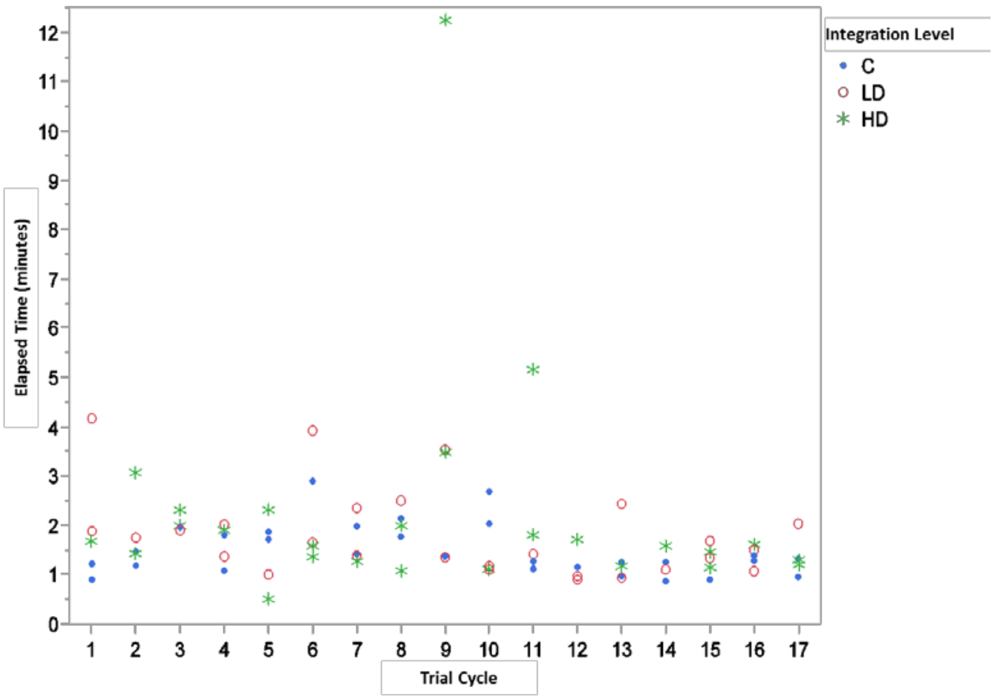


Figure B-12. Prep M240 Coax; Reassembly by Integration Level



B.5.4.3.2 Prepare the M240 Coax by Integration Level Tables and Analysis

The results in the tables below are presented by integration level and include potential influential points. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance, indicated by p-values less than 0.01, along with their percentage differences.

Table B-8. Prep M240 Coax by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Prep M240 Coaxial Machine Gun (minutes) *	C	29	2.73	0.89
	LD	27	3.18	1.23
	HD	27	3.95	2.36
Prep M240 Coaxial Machine Gun; Disassembly (minutes) *	C	29	1.24	0.46
	LD	27	1.39	0.50
	HD	27	1.76	0.70
Prep M240 Coaxial Machine Gun; Re-assembly (minutes)	C	29	1.49	0.52
	LD	27	1.79	0.88
	HD	27	2.19	2.21

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-9. Prep M240 Coax by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Prep M240 Coaxial Machine Gun (minutes)	3.70 (2, 47)	0.03††	LD-C	0.45	16.43%	0.13	0.07	0.82	-0.04	0.93
			HD-C	1.22	44.63%	0.02††	0.59	1.85	0.40	2.04
			HD-LD	0.77	24.22%	0.14	0.10	1.44	-0.09	1.63
Prep M240 Coaxial Machine Gun; Disassembly (minutes)	6.42 (2, 80)	< 0.01*	LD-C	0.14	11.69%	0.60	-0.11	0.40	-0.17	0.46
			HD-C	0.52	42.10%	< 0.01*	0.26	0.78	0.21	0.83
			HD-LD	0.38	27.23%	0.04*	0.11	0.64	0.06	0.69
Prep M240 Coaxial Machine Gun; Re-assembly (minutes)	2.21 (2, 44)	0.12	LD-C	0.30	20.38%	0.13	0.05	0.56	-0.02	0.63
			HD-C	0.70	46.75%	0.12	0.12	1.27	-0.04	1.44
			HD-LD	0.39	21.90%	0.40	-0.21	0.99	-0.38	1.17

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

††Indicates statistical significance using a robust ANOVA (accounts for unequal variances). For comparisons utilizing robust ANOVA, Tukey confidence intervals have been replaced by individual Welch's T-tests with p-values compared to a Bonferroni corrected threshold.

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. For the overall time to prep the M240, the C, LD, and HD group data are all not normally distributed with Shapiro-Wilk Test p-values of less than 0.01. Additionally, since the standard deviation of the HD group is more than twice that of the C group, we recommend using the robust ANOVA results presented above. We proceed with presenting robust ANOVA results, since they are confirmed by a Kruskal-Wallis Test (p-value < 0.01).

For the disassembly portion of the time, the C, LD, and HD group data are all not normally distributed with Shapiro-Wilk Test p-values of less than 0.01. Group standard deviations are sufficiently similar to satisfy the equal variance assumption of ANOVA. Thus, we proceed with presenting ANOVA results, since they are confirmed by a Kruskal-Wallis Test (p-value < 0.01).

For the assembly portion of the time, the C, LD, and HD group data are all not normally distributed with Shapiro-Wilk Test p-values of less than 0.01. Additionally, since the standard deviation of the HD group is more than twice that of the C and LD groups, we recommend using the robust ANOVA results presented above. We proceed with presenting robust ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.24)

For the overall time required to disassemble and assemble the M240 the C group had a mean time of 2.73 minutes, which was statistically significantly faster than the HD groups that recorded a mean time of 3.95 minutes. The C group's 44.63% faster mean time when compared to the HD group was statistically significant. The Disassembly Time metric in Table B-9 also shows a statistical significance when comparing the HD group to the C group and the LD group. All HD and LD groups for the three metrics above show greater variability, as shown by the SD column. See Table B-8 and Table B-9 for detailed analytical results.

B.5.4.3.3 Prepare the M240 Coax by Integration Level Additional Insights

Observations when examined by integration level remain largely the same as by critical billet, although critical billet results in the low-density group are less apparent.

B.5.4.4 Prepare the M240 Coax by Integration Level (Excluding Potential Influential Points)

See B.5.4 for description of task.

B.5.4.4.1 Prepare the M240 Coax by Integration Level (Excluding Potential Influential Points) Scatterplots

The scatter plots below display the results of the various integration levels elapsed times to complete the task, as well as disassembly and reassembly of the M240 coax. Influential points in the following plots are identified by a lightly colored solid black circle.

Figure B-13. Prep M240 Coax by Integration Level (Excluding)

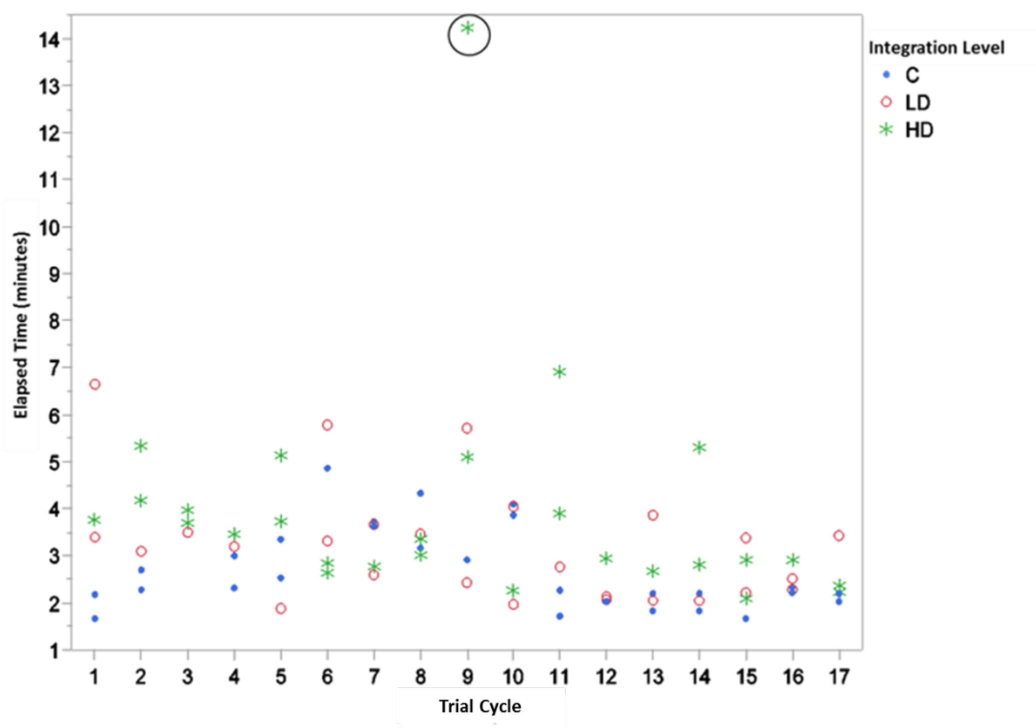


Figure B-14. Prep M240 Coax by Integration Level; Disassembly (Excluding)

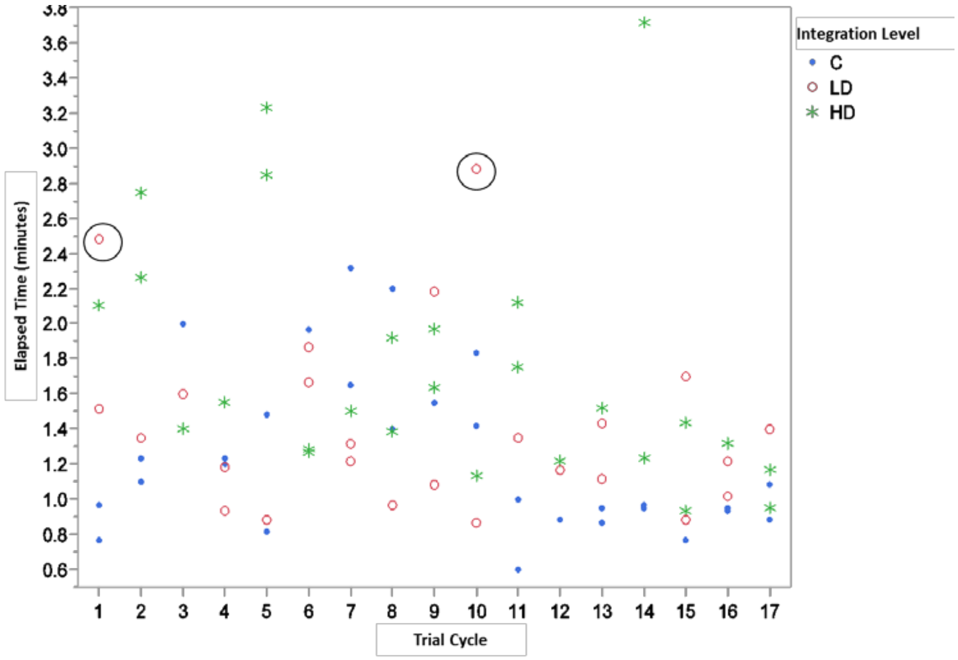
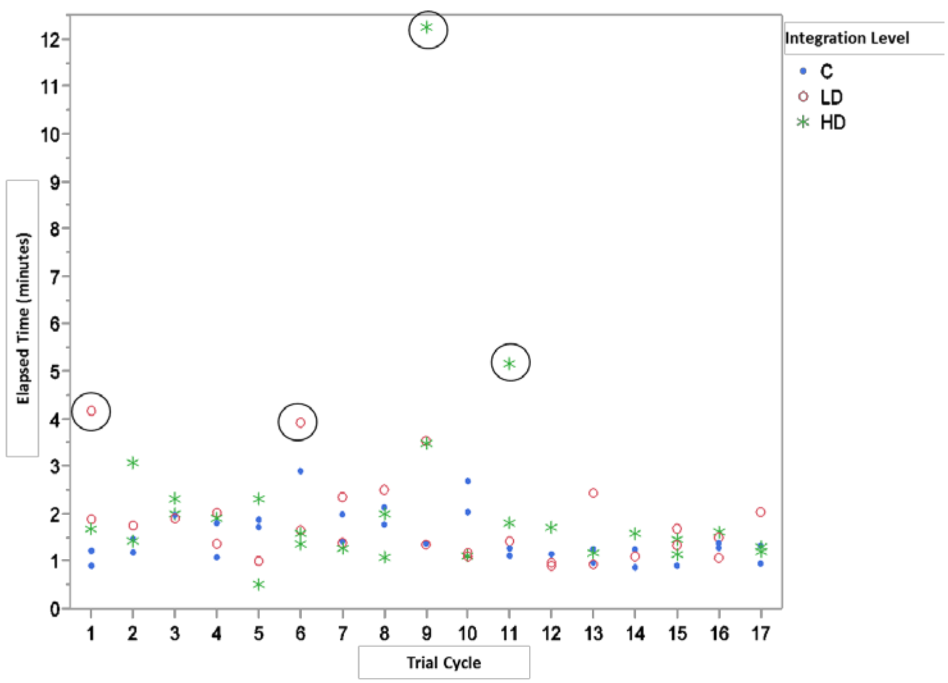


Figure B-15. Prep M240 Coax by Integration Level; Reassembly (Excluding)



B.5.4.4.2 Prepare the M240 Coax by Integration Level (Excluding Potential Influential Points) Data Tables and Analysis

The tables below summarize the results of the task, Prepare the M240 Coax. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-10. Prep M240 Coax by Integration Level (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Prep M240 Coaxial Machine Gun **[excluding potential influential points]** (minutes) *	C	29	2.73	0.89
	LD	27	3.18	1.23
	HD	26	3.55	1.19
Prep M240 Coaxial Machine Gun; Disassembly **[excluding potential influential points]** (minutes) *	C	29	1.24	0.46
	LD	25	1.28	0.33
	HD	27	1.76	0.70
Prep M240 Coaxial Machine Gun; Re-assembly **[excluding potential influential points]** (minutes)	C	29	1.49	0.52
	LD	25	1.61	0.62
	HD	25	1.66	0.63

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-11. Prep M240 Coax by Integration Level ANOVA and Tukey Test (Excluding)

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Prep M240 Coaxial Machine Gun **[excluding potential influential points]** (minutes)	3.82 (2, 79)	0.03*	LD-C	0.45	16.43%	0.29	-0.06	0.96	-0.17	1.06
			HD-C	0.82	30.16%	0.02*	0.31	1.34	0.20	1.45
			HD-LD	0.37	11.79%	0.44	-0.15	0.90	-0.26	1.01
Prep M240 Coaxial Machine Gun; Disassembly **[excluding potential influential points]** (minutes)	5.91 (2, 49)	< 0.01††	LD-C	0.04	3.31%	0.71	-0.10	0.18	-0.14	0.22
			HD-C	0.52	42.10%	< 0.01††	0.31	0.73	0.25	0.79
			HD-LD	0.48	37.54%	< 0.01††	0.29	0.68	0.23	0.73
Prep M240 Coaxial Machine Gun; Re-assembly **[excluding potential influential points]** (minutes)	0.63 (2, 76)	0.53	LD-C	0.12	8.30%	0.72	-0.16	0.40	-0.21	0.46
			HD-C	0.17	11.70%	0.53	-0.10	0.45	-0.16	0.51
			HD-LD	0.05	3.14%	0.95	-0.24	0.34	-0.30	0.40

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

††Indicates statistical significance using a robust ANOVA (accounts for unequal variances). For comparisons utilizing robust ANOVA, Tukey confidence intervals have been replaced by individual Welch's T-tests with p-values compared to a Bonferroni corrected threshold.

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. For the overall time to prep the M240, the C and LD groups are not normally distributed with Shapiro-Wilk Test p-values of less than 0.01, while the HD group is normally distributed with p-value of 0.01, respectively. Group standard deviations are sufficiently similar to satisfy the equal variance assumption of ANOVA. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.01).

For the disassembly portion of the time, the C and HD group data are not normally distributed with Shapiro-Wilk Test p-values of less than 0.01, while the LD group is normally distributed with p-value of 0.09. Additionally, since the standard deviation of the HD group is more than twice that of the LD group, we recommend using the robust ANOVA results presented above. We proceed with presenting robust ANOVA results, since they are confirmed by a Kruskal-Wallis Test (p-value < 0.01).

For the assembly portion of the time, the C group is not normally distributed with Shapiro-Wilk Test p-value of less than 0.01, while the LD and HD groups are normally distributed with p-values of 0.01 and 0.01, respectively. Group standard deviations are sufficiently similar to satisfy the equal variance assumption of ANOVA. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.46).

When excluding potential influential results, the overall time required to disassemble and assemble the M240 the C group had a mean time of 2.73 minutes, which was statistically significantly faster than the HD groups that recorded a mean time of 3.55 minutes. The C group’s 30.16% faster mean time when compared to the HD group was statistically significant. The Disassembly Time metric shown in Table B-11 also shows a statistical significance when comparing the HD group to the C group and the LD group. Additionally, the LD group shows less variability than the C group for disassembly time as shown by the SD column; all other HD and LD groups for the three metrics above show greater variability. See Table B-10 and Table B-11 for detailed analytical results.

B.5.4.4.3 Prepare the M240 Coax by Integration Level (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.5 Manually Traverse and Elevate the LAV-25 Turret Overview

This subtask required the LAV-25 Gunner, without assistance, to manually traverse, elevate, and depress the vehicle turret. The Gunner used the manual traverse and elevation controls within the turret to accomplish the subtask. In the event of power loss within the turret, a Gunner must be able to rapidly manipulate the turret to acquire and engage targets. This skill is regularly assessed with manual engagements during table qualification gunnery. Gunners were evaluated on the time required to traverse and to elevate the vehicle turret.

The data collected for this task was analyzed by critical billet and by integration level.

B.5.5.1 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet

See B.5.5 for task description.

B.5.5.1.1 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet Scatterplots

The following scatter plots depict the full data set for this particular task. The data points contained inside dark black circles were determined to be data collector or Toughbook errors and were excluded from the analysis. The first plot shows the full elapsed time for the critical billet displayed by the gender of the Marine executing the task. The second and third scatter plots focus on the traversing and elevate/depress aspects of the task, respectively.

Figure B-16. Manually Manipulate Turret/Main Gun by Critical Billet

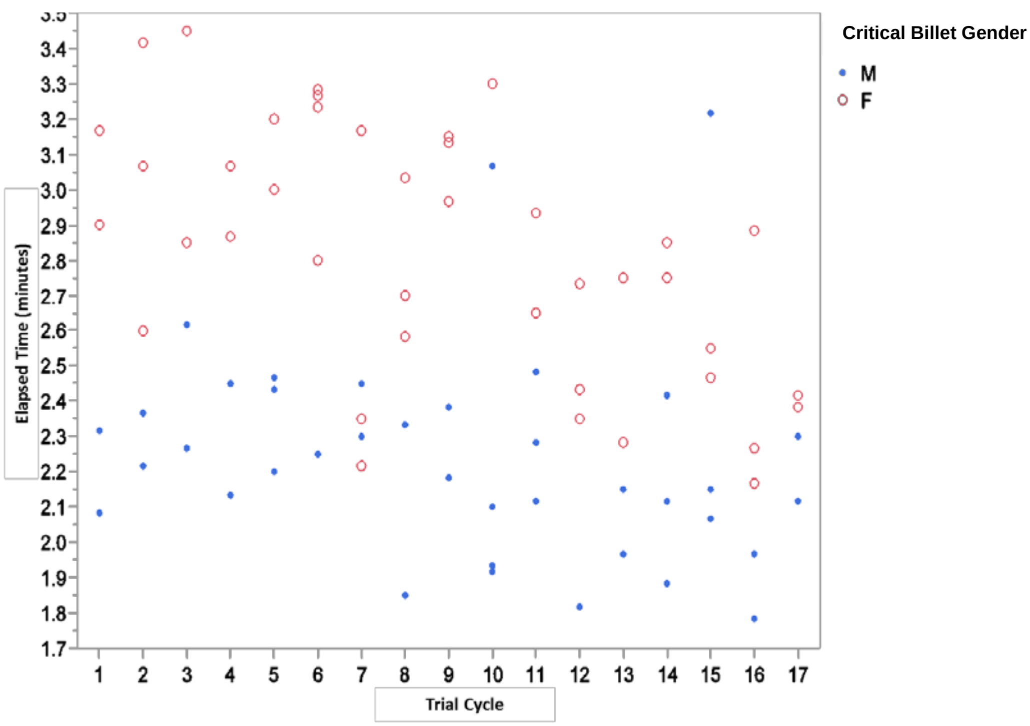


Figure B-17. Manually Manipulate Turret/Main Gun; Traverse by Critical Billet

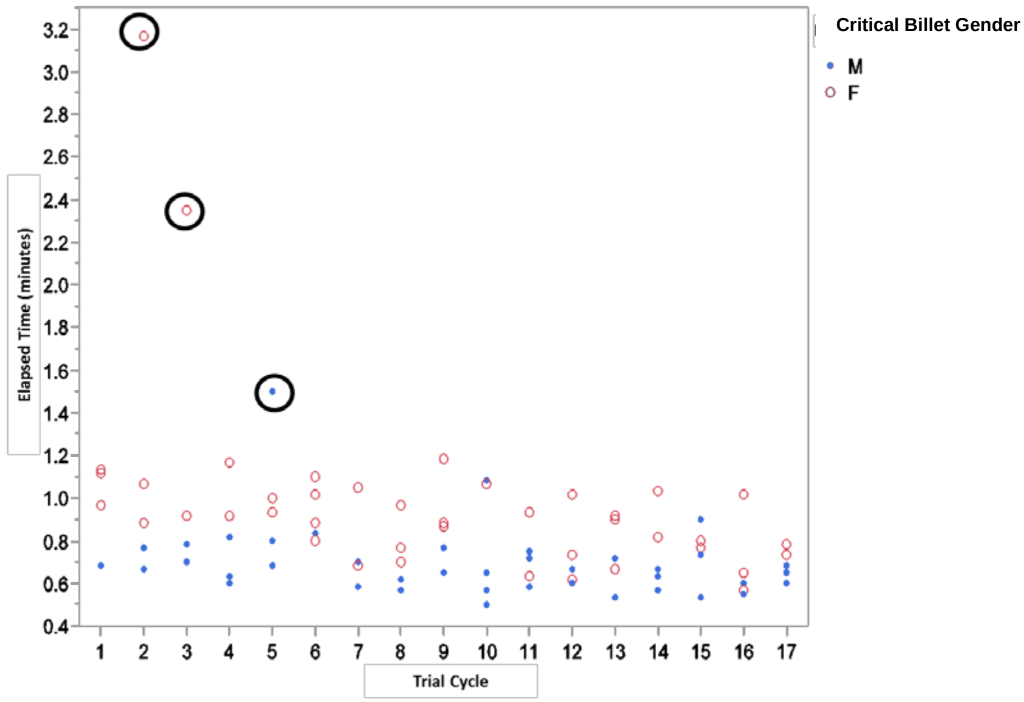
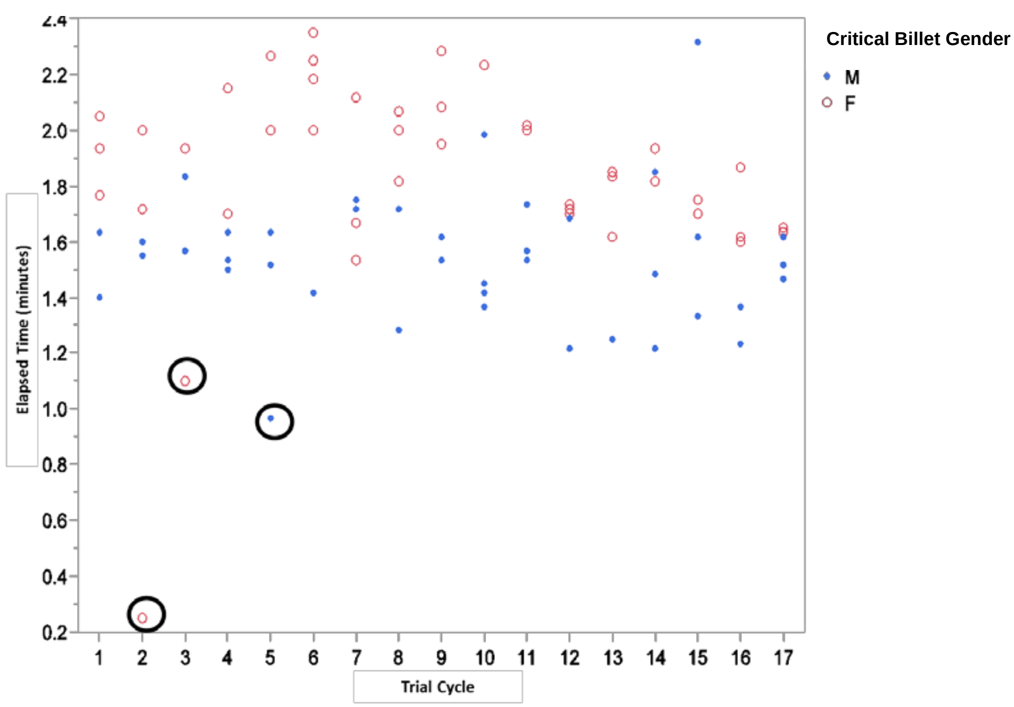


Figure B-18. Manually Manipulate Turret/Main Gun; Elevate/Depress by Critical Billet



B.5.5.1.2 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet Data Tables and Analysis

The tables below summarize the results of the task, Manually Traverse and Elevate the LAV-25 Turret. The results are displayed by critical billet. The first table compares means across metrics and critical billet by gender. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance, as depicted by p-values, along with their percentage differences.

Table B-12. Manually Manipulate Turret/Main Gun by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Manually Manipulate Turret/Main Gun (Gunner) (minutes) *	M	41	2.24	0.29
	F	43	2.82	0.35
Manually Manipulate Turret/Main Gun; Traverse Elements (Gunner) (minutes) *	M	40	0.68	0.11
	F	41	0.89	0.17
Manually Manipulate Turret/Main Gun; Elevate/Depress Main Gun (Gunner) (minutes) *	M	40	1.56	0.22
	F	41	1.90	0.22

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-13. Manually Manipulate Turret/Main Gun by Critical Billet ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Manually Manipulate Turret/Main Gun (Gunner) (minutes)	69.35 (1, 82)	< 0.01*	F-M	0.58	26.04 %	8.37 (80)	< 0.01*	< 0.01*	0.49	0.67	0.47	0.70
Manually Manipulate Turret/Main Gun; Traverse Elements (Gunner) (minutes)	44.94 (1, 79)	< 0.01*	F-M	0.21	31.21 %	6.73 (71)	< 0.01*	< 0.01*	0.17	0.25	0.16	0.26
Manually Manipulate Turret/Main Gun; Elevate/Depress Main Gun (Gunner) (minutes)	51.94 (1, 79)	< 0.01*	F-M	0.35	22.34 %	7.21 (79)	< 0.01*	< 0.01*	0.29	0.41	0.27	0.43

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

For all tasks above, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). Additionally, for all tasks, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For all tasks above, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). For the overall time required to traverse and elevate the turret, the Male group had a mean time of 2.24 minutes, which was statistically significantly faster than the Female group that recorded a mean time of 2.82 minutes. The Male group's 26.04% faster mean time when compared to the Female group was statistically significant. The other two metrics shown in Table B-13 above, both Traverse Time and Elevation Time also show statistical significance when comparing the Male group and Female group. The percentage differences equate to degradation in traverse and elevation times for Female Group when compared to the Male group. Additionally, the Female group for the overall time and traverse time metrics above shows a greater variability as shown by the SD column. See Table B-12 and Table B-13 for detailed analytical results.

B.5.5.1.3 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet Contextual Comments

This is an individual task solely executed by the Gunner. A LD integration crew could have either a male or a female serving as the Gunner. This has the potential to mask an individual Gunner's performance using integration level analysis alone. Therefore, it was necessary to explore the results of this subtask based on the gender of the Gunner. Per MCWP 3-14.1, a Gunner's ability to conduct Manual Engagements is regularly assessed during pre-qualification and qualification tables of gunnery. The speed at which a Gunner is able to manipulate the turret to acquire and engage a target directly affects the effectiveness, survivability, and lethality of an LAV-25 crew.

B.5.5.1.4 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet Additional Insights

Integrated crews performed noticeably slower during the execution of this subtask. Should a vehicle lose power or should the requirement arise to operate without power (i.e., night operations on a screen line) a Gunner must be able to rapidly obtain a firing solution and quickly engage targets, and slower traverse and elevation times can degrade the survivability of the crew and vehicle.

B.5.5.2 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet (Excluding Potential Influential Points)

See B.5.5 for task description.

B.5.5.2.1 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet (Excluding Potential Influential Points) Scatterplots

The following scatter plots depict the full data set for this particular task. The data points contained inside dark black circles were determined to be data errors and were excluded from the analysis. The data points contained inside the lightly colored black circles were considered to be potential influential points and were excluded from the analysis. The first plot shows the full elapsed time to complete the task for the critical billet, displayed by the gender of the Marine executing the task. The second scatter plot displays the traverse aspect of the task.

Figure B-19. Manually Manipulate Turret/Main Gun by Critical Billet (Excluding)

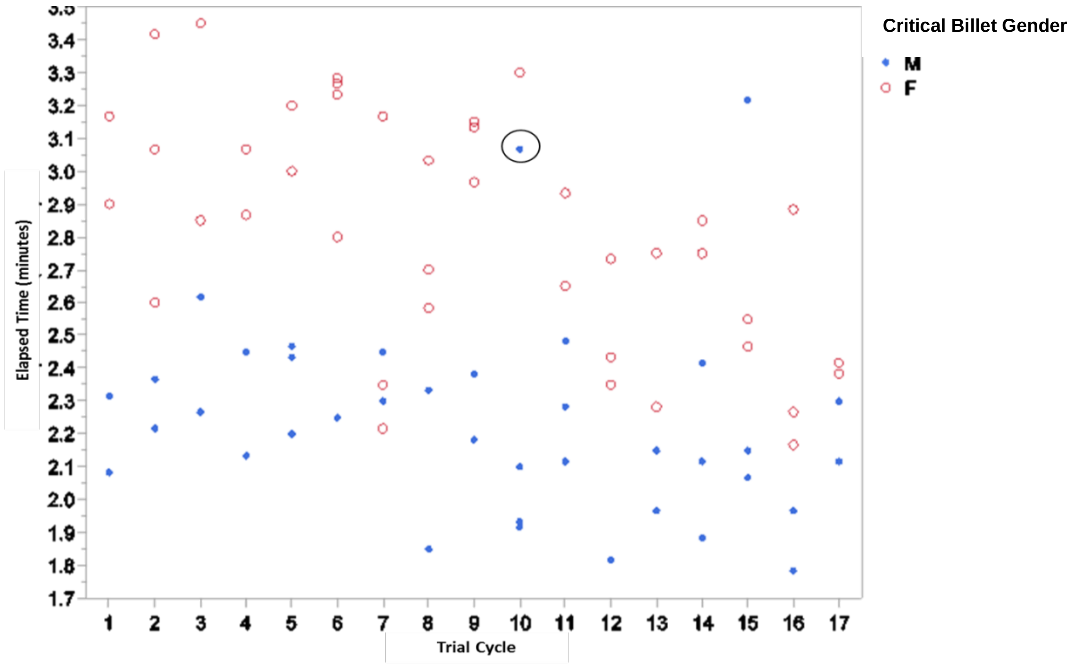
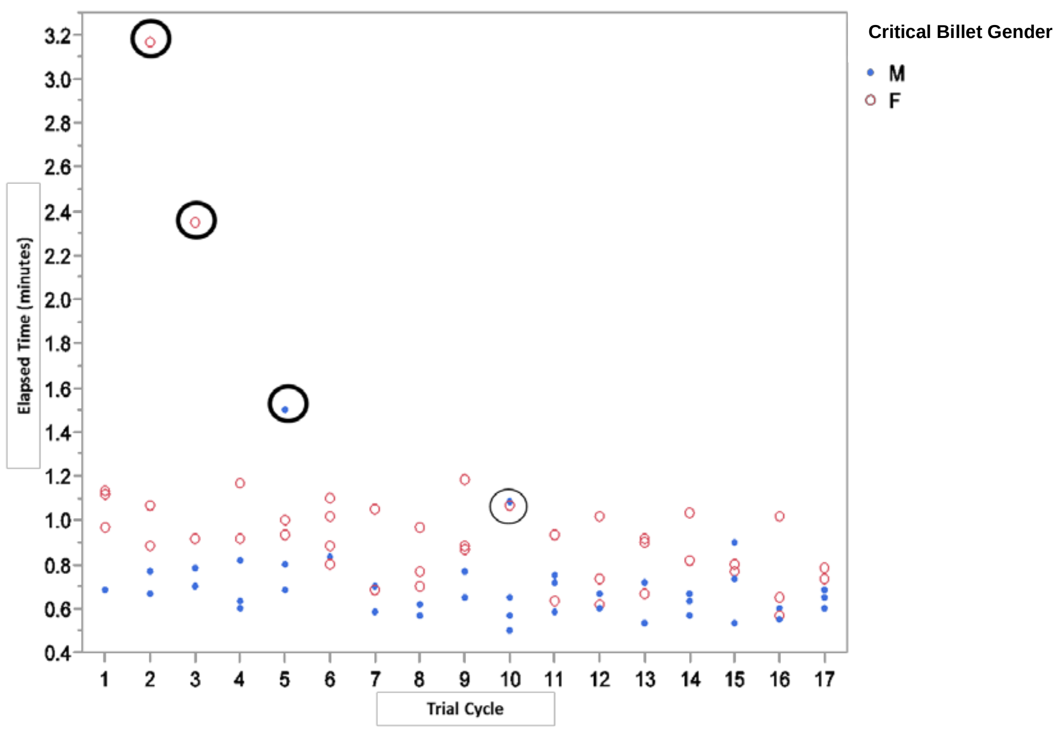


Figure B-20. Manually Manipulate Turret/Main Gun; Traverse by Critical Billet (Excluding)



B.5.5.2.2 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet (Excluding Potential Influential Points) Data Tables and Analysis

The tables below summarize the results of the task, Manually Traverse and Elevate the LAV-25 Turret. The results are presented by critical billet and exclude potential influential points. The first table compares means across metrics and critical billets by gender. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-14. Manually Manipulate Turret/Main Gun Results by Critical Billet (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Manually Manipulate Turret/Main Gun (Gunner) **[excluding potential influential points]** (minutes) *	M	40	2.22	0.26
	F	43	2.82	0.35
Manually Manipulate Turret/Main Gun; Traverse Elements (Gunner) **[excluding potential influential points]** (minutes) *	M	39	0.66	0.09
	F	41	0.89	0.17

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-15. Manually Manipulate Turret/Main Gun by Critical Billet (Excluding) ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Manually Manipulate Turret/Main Gun (Gunner) **[excluding potential influential points]** (minutes)	79.15 (1, 81)	< 0.01*	F-M	0.60	27.22%	8.99 (77)	< 0.01*	< 0.01*)	0.52	0.69	0.49	0.72
Manually Manipulate Turret/Main Gun; Traverse Elements (Gunner) **[excluding potential influential points]** (minutes)	54.1 (1, 78)	< 0.01*	F-M	0.22	33.27%	7.45 (63)	< 0.01*	< 0.01*	0.18	0.26	0.17	0.27

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test..

For all tasks above, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). Additionally, for all tasks, group SDs are sufficiently similar to satisfy the equal variance assumption for ANOVA. For all tasks above, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). For the overall time excluding potential influential points required to traverse and elevate the turret, the Male group had a mean time of 2.22 minutes, which was statistically significantly faster than the Female group that recorded a mean time of 2.82 minutes. The Male group's 27.22% faster mean time when compared to the Female group was statistically significant. The traverse time metric shown in Table B-15 also shows statistical significance when comparing the Male group and Female group. Elevation elapsed time results did not contain potential influential points and are excluded. Additionally, the Female group for the overall time and traverse time metrics above shows a greater variability as shown by the SD column. See Table B-14 and Table B-15 for detailed analytical results.

B.5.5.2.3 Manually Traverse and Elevate the LAV-25 Turret by Critical Billet (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.5.3 Manually Traverse and Elevate the LAV-25 Turret by Integration Level

See B.5.5 for task description.

B.5.5.3.1 Manually Traverse and Elevate the LAV-25 Turret by Integration Level Scatterplots

The scatter plots below display the results of the various integration level elapsed times to complete the task.

Figure B-21. Manually Manipulate Turret/Main Gun by Integration Level

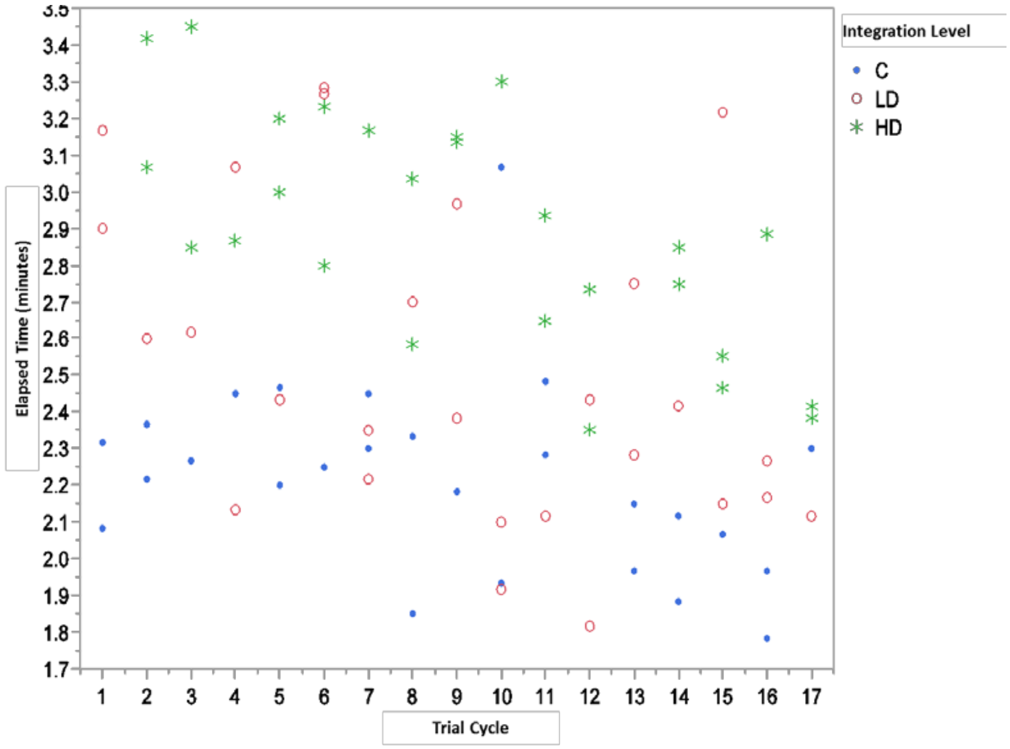


Figure B-22. Manually Manipulate Turret/Main Gun; Traverse by Integration Level

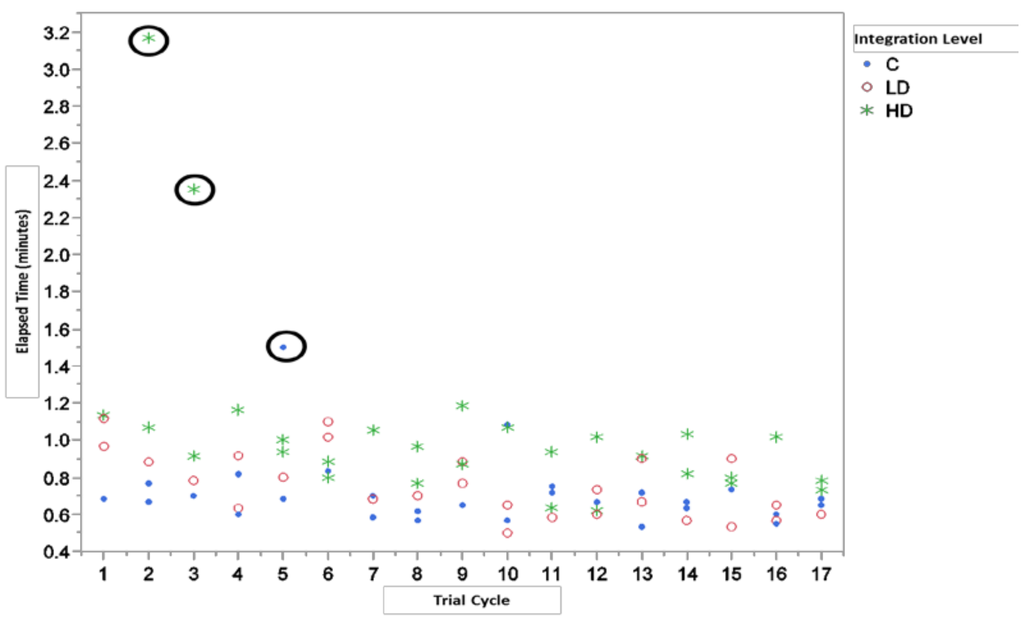
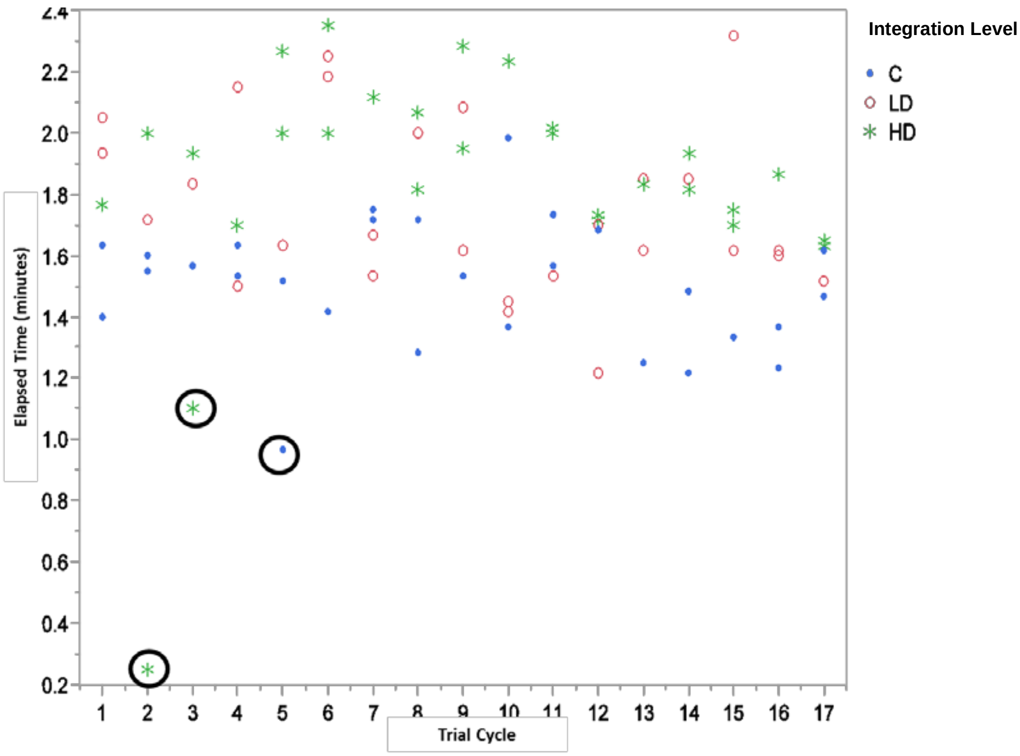


Figure B-23. Manually Manipulate Turret/Main Gun; Elevate/Depress by Integration Level



B.5.5.3.2 Manually Traverse and Elevate the LAV-25 Turret by Integration Level Data Tables and Analysis

The tables below summarize the results of the task, Manually Traverse and Elevate the LAV-25 Turret by integration level. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-16. Manually Manipulate Turret/Main Gun by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Manually Manipulate Turret/Main Gun (minutes) *	C	29	2.22	0.25
	LD	27	2.51	0.43
	HD	28	2.89	0.31
Manually Manipulate Turret/Main Gun; Traverse Elements (minutes) *	C	28	0.68	0.11
	LD	27	0.75	0.17
	HD	26	0.92	0.15
Manually Manipulate Turret/Main Gun; Elevate/Depress Main Gun (minutes) *	C	28	1.53	0.18
	LD	27	1.76	0.28
	HD	26	1.93	0.20

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test

Table B-17. Manually Manipulate Turret/Main Gun by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Manually Manipulate Turret/Main Gun (minutes)	28.4 (2, 81)	< 0.01*	LD-C	0.29	13.25%	< 0.01*	0.14	0.45	0.11	0.48
			HD-C	0.67	30.19%	< 0.01*	0.52	0.82	0.48	0.86
			HD-LD	0.38	14.96%	< 0.01*	0.22	0.53	0.19	0.56
Manually Manipulate Turret/Main Gun; Traverse Elements (minutes)	17.87 (2, 78)	< 0.01*	LD-C	0.07	10.67%	0.17	0.00	0.14	-0.01	0.16
			HD-C	0.24	34.57%	< 0.01*	0.17	0.31	0.15	0.32
			HD-LD	0.16	21.59%	< 0.01*	0.09	0.23	0.08	0.25
Manually Manipulate Turret/Main Gun; Elevate/Depress Main Gun (minutes)	21.45 (2, 78)	< 0.01*	LD-C	0.23	15.06%	< 0.01*	0.12	0.34	0.10	0.36
			HD-C	0.40	26.24%	< 0.01*	0.29	0.51	0.27	0.53
			HD-LD	0.17	9.72%	0.02*	0.06	0.28	0.04	0.30

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. For all tasks above, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the overall time to manipulate main gun, the C, LD, and HD groups are all normally distributed with Shapiro-Wilk Test p-values of 0.03, 0.07, and 0.74, respectively. For the traverse elements of manipulate main gun, the C group is not normally distributed with Shapiro-Wilk Test p-values of less than 0.01, while the LD and HD groups are normally distributed with p-values of 0.12 and 0.73, respectively. We proceed with presenting ANOVA results, since they are confirmed by a Kruskal-Wallis Test (p-value < 0.01). For the elevate/depress portion of manipulate main gun, the C, LD, and HD groups are all normally distributed with Shapiro-Wilk Test p-values of 0.61, 0.25, and 0.18, respectively.

For the overall time required to manually manipulate the turret, the C group had a mean time of 2.22 minutes, which was statistically significantly faster than the HD groups that recorded a mean time of 2.89 minutes. The C group's 30.19% faster mean time when compared to the HD group was statistically significant. The two additional metrics traverse time and elevation time shown in Table B-16 above also shows a statistical significance when comparing the HD group to the C group and the LD group. The 34.57% and 26.24% percentage differences equate to degradation in traverse time and elevation time, respectively, for HD groups when compared to the C group. Additionally, HD and LD groups for the three metrics above show greater variability than the C group as shown by the SD column. See Table B-16 and Table B-17 for detailed analytical results.

B.5.5.3.3 Manually Traverse and Elevate the LAV-25 Turret by Integration Level Additional Insights

Observations when examined by integration level remain largely the same as by critical billet, although the results in the low-density integration group are less apparent.

B.5.5.4 Manually Traverse and Elevate the LAV-25 Turret by Integration Level (Excluding Potential Influential Points)

See B.5.5 for task description.

B.5.5.4.1 Manually Traverse and Elevate the LAV-25 Turret by Integration Level (Excluding Potential Influential Points) Scatterplots

The scatter plots below display the results of integration level elapsed times to complete the Manually Manipulate Turret/Main Gun task. The data points contained inside dark black circles were determined to be data errors and were excluded from the analysis. Lightly colored black circles were determined to be potential influential points and were excluded from the analysis.

Figure B-24. Manually Manipulate Turret/Main Gun by Integration Level (Excluding)

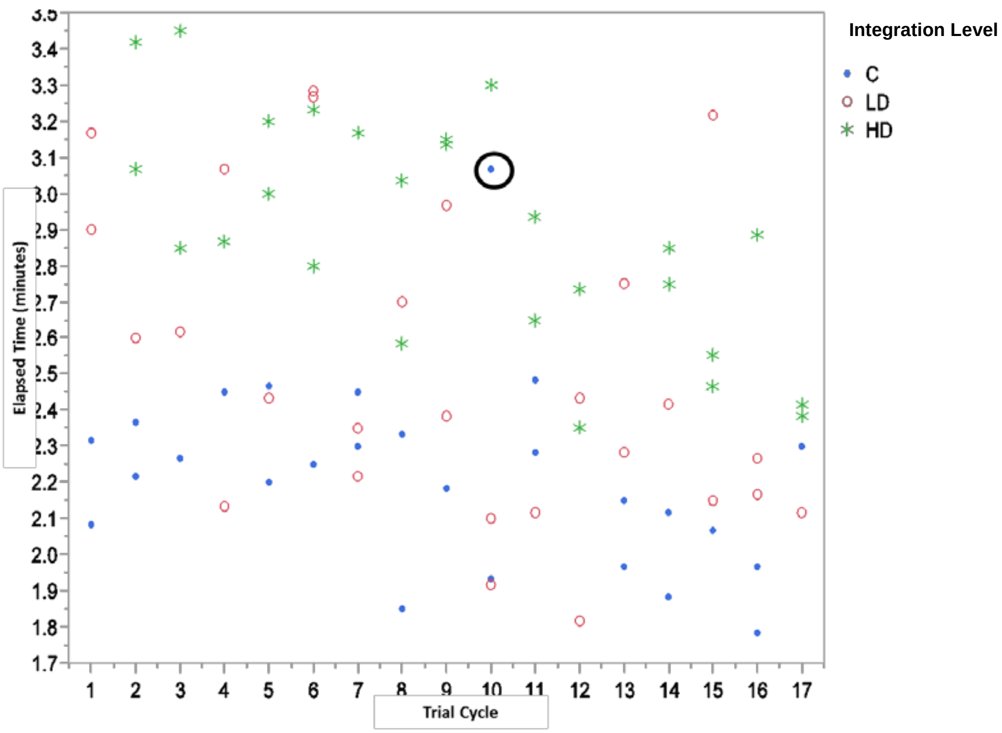
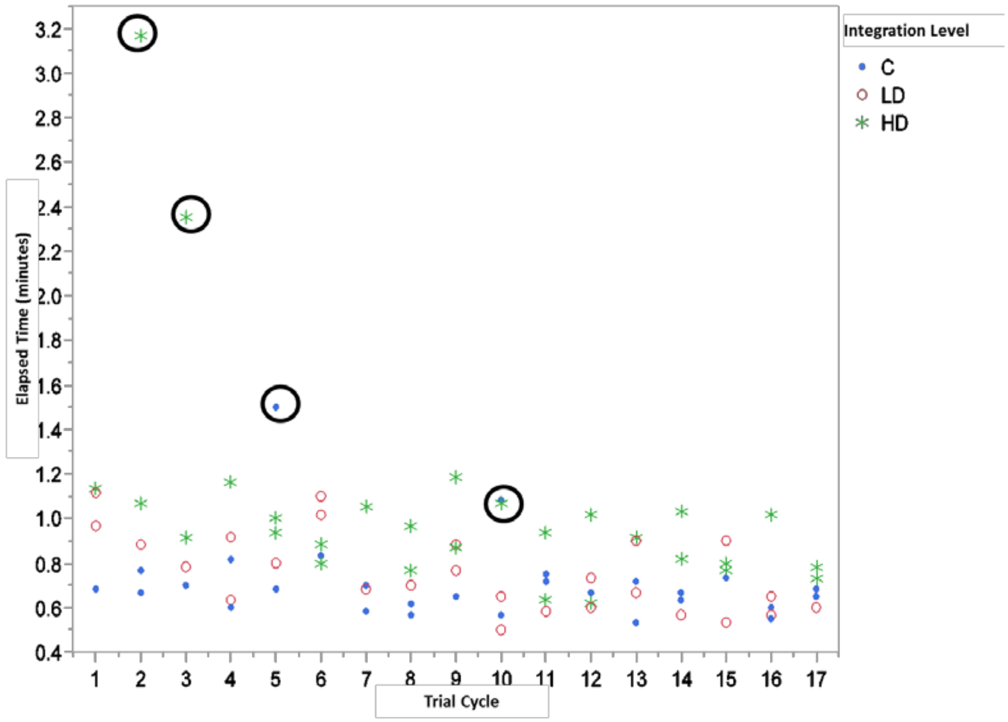


Figure B-25. Manually Manipulate Turret/Main Gun; Traverse by Integration Level (Excluding)



B.5.5.4.2 Manually Traverse and Elevate the LAV-25 Turret by Integration Level (Excluding Potential Influential Points) Data Table and Analysis

The tables below summarize the results of the Manually Manipulate Turret/Main Gun task by integration level. Table B-18 compares means across metrics and integration levels. Table B-19 presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-18. Manually Manipulate Turret/Main Gun by Integration Level (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Manually Manipulate Turret/Main Gun **[excluding potential influential points]** (minutes) *	C	28	2.19	0.19
	LD	27	2.51	0.43
	HD	28	2.89	0.31
Manually Manipulate Turret/Main Gun; Traverse Elements **[excluding potential influential points]** (minutes) *	C	27	0.67	0.08
	LD	27	0.75	0.17
	HD	26	0.92	0.15

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-19. Manually Manipulate Turret/Main Gun by Integration Level (Excluding) ANOVA and Welch's T-Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Manually Manipulate Turret/Main Gun **[excluding potential influential points]** (minutes)	52.37 (2, 48)	< 0.01††	LD-C	0.32	14.82%	< 0.01††	0.21	0.44	0.17	0.48
			HD-C	0.70	31.99%	< 0.01††	0.61	0.79	0.58	0.82
			HD-LD	0.38	14.96%	< 0.01††	0.24	0.51	0.21	0.55
Manually Manipulate Turret/Main Gun; Traverse Elements **[excluding influential points]** (minutes)	27.99 (2, 44)	< 0.01††	LD-C	0.09	13.14%	0.02††	0.04	0.14	0.03	0.15
			HD-C	0.25	37.57%	< 0.01††	0.21	0.29	0.19	0.31
			HD-LD	0.16	21.59%	< 0.01††	0.10	0.22	0.09	0.24

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances). For comparisons utilizing robust ANOVA, Tukey confidence intervals have been replaced by individual Welch's T-tests with p-values compared to a Bonferroni corrected threshold.

Statistical analysis included ANOVA and multiple comparisons using Tukey tests. For the overall time to manipulate main gun, the C, LD, and HD groups are all normally distributed with Shapiro-Wilk Test p-values of 0.40, 0.07, and 0.74, respectively. Since the standard deviation of the LD group is more than twice that of the C group, we recommend using the robust ANOVA results presented above.

For the traverse elements of manipulate main gun, the C, LD, and HD groups are all normally distributed with Shapiro-Wilk Test p-values of 0.75, 0.13, and 0.73, respectively. Since the standard deviation of the LD group is more than twice that of the C group, we recommend using the robust ANOVA results presented above.

When excluding potential influential points for the overall time required to manually manipulate the turret, the C group had a mean time of 2.19 minutes, which was statistically significantly faster than the HD groups that recorded a mean time of 2.89 minutes. The C group's 31.99% faster mean time when compared to the HD group was statistically significant. The additional traverse time metric shown in Table B-19 above also shows a statistical significance when comparing the HD group to the C group and the LD group. The 37.57% percentage difference equates to degradation in traverse time for HD groups when compared to the C group. The elevation time metric contained no potential influential points and is excluded. Additionally, HD and LD groups for both metrics above show greater variability than the C group as shown by the SD column. See Table B-18 and Table B-19 for detailed analytical results

B.5.5.4.3 Manually Traverse and Elevate the LAV-25 Turret by Integration Level (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.6 Prep the M242 Main Gun Overview

This subtask required the LAV-25 Gunner, without assistance, to remove, disassemble, reassemble and install the M242 Bushmaster Cannon, weighing approximately 263 lb. If, after 30 minutes, the Gunner was unable to perform this task, the Driver was then able to assist with the process. Based on the LAV Gunnery Skills Test standards listed in MCWP 3-14.1, a Gunner has 30 minutes to disassemble and reassemble the main gun. Proficient Gunners are expected to far exceed this standard. Disassembly was achieved when the Gunner disassembled the bolt and track assembly and removed the barrel from the weapon system. In many situations, the other members of the crew will be unavailable to assist the Gunner in manipulation and troubleshooting of the main gun. A Gunner's ability to rapidly troubleshoot, identify, and rectify any issues or a malfunction with the primary weapon system of the platform is absolutely vital to the effectiveness and survivability of the vehicle crew. Gunners were evaluated on the time required to disassemble and reassemble the weapon system.

The data collected for this task was analyzed both by critical billet and by integration level.

B.5.6.1 Prep the M242 Main Gun by Critical Billet

Disassembly and reassembly of the M242 Main Gun is an individual task solely executed by the Gunner. Only after the Gunner exceeded 30 minutes was any assistance allowed for this task. As the low-density integration crew could have either a male or female serving as the Gunner, this group masks an individual Gunner's performance. It is therefore beneficial to explore the results of this subtask based on the gender of the Gunner.

B.5.6.1.1 Prep the M242 Main Gun by Critical Billet Scatterplots

The scatter plots below display the results to complete the Prep M242 Main Gun by Gun task by critical billet. The data points contained inside dark black circles were determined to be data errors and were excluded from the analysis.

Figure B-26. Prep M242 Main Gun by Critical Billet

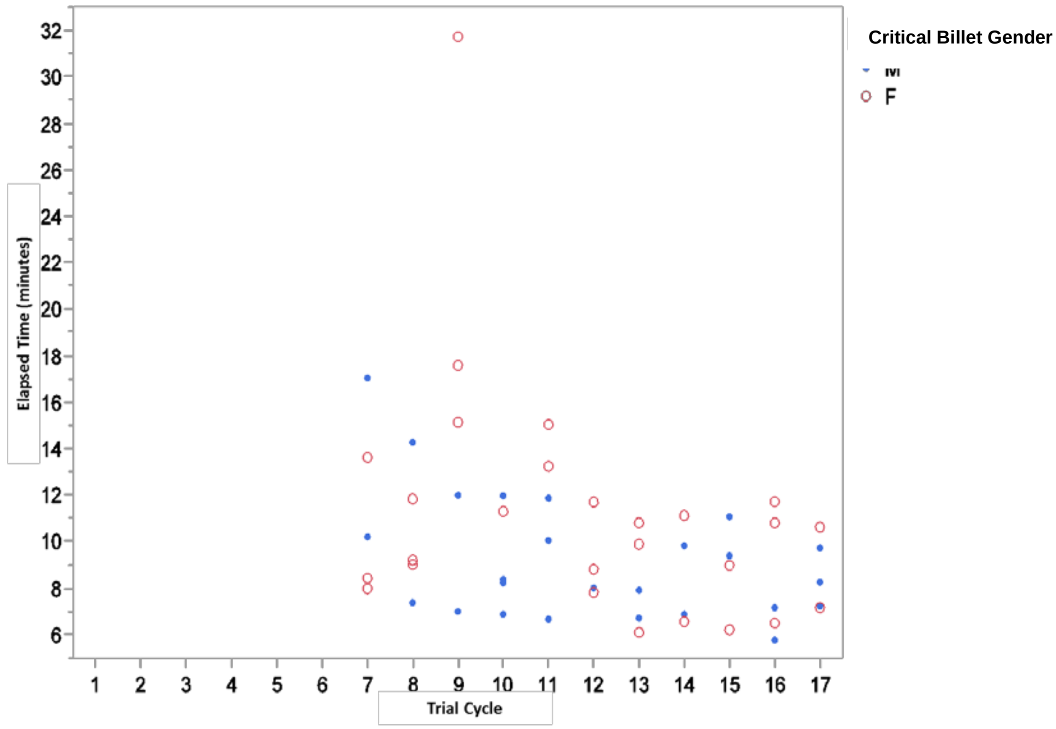


Figure B-27. Prep M242 Main Gun; Disassembly by Critical Billet

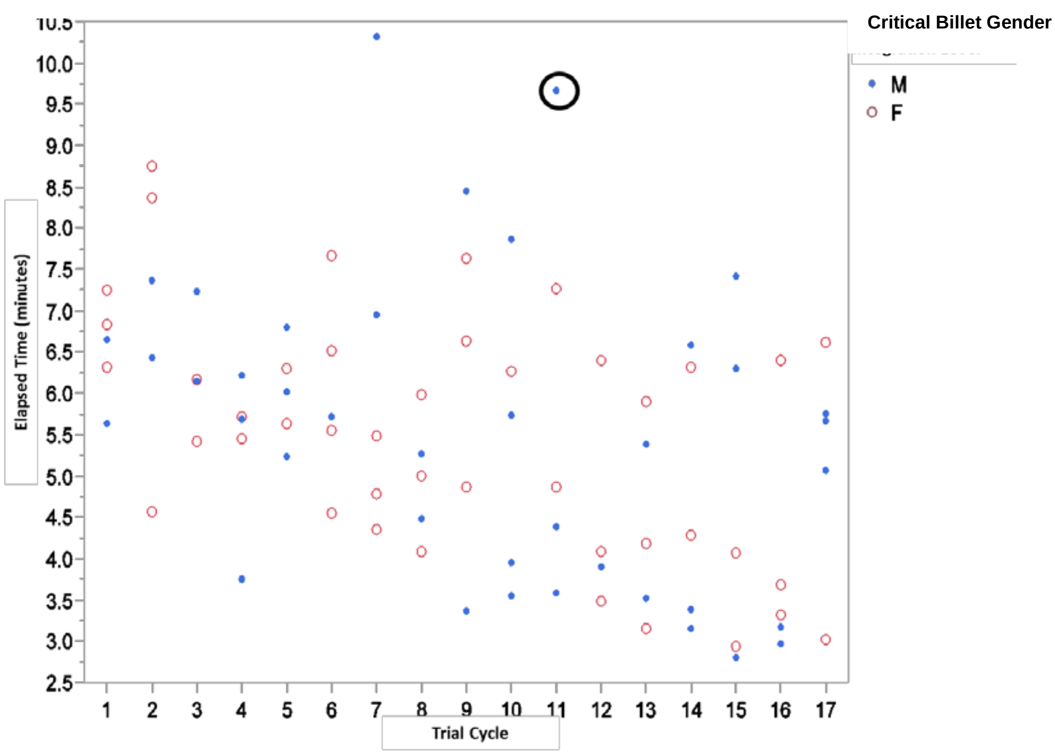
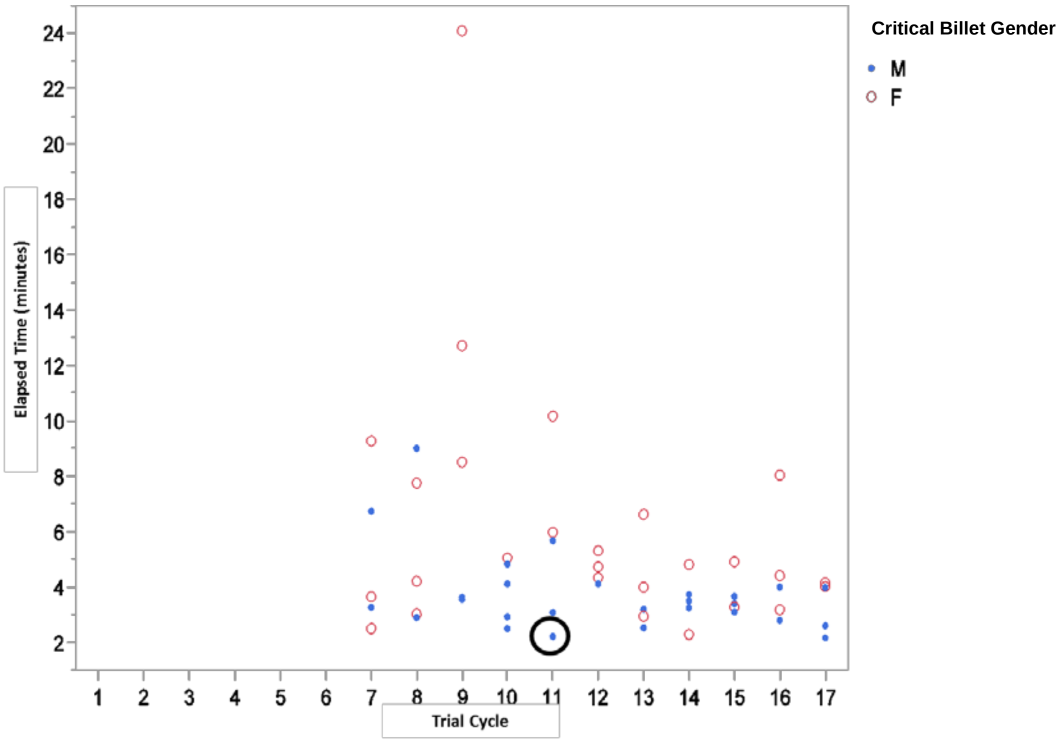


Figure B-28. Prep M242 Main Gun; Reassembly by Critical Billet



B.5.6.1.2 Prep the M242 Main Gun by Critical Billet Data Table and Analysis

The tables below summarize the results of the task by critical billet. Table B-20 compares means across metrics and integration levels. Table B-21 presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-20. Prep M242 Main Gun by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Prep M242 Main Gun (Gunner) (minutes) *	M	27	8.99	2.68
	F	27	11.06	5.06
Prep M242 Main Gun; Disassembly (Gunner) (minutes)	M	39	5.43	1.72
	F	43	5.49	1.46
Prep M242 Main Gun; Re-assembly (Gunner) (minutes) *	M	26	3.76	1.46
	F	27	6.06	4.41

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-21. Prep M242 Main Gun by Critical Billet ANOVA

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Prep M242 Main Gun (Gunner) (minutes)	3.54 (1, 52)	0.07*	F-M	2.07	23.08%	1.88 (40)	0.03*	0.07*	0.64	3.51	0.22	3.93
Prep M242 Main Gun; Disassembly (Gunner) (minutes)	0.03 (1, 80)	0.87	F-M	0.06	1.09%	0.17 (75)	0.43	0.87	-0.40	0.52	-0.53	0.65
Prep M242 Main Gun; Re-assembly (Gunner) (minutes)	6.57 (1, 32)	0.02††	F-M	2.30	61.01%	2.56 (32)	< 0.01*	0.02*	1.12	3.47	0.78	3.81

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Indicates statistical significance using a robust ANOVA (accounts for unequal variances).

For the overall time to prep the M242, the M and F groups are not normally distributed with Shapiro-Wilk p-values of less than 0.01. We proceed with presenting ANOVA results since they are confirmed by a Mann-Whitney Test (p-value = 0.08). For disassembly time of M242, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). For the reassembly time of the M242, the M and F groups are not normally distributed with Shapiro-Wilk p-values of less than 0.01. Since the standard deviation of the F group is more than twice that of the M group, we recommend using robust ANOVA presented above. We proceed with presenting ANOVA results, since they are confirmed by a Mann-Whitney Test (p-value < 0.01).

For the overall time required for disassembly and reassembly, the Male group had a mean time of 8.99 minutes, which was statistically significantly faster than the Female group that recorded a mean time of 11.06 minutes. The Male group's 23.08% faster mean time when compared to the Female group was statistically significant. The reassembly time metric also shows statistical significance when comparing the Male group to the Female group. The percentage differences equate to degradation in reassembly time for the Female Group when compared to the Male group. Additionally, the Female group above shows a greater variability for the overall time and reassembly time metrics as shown by the SD column; the Female group shows less variability for the disassembly time metric. See Table B-20 and Table B-21 for detailed analytical results.

B.5.6.1.3 Prep the M242 Main Gun by Critical Billet Additional Insights

As the primary weapon system of the platform, upkeep and employment of the 25mm cannon is a Gunner's central role. When observing breakout times, males and females performed nearly identically during disassembly of the weapon system. During reassembly, there was a large disparity between male and female gunners.

Reassembly requires the manipulation of heavy components by a single individual in a confined space. Longer times to ensure primary weapon system functionality have the potential to impede operations, especially when the LAV-25 is operating in a direct fire or support by fire role.

B.5.6.2 Prep the M242 Main Gun by Critical Billet (Excluding Potential Influential Points)

See Section B.5.6 for task description.

B.5.6.2.1 Prep the M242 Main Gun by Critical Billet (Excluding Potential Influential Points) Scatterplots

The scatter plots below display the results to complete the Prep M242 Main Gun task by critical billet. The data points contained inside dark black circles were determined to be data errors and were excluded from the analysis.

Figure B-29. Prep M242 Main Gun by Critical Billet (Excluding)

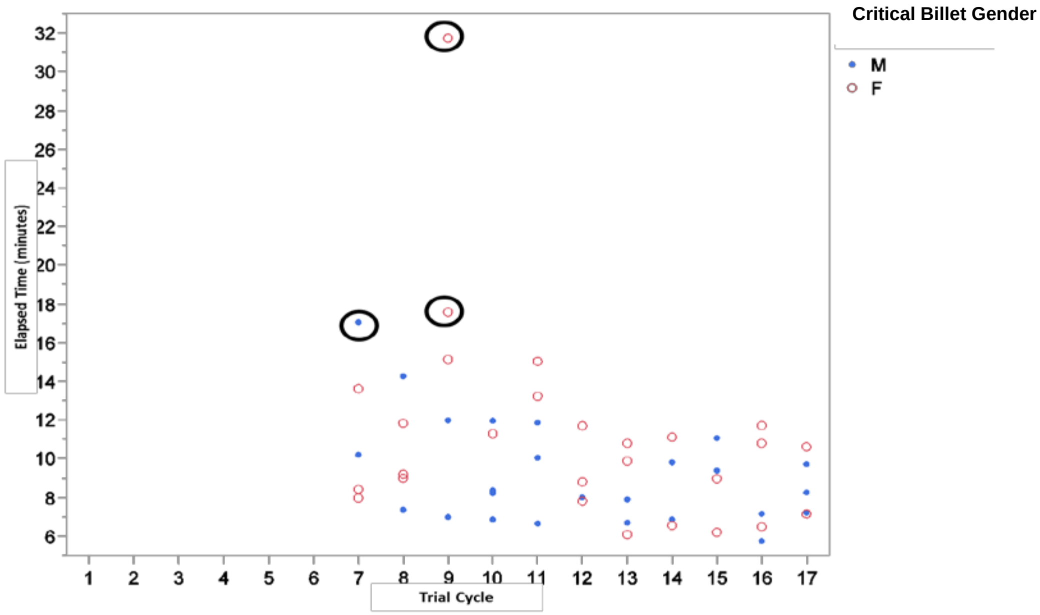


Figure B-30. Prep M242 Main Gun; Disassembly by Critical Billet (Excluding)

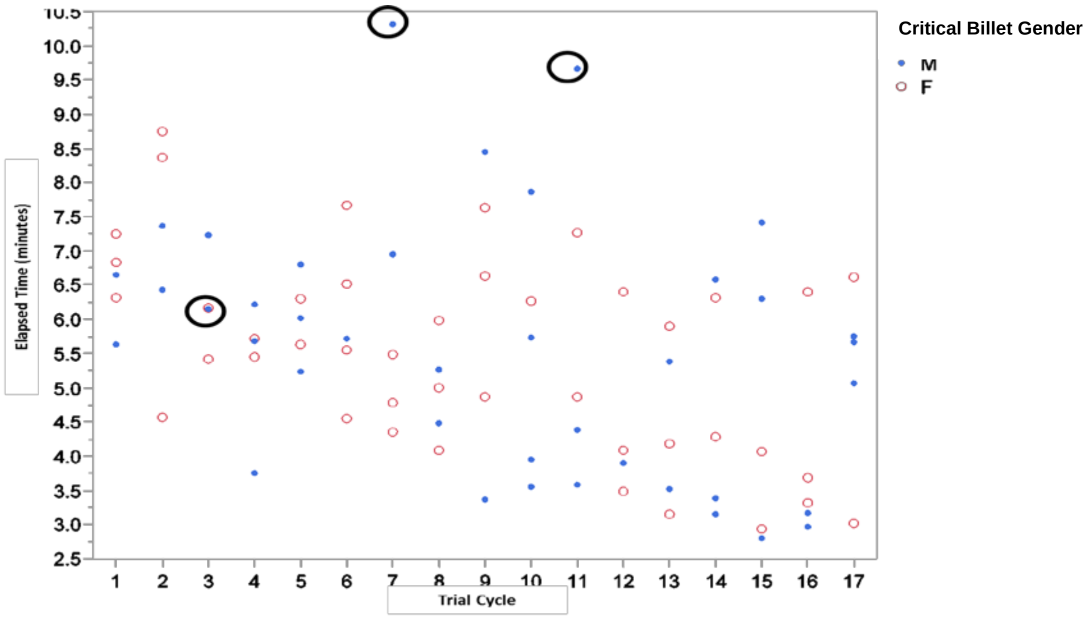
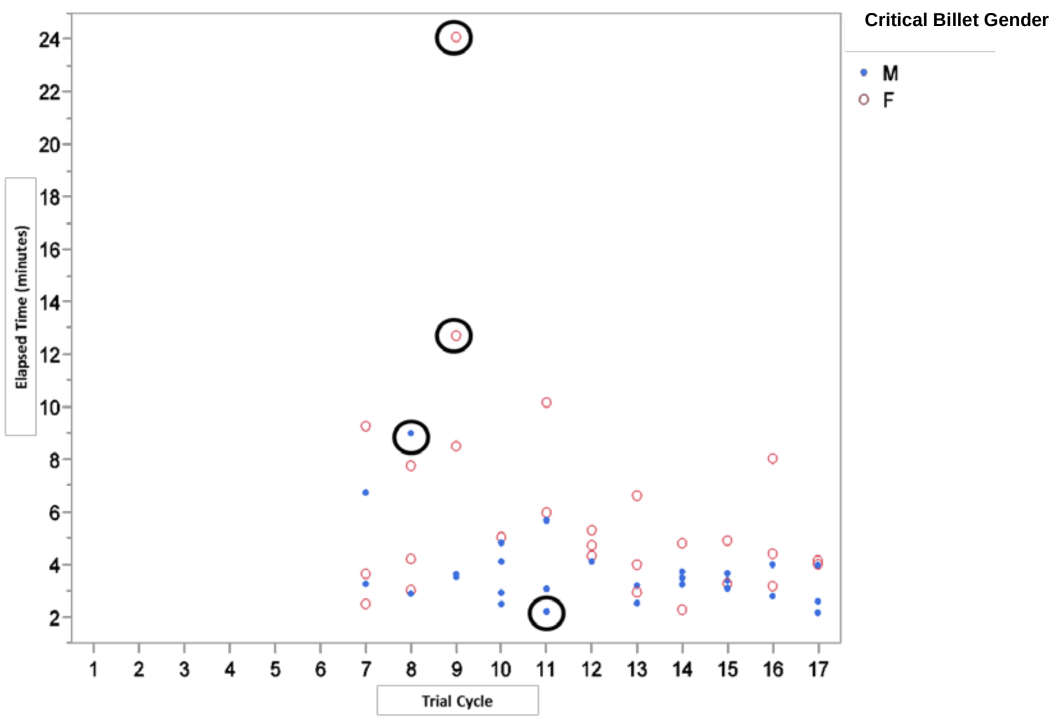


Figure B-31. Prep M242 Main Gun; Reassembly by Critical Billet (Excluding)



B.5.6.2.2 Prep the M242 Main Gun by Critical Billet (Excluding Potential Influential Points) Data Table and Analysis

The tables below summarize the results of the task by critical billet. Table B-22 compares means across metrics and integration levels. Table B-23 presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-22. Prep M242 Main Gun by Critical Billet (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Prep M242 Main Gun (Gunner) **[excluding potential influential points]** (minutes) *	M	26	8.68	2.18
	F	25	9.98	2.64
Prep M242 Main Gun; Disassembly (Gunner) **[excluding potential influential points]** (minutes)	M	38	5.30	1.54
	F	42	5.47	1.47
Prep M242 Main Gun; Re-assembly (Gunner) **[excluding potential influential points]** (minutes) *	M	25	3.56	1.01
	F	25	5.07	2.16

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-23. Prep M242 Main Gun by Critical Billet (Excluding) ANOVA

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Prep M242 Main Gun (Gunner) **[excluding potential influential points]** (minutes)	3.67 (1, 49)	0.06*	F-M	1.30	14.95%	1.91 (47)	0.03*	0.06*	0.41	2.18	0.16	2.44
Prep M242 Main Gun; Disassembly (Gunner) **[excluding potential influential points]** (minutes)	0.26 (1, 78)	0.61	F-M	0.17	3.23%	0.51 (76)	0.31	0.61	-0.26	0.61	-0.39	0.73
Prep M242 Main Gun; Re-assembly (Gunner) **[excluding potential influential points]** (minutes)	10.13 (1, 34)	< 0.01††	F-M	1.52	42.72%	3.18 (34)	< 0.01*	< 0.01*	0.90	2.14	0.71	2.33

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances)

For the overall time to prep the M242, the M and F groups are normally distributed with Shapiro-Wilk p-values of 0.03 and 0.38, respectively. For disassembly time of M242, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). For the reassembly time of the M242, the M group is not normally distributed with Shapiro-Wilk p-value of less than 0.01, while the F group is normally distributed with p-value of 0.02. Since the standard deviation of the F group is more than twice that of the M group, we recommend using robust ANOVA presented above. We proceed with presenting ANOVA results, since they are confirmed by a Mann-Whitney Test ($p\text{-value} < 0.01$).

When excluding potential influential results for the overall time required while disassembling and reassembling the main gun, the Male group had a mean time of 8.68 minutes, which was statistically significantly faster than the Female group that recorded a mean time of 9.98 minutes. The Male group's 14.95% faster mean time when compared to the Female group was statistically significant. The reassembly time metric also shows statistical significance when comparing the Male group and Female group. The percentage differences equate to degradation in reassembly time for the Female Group when compared to the Male group. The Female group above shows a greater variability for the overall time and reassembly time metrics as shown by the SD column; the Female group shows less variability for the disassembly time metric. See Table B-22 and Table B-23 for detailed analytical results.

B.5.6.2.3 Prep the M242 Main Gun by Critical Billet (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.6.3 Prepare the M242 Main Gun by Integration Level

See Section B.5.6 for task description.

B.5.6.3.1 Prepare the M242 Main Gun by Integration Level Scatterplots

The scatter plots below display the results to complete the Prep M242 Main Gun task by integration level. The data points contained inside dark black circles were determined to be data errors and were excluded from the analysis.

Figure B-32. Prep M242 Main Gun by Integration Level

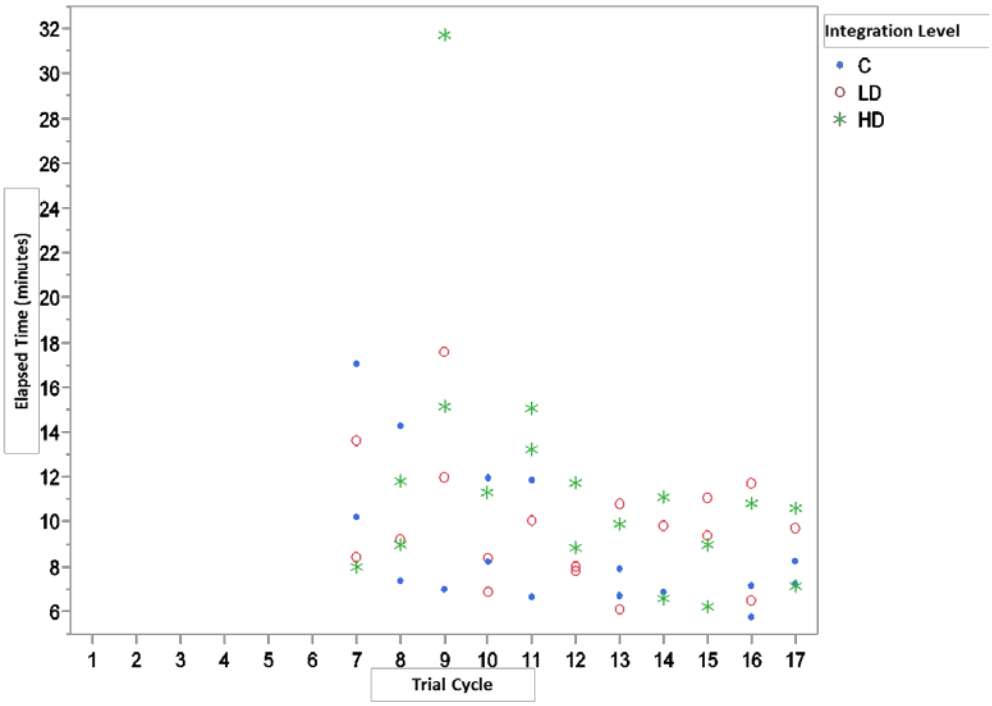


Figure B-33. Prep M242 Main Gun; Disassembly by Integration Level

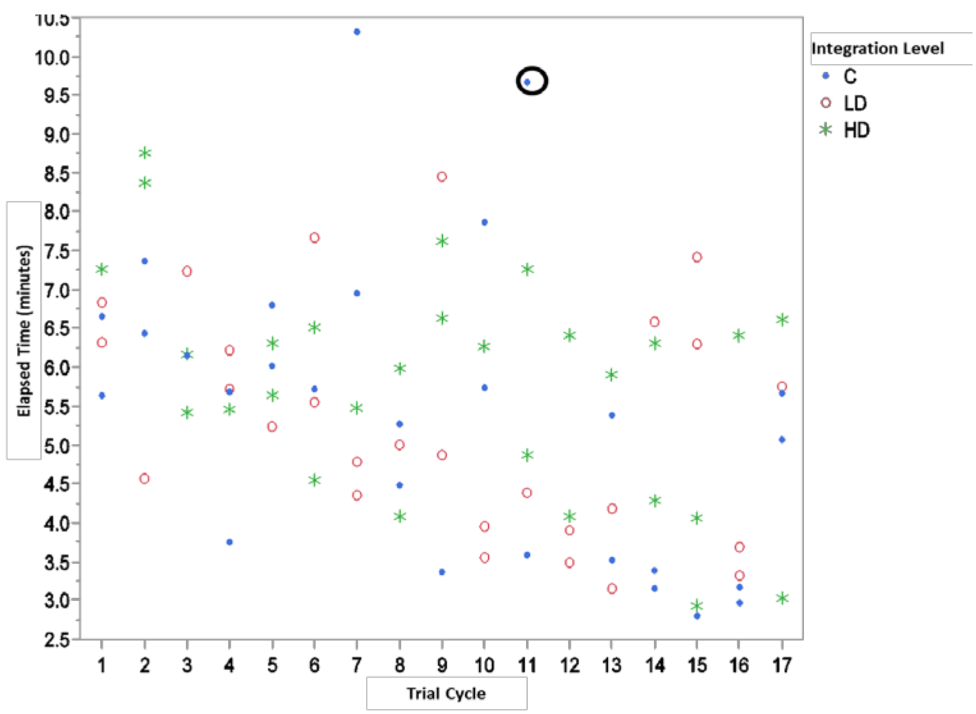
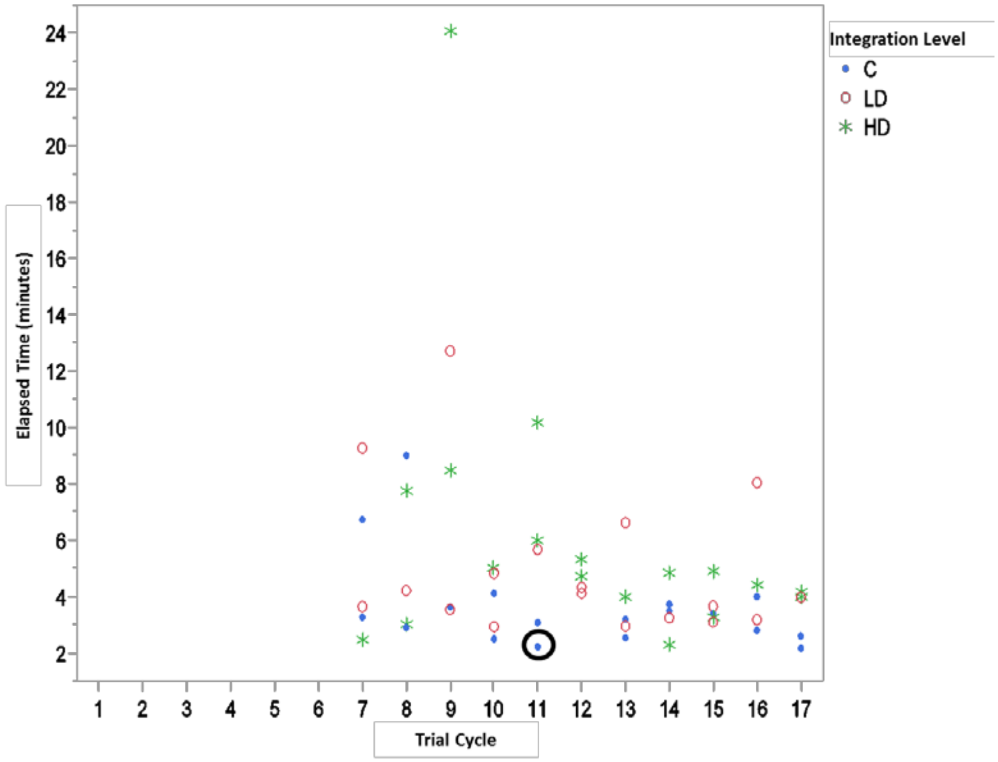


Figure B-34. Prep M242 Main Gun; Reassembly by Integration Level



B.5.6.3.2 Prepare the M242 Main Gun by Integration Level Data Table and Analysis

The tables below summarize the results of the task by integration level. Table B-24 compares means across metrics and integration levels. Table B-25 presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-24. Prep M242 Main Gun by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Prep M242 Main Gun (minutes)	C	18	8.75	3.11
	LD	18	9.83	2.78
	HD	18	11.50	5.67
Prep M242 Main Gun; Disassembly (minutes)	C	27	5.29	1.79
	LD	27	5.28	1.47
	HD	28	5.81	1.45
Prep M242 Main Gun; Re-assembly (minutes)	C	17	3.70	1.71
	LD	18	4.99	2.63
	HD	18	6.04	4.95

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-25. Prep M242 Main Gun by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Prep M242 Main Gun (minutes) †	1.72 (2, 32)	0.20	LD-C	1.08	12.38%	0.12	0.28	-0.20	2.37	-0.58
			HD-C	2.75	31.48%	0.04	0.08	0.75	4.76	0.16
			HD-LD	1.67	16.99%	0.36	0.27	-0.29	3.63	-0.87
Prep M242 Main Gun; Disassembly (minutes)	1.02 (2, 79)	0.36	LD-C	-0.02	-0.30%	1.00	-0.76	0.73	-0.91	0.88
			HD-C	0.52	9.77%	0.45	-0.22	1.25	-0.37	1.40
			HD-LD	0.53	10.10%	0.43	-0.20	1.27	-0.35	1.42
Prep M242 Main Gun; Re-assembly (minutes) †	2.07 (2, 50)	0.02*	LD-C	1.29	34.87%	0.02	0.35	1.30	0.22	1.48
			HD-C	2.34	63.30%	0.01	0.78	1.92	0.62	2.12
			HD-LD	1.05	21.07%	0.39	-0.32	1.17	-0.63	1.33

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances). For comparisons utilizing robust ANOVA, Tukey confidence intervals have been replaced by individual Welch's T-tests with p-values compared to a Bonferroni corrected threshold.

For the overall time to prep the M242, the C and LD groups are not normally distributed with Shapiro-Wilk Test p-values of less than 0.01, while the HD group is normally distributed with p-value of 0.01, respectively. Since the standard deviation of the HD group is more than twice that of the LD group, we recommend using the robust ANOVA results presented above. We proceed with presenting robust ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.01). We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.01). For the disassembly portion of the time, the C, LD, and HD groups are normally distributed with Shapiro-Wilk Test p-values of 0.08, 0.33, and 0.63, respectively. For the assembly portion of the time, the C, LD, and HD groups are not normally distributed with Shapiro-Wilk Test p-values of less than 0.01. Because of lack of normality, we recommend using the results of the Kruskal-Wallis test presented in the table above.

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. For the overall time required to disassemble and assemble the main gun the C group had a mean time of 8.75 minutes, which was faster than the HD groups that recorded a mean time of 11.50 minutes. However, the C group's 12.38% faster mean time when compared to the HD group was not statistically significant. No metric in Table B-25 shows a statistical significance when comparing the HD group to the C group and the LD group. See Table B-24 and Table B-25 for detailed analytical results.

B.5.6.3.3 Prepare the M242 Main Gun by Integration Level Additional Insights

Observations when examined by integration level in the low-density group are less apparent, reveal a variance increase, and fall below the threshold required for statistical significance. The analysis above shows the masking effect that analysis by integration level alone has on the results. When examining specific tasks for LAV gunnery it is imperative to consider analysis by critical billet as it exposes the masking of analysis by the integration level alone.

B.5.6.4 Prepare the M242 Main Gun by Integration Level (Excluding Potential Influential Points)

See B.5.6 for task description.

B.5.6.4.1 Prepare the M242 Main Gun by Integration Level (Excluding Potential Influential Points) Scatterplots

The scatter plots below display the results to complete the Prep M242 Main Gun task by integration level. The data points in lightly colored black circles were determined to be potential influential points. They were excluded from the analysis for this sub-section. The data points contained inside dark black circles were determined to be data errors and were excluded from the analysis.

Figure B-35. Prep M242 Main Gun by Integration Level (Excluding)

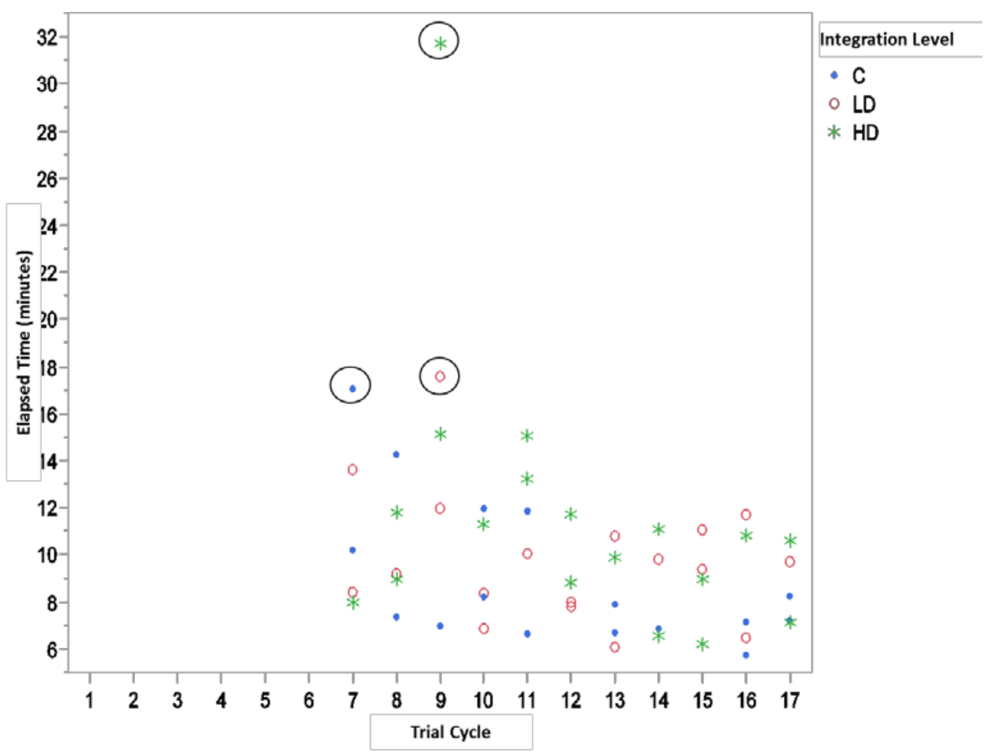


Figure B-36. Prep M242 Main Gun; Disassembly by Integration Level (Excluding)

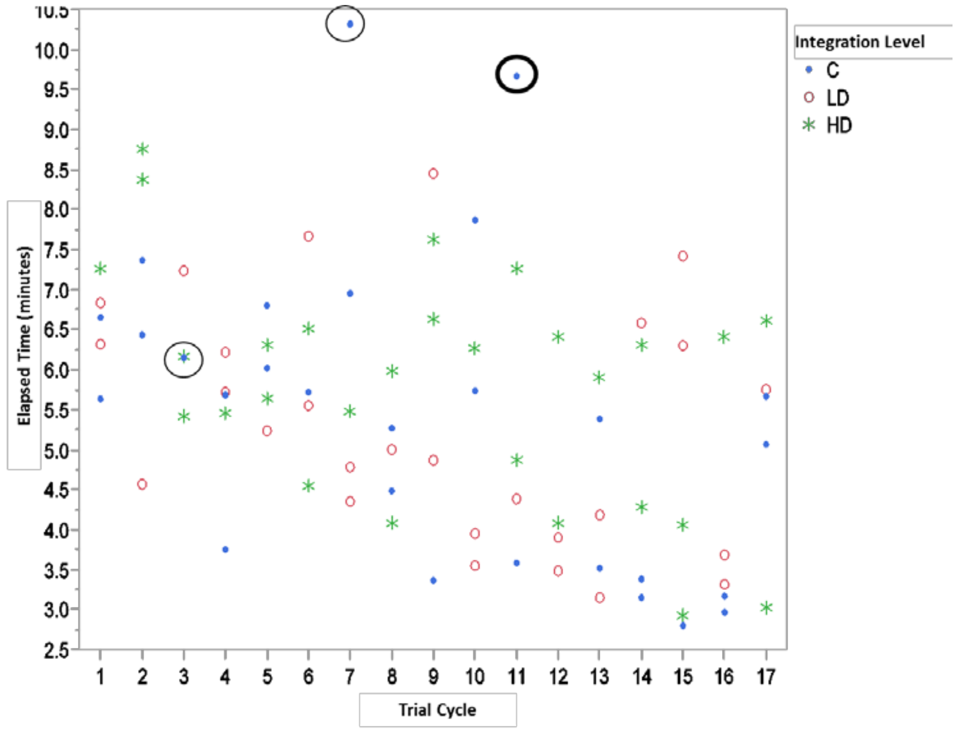
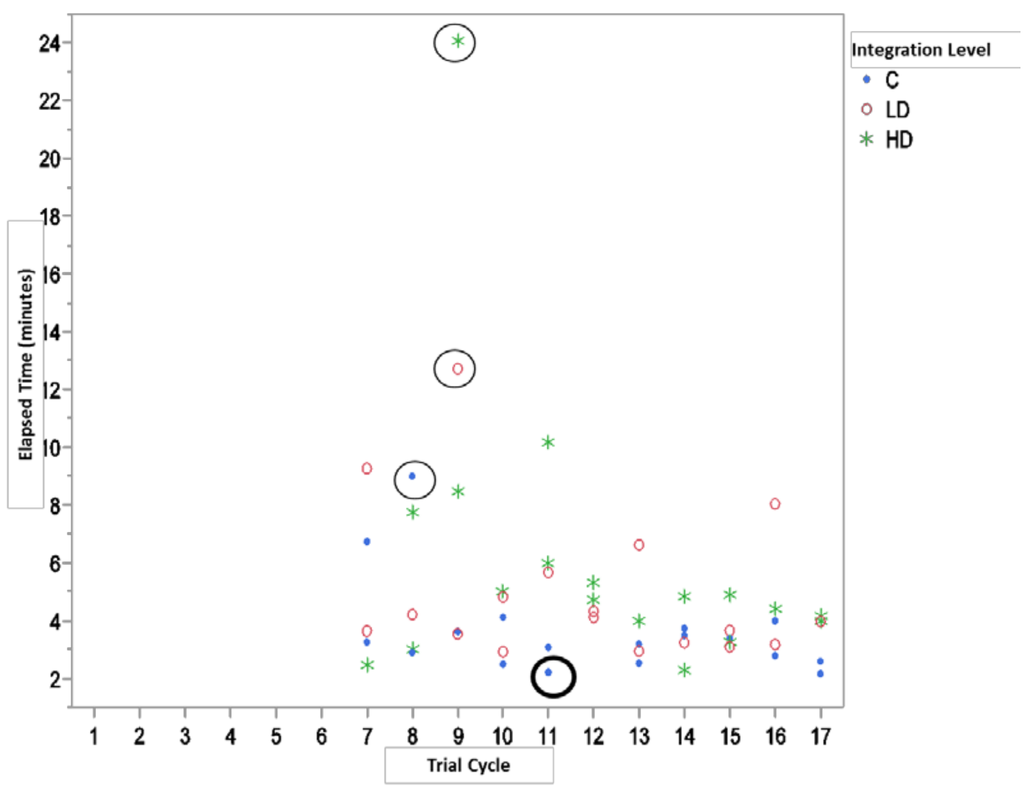


Figure B-37. Prep M242 Main Gun; Reassembly by Integration Level (Excluding)



B.5.6.4.2 Prepare the M242 Main Gun by Integration Level (Excluding Potential Influential Points) Data Table and Analysis

The tables below summarize the results of the Prep M242 Main Gun task by integration level. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-26. Prep M242 Main Gun by Integration Level (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Prep M242 Main Gun **[excluding potential influential points]** (minutes) *	C	17	8.26	2.38
	LD	17	9.37	2.06
	HD	17	10.31	2.66
Prep M242 Main Gun; Disassembly **[excluding potential influential points]** (minutes)	C	26	5.10	1.52
	LD	27	5.28	1.47
	HD	27	5.80	1.47
Prep M242 Main Gun; Re-assembly **[excluding potential influential points]** (minutes) *	C	16	3.37	1.06
	LD	17	4.54	1.84
	HD	17	4.98	2.11

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-27. Prep M242 Main Gun by Integration Level (Excluding) ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Prep M242 Main Gun **[excluding influential points]** (minutes)	3.17 (2, 48)	0.05*	LD-C	1.12	13.51%	0.37	-0.31	2.54	-0.60	2.83
			HD-C	2.05	24.85%	0.04*	0.63	3.47	0.34	3.77
			HD-LD	0.94	9.99%	0.49	-0.49	2.36	-0.78	2.65
Prep M242 Main Gun; Disassembly **[excluding influential points]** (minutes)	1.58 (2, 48)	0.21	LD-C	0.18	3.48%	0.90	-0.53	0.88	-0.67	1.03
			HD-C	0.70	13.67%	0.21	-0.01	1.40	-0.15	1.55
			HD-LD	0.52	9.85%	0.41	-0.18	1.22	-0.32	1.36
Prep M242 Main Gun; Re-assembly **[excluding influential points]** (minutes)	3.73 (2, 47)	0.03*	LD-C	1.17	34.65%	0.14	0.11	2.23	-0.11	2.44
			HD-C	1.61	47.82%	0.03*	0.55	2.67	0.33	2.89
			HD-LD	0.44	9.79%	0.74	-0.60	1.49	-0.81	1.70

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

For all tasks above, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the overall time to prep the M242, the C group is not normally distributed with Shapiro-Wilk Test p-values of less than 0.01, while the LD and HD groups are normally distributed with p-values of 0.97 and 0.66, respectively. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-

Wallis Test (p-value = 0.05). For the disassembly portion of the time, the C, LD, and HD group data are normally distributed with Shapiro-Wilk Test p-values of 0.09, 0.33, and 0.71, respectively. For the assembly portion of the time, the C and LD groups are not normally distributed with Shapiro-Wilk Test p-values of less than 0.01, while the HD group is normally distributed with a p-value of 0.06. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.01).

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. When excluding potential influential results for the overall time required disassembling and assembling the main gun, the C group had a mean time of 8.26 minutes, which was statistically significantly faster than the LD and HD groups that recorded a mean time of 9.37 and 10.31 minutes, respectively. The C group's 24.85% faster mean time when compared to the HD group was statistically significant. The reassembly time metric shown in Table B-27 also shows a statistical significance when comparing the HD group to the C group. The 47.82% percentage difference equates to degradation in reassembly time for HD groups when compared to the C group. See Table B-26 and Table B-27 for detailed analytical results.

B.5.6.4.3 Prepare the M242 Main Gun by Integration Level (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.7 Load and Stow Additional Ammunition Overview

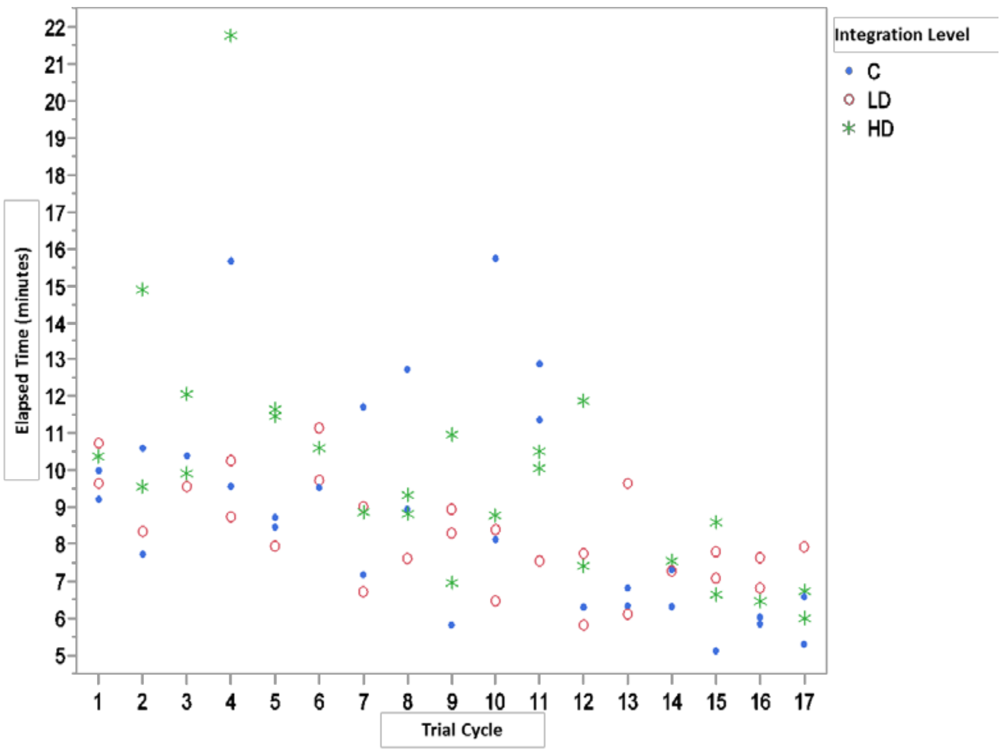
This subtask required the LAV-25 crew to load and stow a full complement of 25 mm ammunition into the vehicle. An LAV-25 can hold 150 rounds of High Explosive (HE) ammunition and 60 rounds of Armor Piercing (AP) ammunition linked and prepared in the vehicle's HE and AP ready-boxes. Two additional full uploads of each ammo type is also stored within the vehicle in ammunition cans for a total of 180 AP and 450 HE rounds. Dummy rounds weighing just over a pound each were used to simulate the linked rounds. These dummy rounds were broken out of their containers and linked prior to the start of the subtask. Crews had a tendency to break apart the large belt of 150 HE rounds to more rapidly upload the rounds into the vehicle. The two additional uploads were comprised of ten cans of HE training rounds, each weighing 50 lbs., and four cans of AP training rounds, each weighing 55 lb. Without assistance, a crew must be able to properly load and stow their ready rounds and additional uploads on their vehicle as part of normal combat preparation. Prior to any typical operation, a crew would depart carrying a full complement of ammunition.

B.5.7.1 Load and Stow Additional Ammunition by Integration Level

B.5.7.1.1 Load and Stow Additional Ammunition by Integration Level Scatterplots

The scatter plots below display the results to complete the Load and Stow Additional Ammo task by integration level. No outliers were identified for this sub-section.

Figure B-38. Load and Stow Additional Ammo by Integration Level



B.5.7.1.2 Load and Stow Additional Ammunition by Integration Level Data Table and Analysis

The tables below summarize the results of the Load and Stow Additional Ammo by integration level. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-28. Load and Stow Additional Ammo by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Load & Stow Additional Ammo (minutes)	C	29	8.84	2.90
	LD	27	8.26	1.39
	HD	28	9.68	3.18

*Indicates there is a statistically significant difference in a two-sided hypothesis test between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-29. Load and Stow Additional Ammo ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Load & Stow Additional Ammo (minutes)	2.47 (2, 48)	0.10††	LD-C	-0.58	-6.56%	0.34	-1.36	0.20	-1.59	0.43
			HD-C	0.84	9.51%	0.30	-0.21	1.89	-0.51	2.19
			HD-LD	1.42	17.20%	0.04	0.56	2.28	0.31	2.53

*Indicates there is a statistically significant difference in a two-sided hypothesis test between integration levels according to ANOVA or a non-parametric equivalent test.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances). For comparisons utilizing robust ANOVA, Tukey confidence intervals have been replaced by individual Welch's T-tests with p-values compared to a Bonferroni corrected threshold

For the time to load and stow additional ammo, the C and LD groups are normally distributed with Shapiro-Wilk Test p-values of 0.03 and 0.83, respectively, while the HD group is not normally distributed with a p-value of less than 0.01. Since the standard deviation of the HD group is more than twice that of the LD group, we recommend using the robust ANOVA results presented above. We proceed with presenting robust ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.22).

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. For the overall time load and stow ammunition the C group had a mean time of 8.84 minutes, which was slower than the LD group that recorded a time of 8.26 minutes and faster than the HD group that recorded a time of 9.68 minutes. No results were statistically significant. Additionally, the HD group shows greater variability when compared to the C and LD groups as shown by the SD column; the LD group shows the lowest variability. See Table B-28 and Table B-29 for detailed analytical results.

B.5.7.1.3 Load and Stow Additional Ammunition by Integration Level Contextual Comments

The observed output from this task contained a high level of variance, making it difficult to draw a definitive conclusion. Like maintenance tasks, in real-world operations, longer times can translate into increased times required to prepare for a mission.

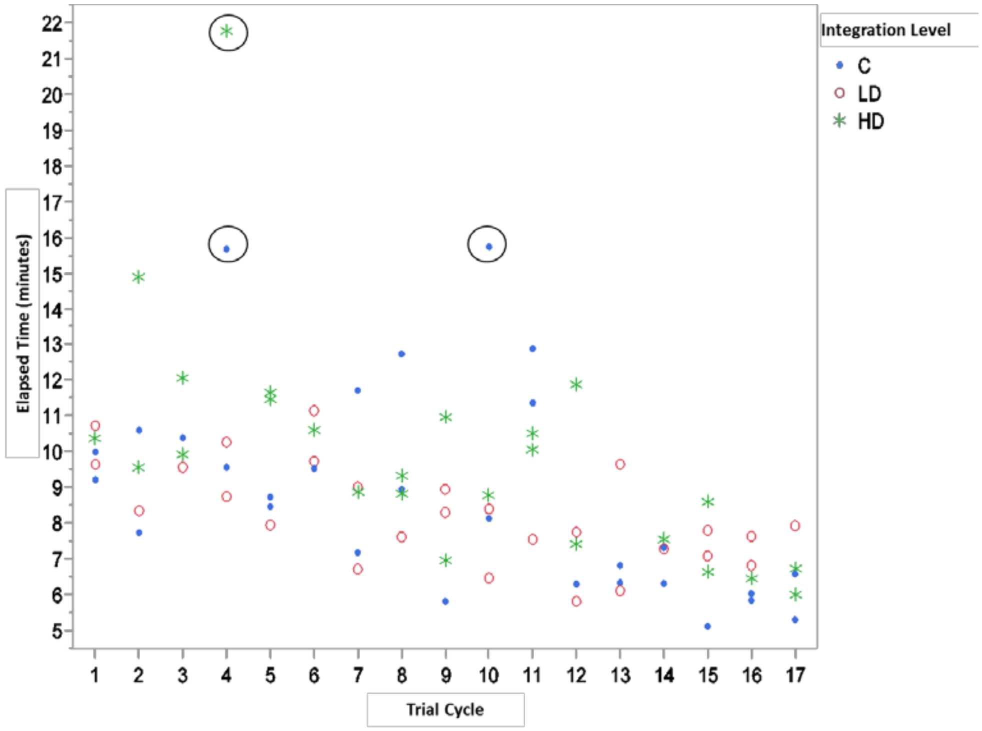
B.5.7.2 Load and Stow Additional Ammunition by Integration Level (Excluding Potential Influential Points) Overview

See B.5.7 for task description.

B.5.7.2.1 Load and Stow Additional Ammunition by Integration Level (Excluding Potential Influential Points) Scatterplots

The scatter plots below display the results to complete the Load and Stow Additional Ammo task by integration level. The data points contained inside lightly colored black circles were determined to be potential influential points; they are excluded from the analysis. No outliers were identified for this sub-section.

Figure B-39. Load and Stow Additional Ammo by Integration Level (Excluding)



B.5.7.2.2 Load and Stow Additional Ammunition by Integration Level (Excluding Potential Influential Points) Data Table and Analysis

The tables below summarize the results of the Load and Stow Additional Ammo task. This portion of the analysis excluded potential influential points. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-30. Load and Stow Additional Ammo by Integration Level (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Load & Stow Additional Ammo **[excluding potential influential points]** (minutes)	C	27	8.33	2.28
	LD	27	8.26	1.39
	HD	27	9.24	2.17

*Indicates there is a statistically significant difference in a two-sided hypothesis test between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-31. Load and Stow Additional Ammo by Integration Level (Excluding)

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Load & Stow Additional Ammo **[excluding potential influential points]** (minutes)	7.95 (2, 78)	0.14	LD-C	-0.07	-0.87%	0.99	-1.01	0.86	-1.20	1.05
			HD-C	0.90	10.82%	0.22	-0.03	1.84	-0.22	2.03
			HD-LD	0.97	11.79%	0.18	0.04	1.91	-0.15	2.10

*Indicates there is a statistically significant difference in a two-sided hypothesis test between integration levels according to ANOVA or a non-parametric equivalent test.

For the task above, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the time to load and stow additional ammo, the C, LD, and HD groups are normally distributed with Shapiro-Wilk p-values of 0.16, 0.83, and 0.31, respectively.

Statistical analysis included ANOVA and multiple comparisons using Tukey tests. When excluding potential influential results for the overall time required to load and stow additional ammunition, the C group had a mean time of 8.33 minutes, which was slower than the LD group mean time of 8.26 minutes, but faster than the HD group mean time of 9.24 minutes. No comparisons revealed statistically significant results. Additionally, the C group shows a greater variability when compared to the LD and HD groups as shown by the SD column. See Table B-30 and Table B-31 for detailed analytical results.

B.5.7.2.3 Load and Stow Additional Ammunition by Integration Level (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.8 Engage Main Gun Targets Overview

The main gun engagements portion of the live-fire trial day were conducted as crew events, with a large amount of the physical workload placed on the Gunner, designated below as the critical billet. Live-fire engagements were adapted from table VI gunnery qualification engagements found in MCWP 3-14.1, which all vehicle crews are required to fire to achieve qualified status. The culmination of a crew's proficiency is their ability during gunnery to qualify and excel at identifying, engaging, and destroying targets. Gunnery engagements, though not nearly as physically demanding as many of the maintenance and prep for combat tasks, reveal the knowledge and proficiency of a Marine as it pertains to their role as a crewman.

B.5.9 Main Gun Remedial Action Overview

This subtask required the LAV-25 Gunner, with assistance from the VC and Driver, to conduct remedial action on the main gun. It is similar in nature to the Prepare the M242 Main Gun prep for combat task. The crew must download the ammunition from the main gun, remove the feeder assembly and ensure it was correctly timed, and then reassemble the weapon system and upload ammunition. The Vehicle Commander and Driver were able to assist in any way, but due to the nature of crew responsibilities and the position of the gun in relation to the crew members, the Gunner bore most of the physical workload. Disassembly was achieved when the Gunner removed the feeder assembly from the weapon system. This task simulates a misfire on the main gun, and the actions that the crew would have to take to get their vehicle back in the fight. Rapidly restoring the functionality of the vehicle's primary weapon system is a critical task for a Gunner. The crew was evaluated on the time required to disassemble and to reassemble the weapon system. Times were calculated by adding the Gunner's overall time with the assistance time of the Driver and the Vehicle Commander.

B.5.9.1 Main Gun Remedial Action by Critical Billet

B.5.9.1.1 Main Gun Remedial Action by Critical Billet Scatterplots

The scatter plots below display the results to complete the Main Gun Remedial Action Subtask by critical billet. No outliers were identified for this sub-section. No data points were determined to be influential.

Figure B-40. Main Gun Remedial Action by Critical Billet

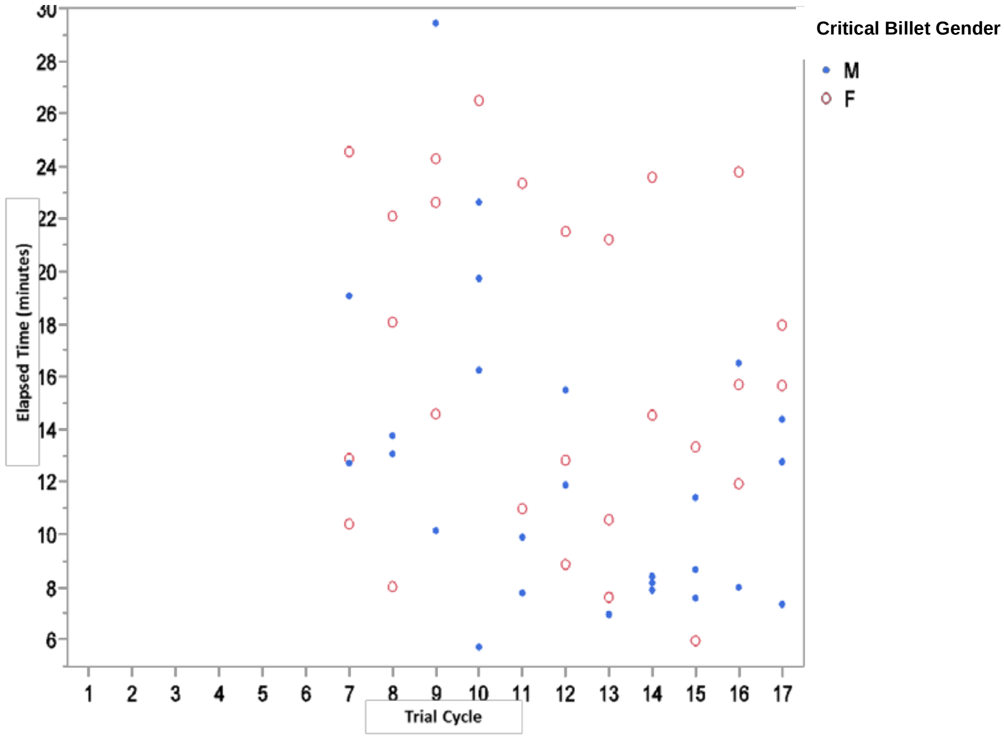


Figure B-41. Main Gun Remedial Action; Disassembly by Critical Billet

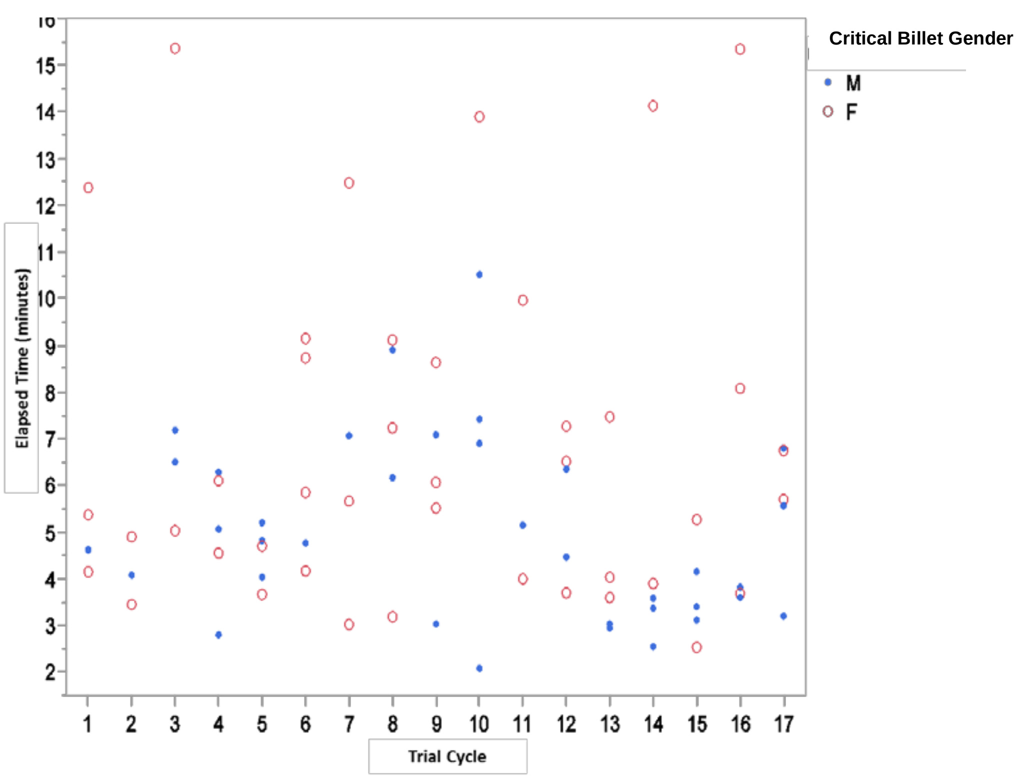
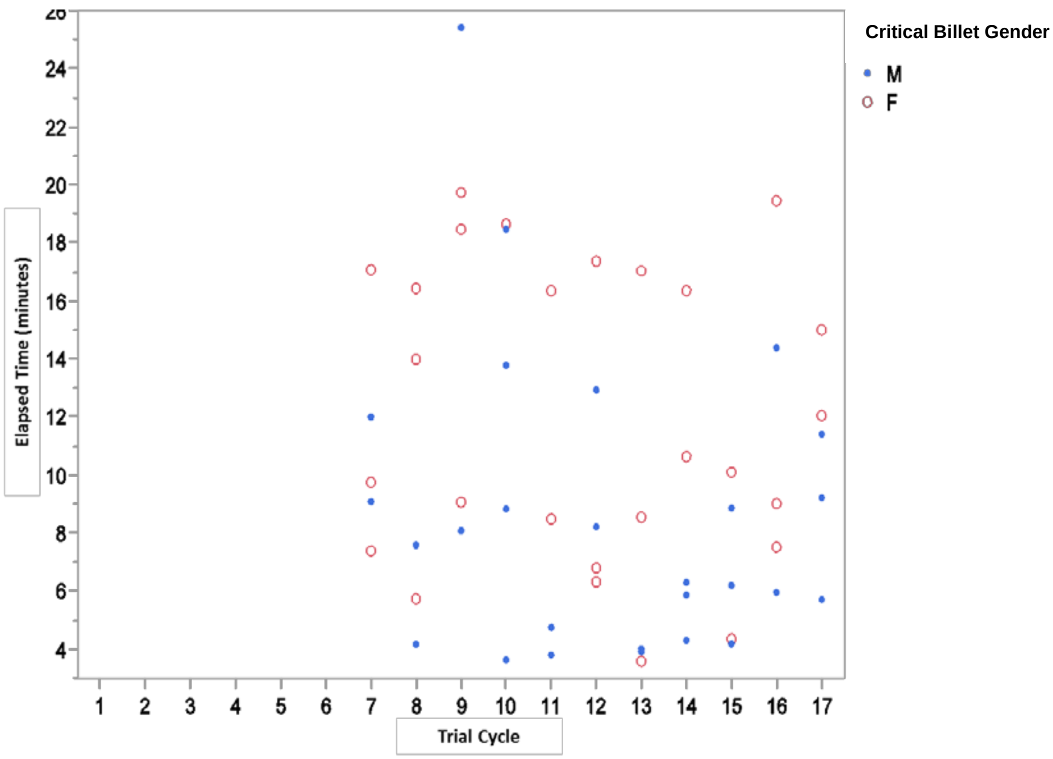


Figure B-42. Main Gun Remedial Action; Reassembly by Critical Billet



B.5.9.1.2 Main Gun Remedial Action by Critical Billet Data Table and Analysis

The tables below summarize the results of Main Gun Remedial Action task. The first table compares means across metrics and critical billets. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-32. Main Gun Remedial Action by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Perform Main Gun Remedial Action (with assistance times) (Gunner) (minutes) *	M	27	12.32	5.60
	F	27	16.42	6.16
Perform Main Gun Remedial Action; Disassembly (with assistance times) (Gunner) (minutes) *	M	39	4.97	1.87
	F	42	6.77	3.51
Perform Main Gun Remedial Action; Re-assembly (with assistance times) (Gunner) (minutes) *	M	27	8.56	5.11
	F	27	12.05	5.09

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-33. Main Gun Remedial Action by Critical Billet ANOVA

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Diff	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Perform Main Gun Remedial Action (Gunner) (minutes)	6.57 (1, 52)	0.01*	F-M	4.10	33.30%	2.56 (52)	< 0.01*	0.01*	2.02	6.18	1.42	6.79
Perform Main Gun Remedial Action; Disassembly (Gunner) (minutes)	8.1 (1, 79)	< 0.01*	F-M	1.80	36.18%	2.91 (63)	< 0.01*	< 0.01*	1.00	2.60	0.77	2.83
Perform Main Gun Remedial Action; Re-assembly (Gunner) (minutes)	6.32 (1, 52)	0.02*	F-M	3.49	40.75%	2.51 (52)	< 0.01*	< 0.02*	1.69	5.29	1.16	5.81

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

For all tasks above, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the overall time to perform main gun remedial action, the M group is not normally distributed with Shapiro-Wilk Test p-value of less than 0.01, while the F group is normally distributed with p-value of 0.11. We proceed with presenting the ANOVA results, since they are confirmed by a Mann-Whitney Test (p-value < 0.01). For the disassembly portion of perform main gun remedial action, sample sizes are sufficient to satisfy normality assumptions. For the reassembly portion of perform main gun remedial action, the M group is not normally distributed with Shapiro-Wilk Test p-value of less than 0.01, while the F group is normally distributed with p-value of 0.05. We proceed with presenting the ANOVA results, since they are confirmed by a Mann-Whitney Test (p-value < 0.01).

For the overall time required to conduct remedial actions, the Male group had a mean time of 12.32 minutes, which was statistically significantly faster than the Female group that recorded a mean time of 16.42 minutes. The Male group's 33.30% faster mean time when compared to the Female group was statistically significant. The disassembly and reassembly time metrics also show statistical significance when comparing the Male group and Female group. The percentage differences equate to degradation in disassembly and reassembly time for the Female Group when compared to the Male group. Additionally, the Female group above shows a greater variability for the overall time and disassembly time metrics as shown by the SD column; the Female group shows less variability for the reassembly time metric. See Table B-32 and Table B-33 for detailed analytical results.

B.5.9.1.3 Main Gun Remedial Action by Critical Billet Contextual Comments

Remedial action of the M242 Main Gun is a task primarily executed by the Gunner. Although the Driver and Vehicle Commander can assist, the Gunner still assumes most of the physical demand. As the low-density integration crew could have either a male or female serving as the Gunner, this group masks an individual Gunner's performance. It is therefore beneficial to explore the results of this subtask based on the gender of the Gunner.

B.5.9.1.4 Main Gun Remedial Action by Critical Billet Additional Insights

Observations and context for this task are similar to those for the Prepare the M242 Main Gun Task due to the almost identical nature of the subtasks. The key difference is that remedial action allows assistance from the additional crewmembers from the start, and Gunners are required to download ammunition prior to beginning disassembly of the weapon system. Longer times to ensure primary weapon system functionality have the potential to impede operations, especially when the LAV-25 is operating in a direct fire or support by fire role.

B.5.9.2 Main Gun Remedial Action by Critical Billet (Excluding Potential Influential Points)

See B.5.9 and B.5.9.1 for task description.

B.5.9.2.1 Main Gun Remedial Action by Critical Billet (Excluding Potential Influential Points) Scatterplots

The scatter plots below display the results to complete the Main Gun Remedial Action Subtask by critical billet. The data points determined to be influential are identified by lightly colored black circles. No outliers were identified for this sub-section.

Figure B-43. Main Gun Remedial Action; Disassembly by Critical Billet (Excluding)

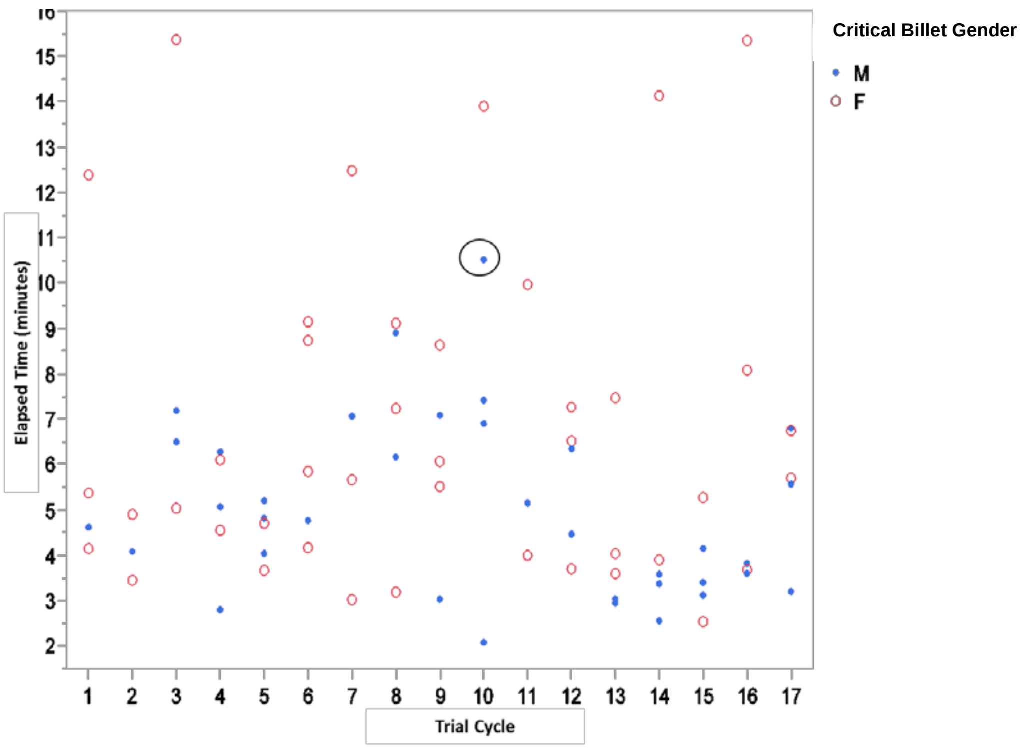
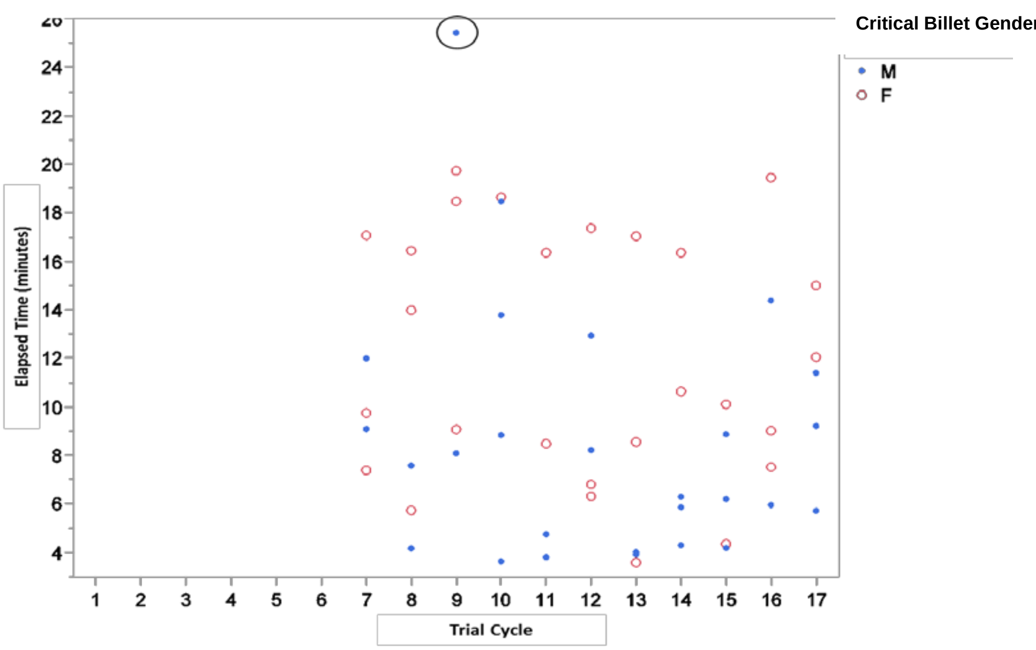


Figure B-44. Main Gun Remedial Action; Disassembly by Critical Billet (Excluding)



B.5.9.2.2 Main Gun Remedial Action by Critical Billet (Excluding Potential Influential Points) Data Table and Analysis

The tables below summarize the results of Main Gun Remedial Action task. The first table compares means across metrics and critical billets. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-34. Main Gun Remedial Action by Critical Billet (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Perform Main Gun Remedial Action; Disassembly**[excluding potential influential points]** (Gunner) (minutes)*	M	38	4.83	1.65
	F	42	6.77	3.51
Perform Main Gun Remedial Action; Re-assembly (Gunner) **[excluding potential influential points]** (minutes)*	M	26	7.91	3.91
	F	27	12.05	5.09

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-35. Main Gun Remedial Action by Critical Billet (Excluding)

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Perform Main Gun Remedial Action; Disassembly **[excluding potential influential points]** (Gunner) (minutes)	10.35 (1, 60)	< 0.01††	F-M	1.94	40.29%	3.22 (60)	< 0.01*	< 0.01*	1.16	2.73	0.93	2.95
Perform Main Gun Remedial Action; Re-assembly **[excluding potential influential points]** (Gunner) (minutes)	10.94 (1, 51)	< 0.01*	F-M	4.14	52.28%	3.32 (49)	< 0.01*	< 0.01*	2.52	5.75	2.05	6.22

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances)

For the disassembly portion of perform main gun remedial action, sample sizes are sufficient to satisfy normality assumptions. Since the standard deviation of the F group is more than twice that of the M group, we recommend using the robust ANOVA results presented above. For the reassembly portion of perform main gun remedial action, the

M and F groups are normally distributed with Shapiro-Wilk Test p-values of 0.01 and 0.05, respectively. Group standard deviations for the re-assembly portion are sufficiently similar to satisfy the equal variance assumption of ANOVA.

When excluding potential influential results for the remedial action disassembly time the Male group had a mean time of 4.83 minutes, which was statistically significantly faster than the Female group that recorded a mean time of 6.77 minutes. The Male group's 40.29% faster mean time when compared to the Female group was statistically significant. The reassembly time metric also shows statistical significance when comparing the Male group and Female group. The percentage differences equate to degradation in reassembly time for the Female Group when compared to the Male group. Overall remediation time contained no influential points and is omitted. Additionally, the Female group above shows a greater variability for disassembly and reassembly time metrics as shown by the SD column. The overall time to conduct remedial actions metric contained no potential influential points. See Table B-34 and Table B-35 for detailed analytical results.

B.5.9.2.3 Main Gun Remedial Action by Critical Billet (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.9.3 Main Gun Remedial Action by Integration Level

See B.5.9 for task description.

B.5.9.3.1 Main Gun Remedial Action by Integration Level Scatterplots

The scatter plots below display the results to complete the Main Gun Remedial Action Subtask by integration level. No outliers were identified for this sub-section. The analysis included all potential influential points.

Figure B-45. Main Gun Remedial Action by Integration Level

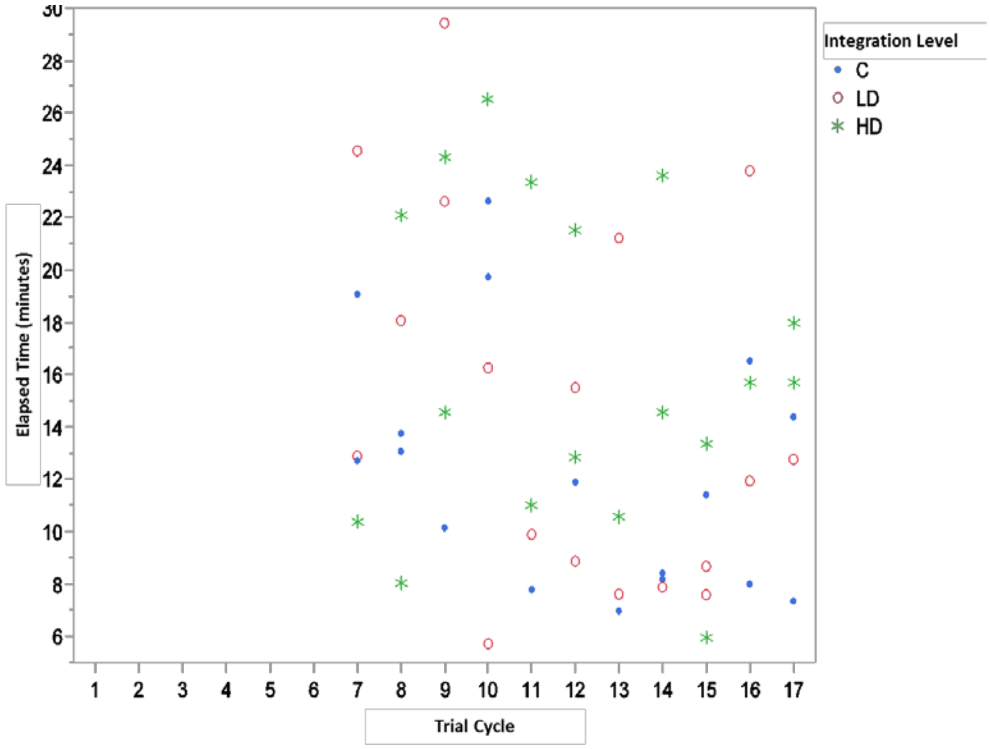


Figure B-46. Main Gun Remedial Action; Disassembly by Integration Level

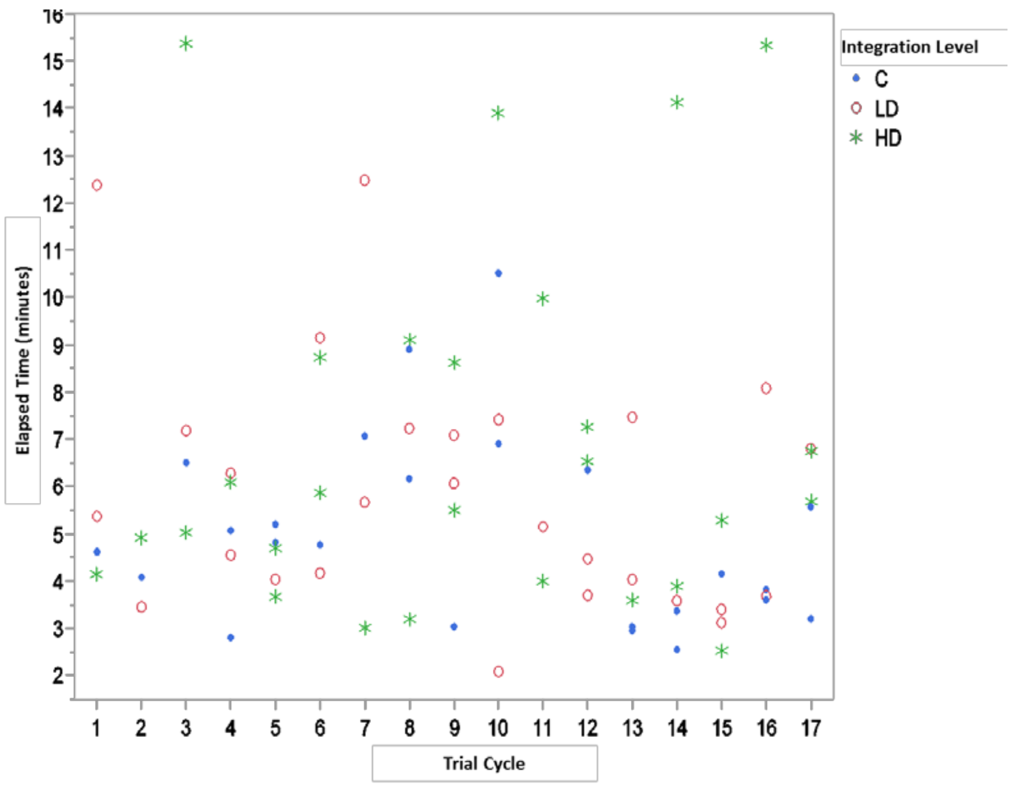
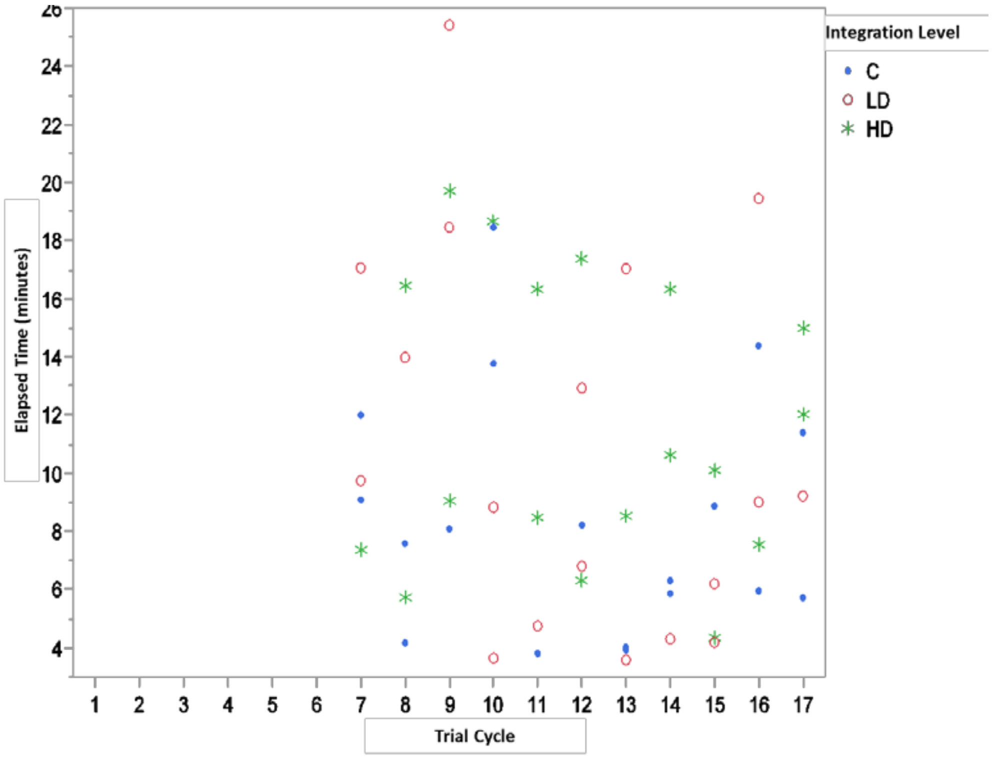


Figure B-47. Main Gun Remedial Action; Reassembly by Integration Level



B.5.9.3.2 Main Gun Remedial Action by Integration Level Data Table and Analysis

The tables below summarize the results of Main Gun Remedial Action task. The first table compares means across metrics and integration level. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-36. Main Gun Remedial Action by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Perform Main Gun Remedial Action (minutes)	C	18	12.16	4.79
	LD	18	14.74	7.08
	HD	18	16.22	6.10
Perform Main Gun Remedial Action; Disassembly (minutes) *	C	27	4.94	1.91
	LD	27	5.85	2.60
	HD	27	6.92	3.82
Perform Main Gun Remedial Action; Re-assembly (minutes)	C	18	8.42	4.14
	LD	18	10.81	6.48
	HD	18	11.67	4.89

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-37. Main Gun Remedial Action by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Perform Main Gun Remedial Action (minutes)	2.06 (2, 51)	0.14	LD-C	2.57	21.15%	0.42	-0.95	6.09	-1.67	6.82
			HD-C	4.05	33.32%	0.12	0.53	7.57	-0.19	8.30
			HD-LD	1.48	10.05%	0.75	-2.04	5.00	-2.76	5.72
Perform Main Gun Remedial Action; Disassembly (minutes) †	3.19 (2, 78)	0.12	LD-C	0.92	18.61%	0.16	0.08	1.43	-0.13	1.62
			HD-C	1.98	40.19%	0.05	0.37	2.08	0.17	2.35
			HD-LD	1.07	18.20%	0.48	-0.40	1.53	-0.57	1.75
Perform Main Gun Remedial Action; Re-assembly (minutes)	1.84 (2, 51)	0.17	LD-C	2.39	28.37%	0.37	-0.67	5.44	-1.29	6.07
			HD-C	3.25	38.57%	0.16	0.19	6.30	-0.44	6.93
			HD-LD	0.86	7.94%	0.88	-2.20	3.91	-2.83	4.54

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality

For all tasks above, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the overall time to perform main gun remedial action, the C, LD, and HD groups are normally distributed with Shapiro-Wilk Test p-values of 0.06, 0.14, and 0.46, respectively. For the disassembly portion of the time, the C and LD group data are normally distributed with Shapiro-Wilk Test p-values of 0.02 and 0.01, respectively, while the HD group is not normally distributed with p-value of less than 0.01. We recommend using the Kruskal-Wallis Test results shown in the table above. For the assembly portion of the time, the C, LD, and HD groups are normally distributed with Shapiro-Wilk Test p-values of 0.08, 0.09, and 0.16, respectively.

Statistical analysis included ANOVA and multiple comparisons using Tukey tests. For the overall time to conduct remedial action, the C group had a mean time of 12.16 minutes, which was faster than the LD and HD groups that recorded times of 14.74 minutes and 16.22 minutes, respectively. The only statistically significant result is the C group's 40.19% faster mean time when compared to the HD group during disassembly. Additionally, the HD and LD groups show greater variability when compared to the C group as shown by the SD column. See Table B-36 and Table B-37 for detailed analytical results.

B.5.9.3.3 Main Gun Remedial Action by Integration Level Additional Insights

Observations when examined by integration level remain largely the same as by critical billet, although results in the low-density group are less apparent and many results fall below the threshold required for statistical significance.

B.5.9.4 Main Gun Remedial Action by Integration Level (Excluding Potential Influential Points)

See B.5.9 for task description.

B.5.9.4.1 Main Gun Remedial Action by Integration Level (Excluding Potential Influential Points) Scatterplots

The scatter plots below display the results to complete the Main Gun Remedial Action Subtask by integration level. No outliers were identified for this sub-section. Influential points are identified by lightly colored black circles.

Figure B-48. Main Gun Remedial Action; Disassembly by Integration Level (Excluding)

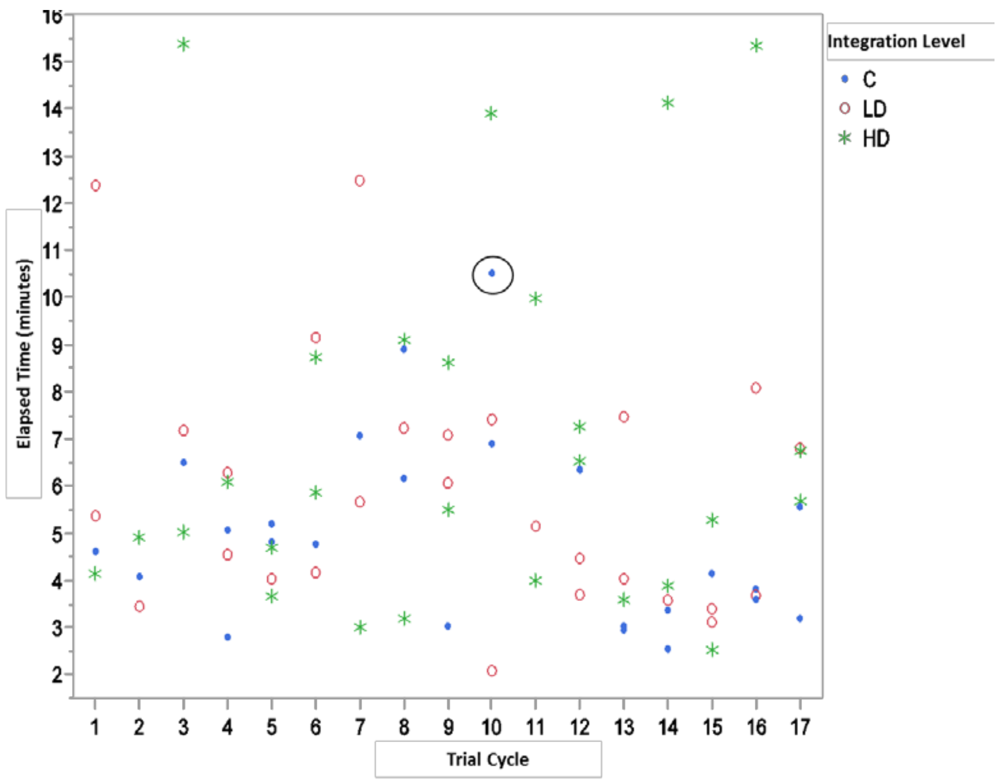
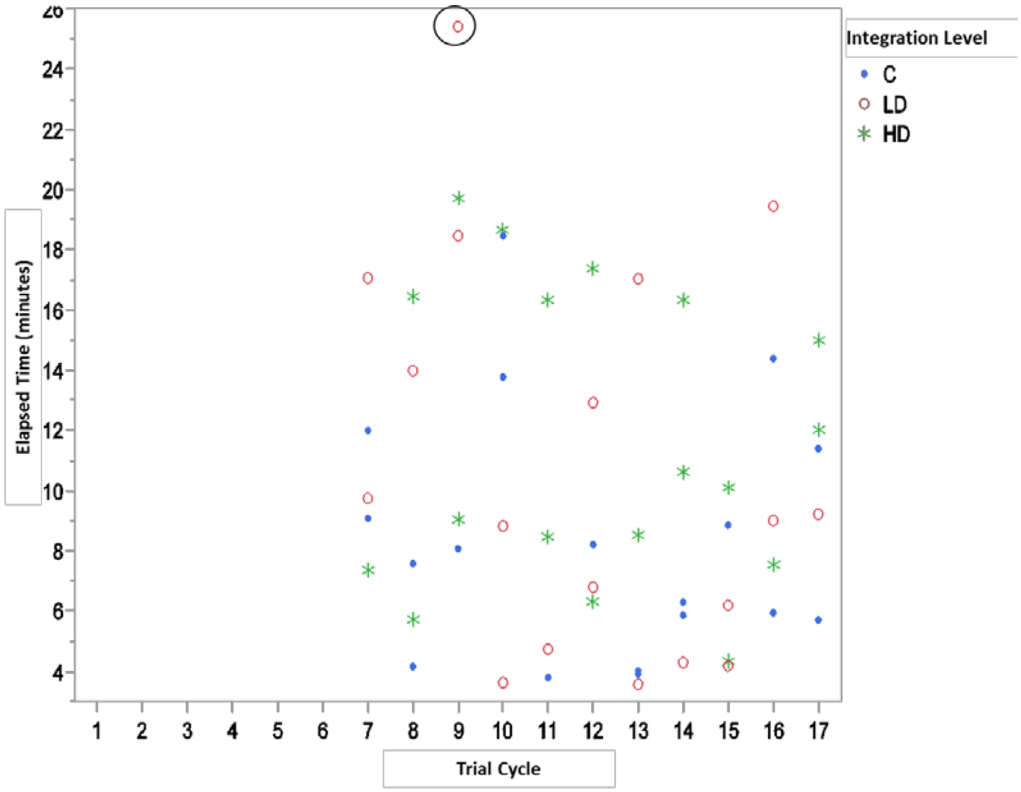


Figure B-49. Main Gun Remedial Action; Reassembly by Integration Level (Excluding)



B.5.9.4.2 Main Gun Remedial Action by Integration Level (Excluding Potential Influential Points) Data Table and Analysis

The tables below summarize the results of Main Gun Remedial Action task. The first table compares means across metrics and integration level. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-38. Main Gun Remedial Action (Excluding)

Metric	Integration Level	Sample Size	Mean	SD
Perform Main Gun Remedial Action; Disassembly (with assistance times) **[excluding potential influential points]** (minutes)*	C	26	4.72	1.58
	LD	27	5.85	2.60
	HD	27	6.92	3.82
Perform Main Gun Remedial Action; Re-assembly (with assistance times) **[excluding potential influential points]** (minutes)	C	18	8.42	4.14
	LD	17	9.95	5.52
	HD	18	11.67	4.89

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-39. Main Gun Remedial Action (Excluding) ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Perform Main Gun Remedial Action; Disassembly **[excluding potential influential points]** (minutes) †	3.99 (2, 77)	0.07*	LD-C	1.13	24.00%	0.10	0.20	1.52	0.02	1.75
			HD-C	2.20	46.56%	0.03*	0.50	2.15	0.28	2.47
			HD-LD	1.07	18.20%	0.48	-0.40	1.53	-0.57	1.75
Perform Main Gun Remedial Action; Re-assembly **[excluding potential influential points]** (minutes)	2 (2, 50)	0.15	LD-C	1.53	18.17%	0.62	-1.34	4.40	-1.93	4.99
			HD-C	3.25	38.57%	0.12	0.42	6.08	-0.16	6.66
			HD-LD	1.72	17.26%	0.55	-1.15	4.59	-1.74	5.18

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality.

For the disassembly portion of perform main gun remedial action, the C and LD groups are normally distributed with Shapiro-Wilk Test p-values of 0.19 and 0.01, respectively, while the HD group is not normally distributed with a p-value of less than 0.01. Additionally, the standard deviation of the HD group is more than twice that of the C group, which would indicate that at a minimum, a robust ANOVA is needed. However, due to lack of normality, we recommend using the Kruskal-Wallis results shown in the table above. For the assembly portion of the time, the C, LD, and HD groups are normally distributed with Shapiro-Wilk Test p-values of 0.08, 0.06, and 0.16, respectively.

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. When excluding potential influential results for disassembly time, the C group had a mean time of 4.72 minutes, which was faster than the LD group and HD group mean times of 5.85 minutes and 6.92 minutes, respectively. The 46.56% difference between the C group and HD group during disassembly was the only statistically significant result. Additionally, the HD and LD groups show greater variability when compared to the C group as shown by the Standard Deviation (SD) column. See Table B-38 and Table B-39 for detailed analytical results.

B.5.9.4.3 Main Gun Remedial Action by Integration Level (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.5.10 Gunner’s Defensive Manual Engagement Overview

The second live-fire defensive engagement was conducted from a defensive battle position. This engagement required the gunner to use AP ammunition to engage a vehicle target without using the Gunner’s hand control. The Gunner used the manual traverse and elevation controls within the turret to acquire and destroy the target. Crews had 80 seconds to acquire, identify, engage, and destroy the target. The time to engage the target and number of bursts required to kill the target was recorded. Live-fire engagements were adapted from MCWP 3-14.1 Table VI gunnery qualification engagements, which all vehicle crews are required to fire to achieve qualified status. Manual engagements require a greater amount of physical exertion because the Gunner does not rely on the electric turret to traverse and elevate.

B.5.10.1 Gunner’s Defensive Manual Engagement by Critical Billet

See B.5.9 for task description.

B.5.10.1.1 Gunner’s Defensive Manual Engagement by Critical Billet Scatterplots

Figure B-50 shows a scatter plot displaying the results to complete the Defensive Manual Engagement by critical billet task. No outliers were identified for this sub-section. Important to note is only successful engagements are displayed on the scatter plot and analyzed for time comparison. Figure B-51 shows a mosaic plot displaying engagement outcomes on a pass or fail scale. This plot displays the pass rate for each group and includes all data points.

Figure B-50. Defensive Manual Engagement by Critical Billet - Passed Engagement

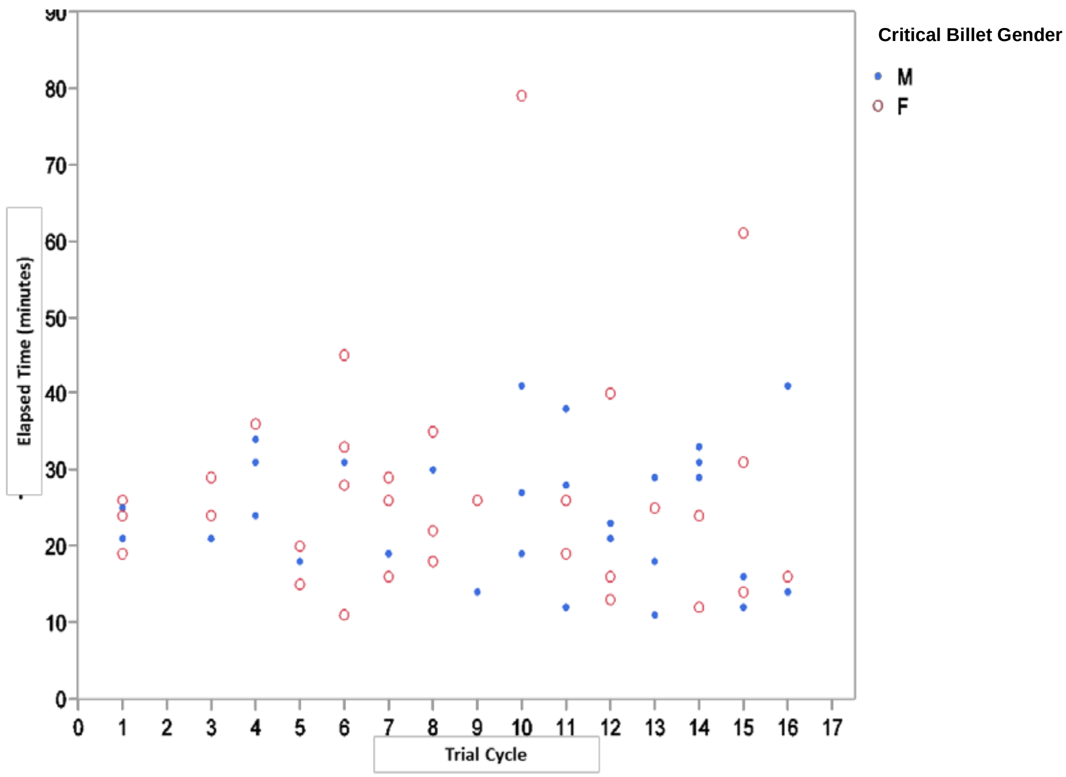
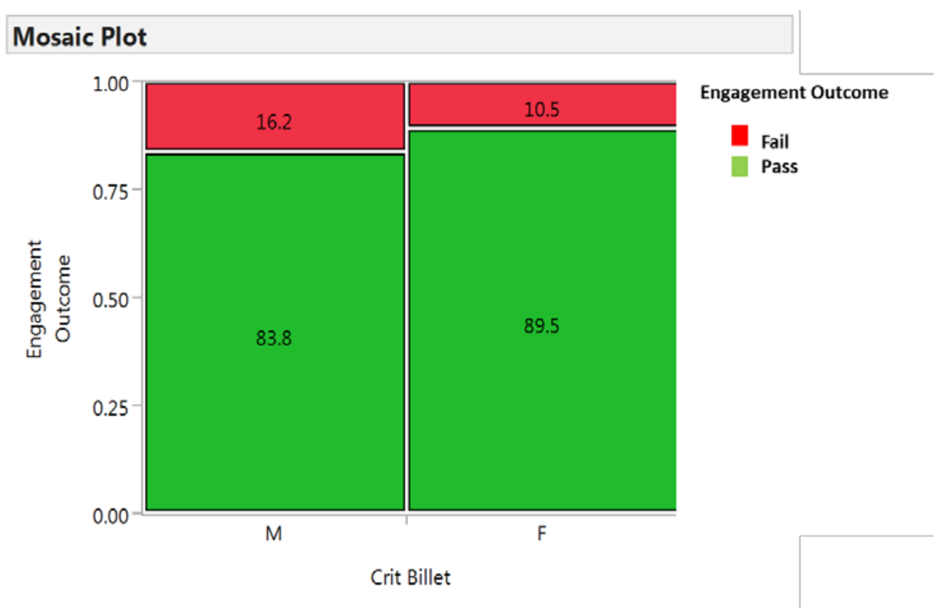


Figure B-51. Defensive Manual Engagement by Critical Billet - Pass/Fail



B.5.10.1.2 Gunner's Defensive Manual Engagement by Critical Billet Data Table and Analysis

The tables below summarize the results of Defensive Manual Engagement task. The first table compares means across metrics and critical billets. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences. Each table only examines successful (passed) engagements.

Table B-40. Defensive Manual Engagement by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Engage Defensive Targets (#2) - Manual (Gunner) (seconds)	M	31	24.13	8.46
	F	34	26.68	13.84

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-41. Defensive Manual Engagement by Critical Billet

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	T-Test Statistic (df)	1-Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Engage Defensive Targets (#2) - Manual (Gunner) (seconds)	0.78 (1, 63)	1.38	F-M	2.55	10.56%	0.9 (55)	0.19	0.37	-1.11	6.20	-2.17	7.26

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

For the task above, we proceed with presenting ANOVA results since all samples are sufficiently large (n > 30). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

During successful live-fire engagements, M group had a mean time of 24.13 seconds, which was faster than the F group mean time of 26.68 seconds. The 10.56% difference is not statistically significant. When observing the binary pass/fail criteria for the engagement, there is no difference in percent kill between M and F gunners in a Chi-Square (likelihood ratio) Test (p-value = 0.47).

B.5.10.1.3 Gunner’s Defensive Manual Engagement by Critical Billet Contextual Comments

Live-fire engagements are primarily executed by the Gunner. Although the Driver and Vehicle Commander can assist, the Gunner still assumes most of the physical demand. As the low-density integration crew could have either a male or female serving as the Gunner, this group masks an individual Gunner’s performance. It is therefore beneficial to explore the results of this subtask based on the gender of the Gunner.

B.5.10.1.4 Gunner’s Defensive Manual Engagement by Critical Billet Additional Insights

The observed output from this task contained a high level of variance, making it difficult to draw a definitive conclusion. A high pass rate and a proficient Gunner are desired to ensure a vehicle crew's effectiveness. Male and female Gunners performed similarly, with males acquiring and engaging targets slightly faster while females having a slightly higher engagement pass rate.

B.5.10.2 Gunner's Defensive Manual Engagement by Integration Level

See B.5.10 for task description

B.5.10.2.1 Gunner's Defensive Manual Engagement by Integration Level Scatterplot and Mosaic Plot

Figure B-52 shows a scatter plot displaying the results of passed engagement results for the Defensive Manual Engagement by integration level task. No outliers were identified for this sub-section. Only successful engagements are displayed on the scatter plot and analyzed for time comparison. Figure B-53 shows a mosaic plot displaying engagement outcomes on a pass or fail scale. This plot displays the pass rate for each group and includes all data points.

Figure B-52. Defensive Manual Engagement- Passed Engagements

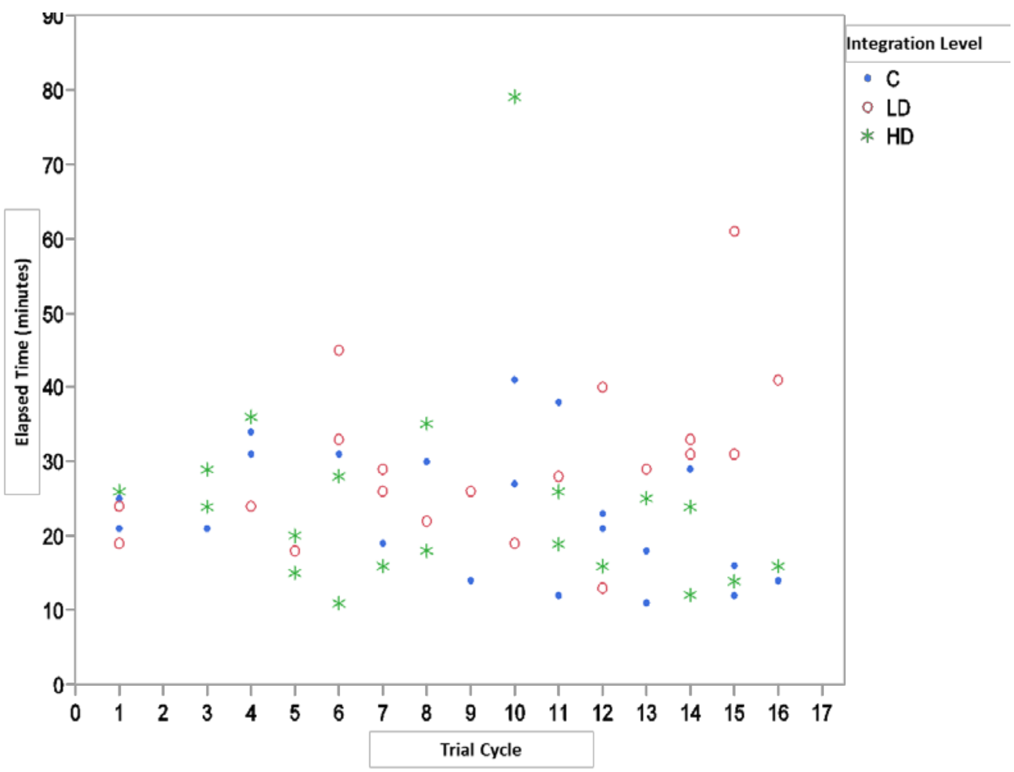
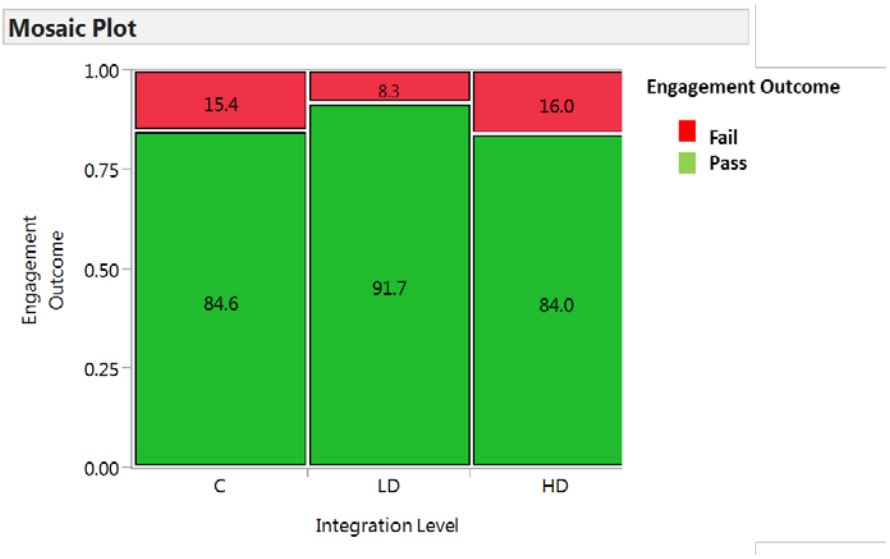


Figure B-53. Defensive Manual Engagement Pass/Fail



B.5.10.2.2 Gunner's Defensive Manual Engagement Data by Integration Level Table and Analysis

The tables below summarize the results of Engage Defensive Targets task. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences. Each table only examines successful (passed) engagements.

Table B-42. Engage Defensive Targets-Manual

Metric	Integration Level	Sample Size	Mean	SD
Engage Defensive Targets (#2) - Manual (seconds)	C	22	23.00	8.67
	LD	22	29.18	10.65
	HD	21	24.14	14.39

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-43. Engage Defensive Targets-Manual ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Engage Defensive Targets (#2) - Manual (seconds)	1.82 (2, 62)	0.17	LD-C	6.18	26.88%	0.18	0.20	12.17	-1.02	13.39
			HD-C	1.14	4.97%	0.94	-4.91	7.20	-6.15	8.43
			HD-LD	-5.04	-17.27%	0.32	-11.09	1.01	-12.33	2.25

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

For the task above, the C, LD, and HD groups are not normally distributed with Shapiro-Wilk Test p-values of less than 0.01. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.39).

During successful live-fire engagements, C group had a mean time of 23.00 seconds, which was faster than the LD and HD group mean times of 29.18 seconds and 24.14 seconds, respectively. No result comparisons are statistically significant. When observing the binary pass/fail criteria for the engagement, there is no difference in percent kill between C, LD, and HD groups in a binomial logistic regression/effect likelihood ratio test (P-value = 0.66)

B.5.10.2.3 Additional Insights

Observations when examined by integration level remain largely the same as by critical billet.

B.5.11 Conduct Maintenance Actions Overview

While a maintenance task does not always convey the same sense of urgency typically found in a casualty evacuation or live-fire task, they remain no less important. The ability to effectively perform continuing maintenance actions is a vital necessity of sustained operations. Vehicles require extensive preventative maintenance even in garrison, and austere terrain and expeditionary operating conditions only magnify these demands. This work is often physically demanding because the parts, tools, and equipment organic to LAR units are large and heavy. The ability of a vehicle crew to work with their maintainers to keep their vehicles in the fight is an everyday necessity. The maintenance actions portion of the non-live-fire trial day consisted of three separate subtasks; remove scout hatch armor, remove side panel armor, and a tire change. The tire change subtask was modeled, and is subsequently discussed below. For descriptive statistics on remove scout hatch and side panel armor, reference Appendix to Annex B. Each of these tasks was evaluated on the time to reach the mid-point of the task and the overall time to completion.

B.5.12 Mount/Remove Spare Tire Overview

The vehicle crew used the SL-3 associated with the LAV-25 to remove a tire, weighing approximately 200 lb, and mount the spare. For the first half of the subtask, the crew would jack up the vehicle, remove a vehicle tire, mount the vehicle’s spare tire and completely tighten the lug nuts. The crew would then remove the spare tire, mount the original tire and completely tighten the lug nuts, and finally return the spare tire and all additional equipment to the proper location. The crew would manually lift the spare tire from the ground to its mounting bracket on the side of the LAV-25. This was accomplished in a variety of ways, although two methods were most often observed. If the Gunner and Driver possessed the required strength, they would lift the tire straight up and mount it on the side of the vehicle. If one or both did not have the strength to conduct the lift, they would roll the vehicle up the side of the vehicle, a slower method of mounting the tire. If the two crewmen were unable to accomplish either method, they would request assistance from the Vehicle Commander to complete the tire re-mount.

B.5.12.1 Mount/Remove Spare Tire by Integration Level

B.5.12.1.1 Mount/Remove Spare Tire by Integration Level Scatterplots

The scatter plots below display the results of Mount/Remove Spare Tire by integration level task. No outliers were identified for this task.

Figure B-54. Mount/Remove Spare Tire

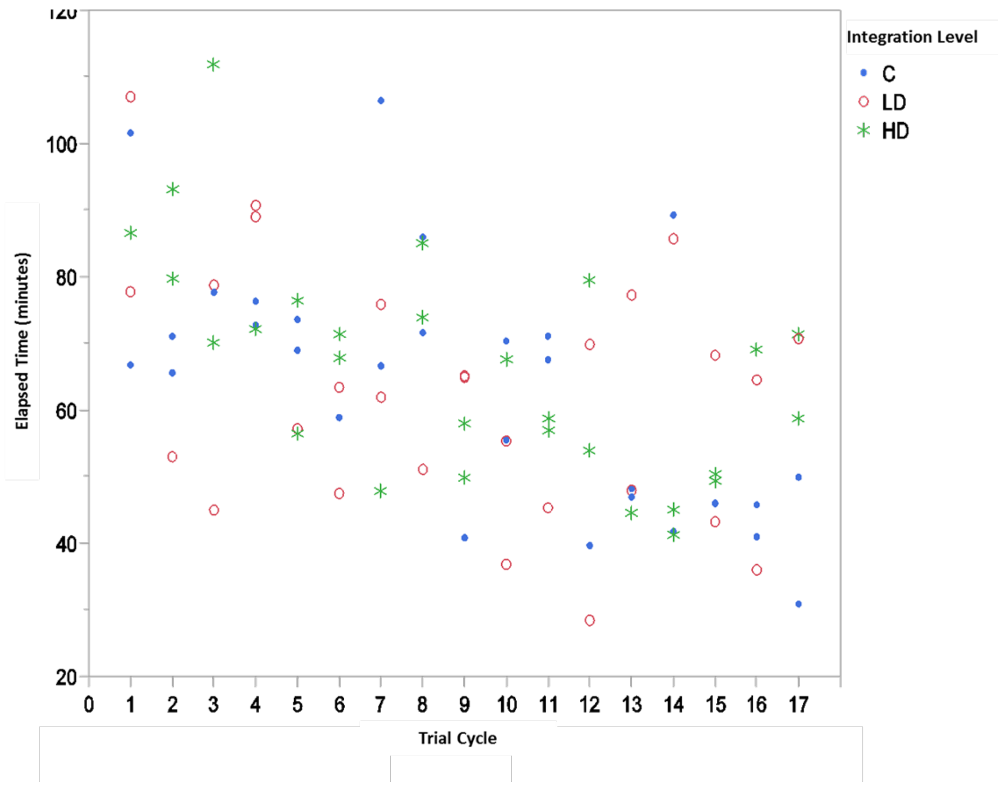


Figure B-55. Mount/Remove Spare Tire; Mounting Spare

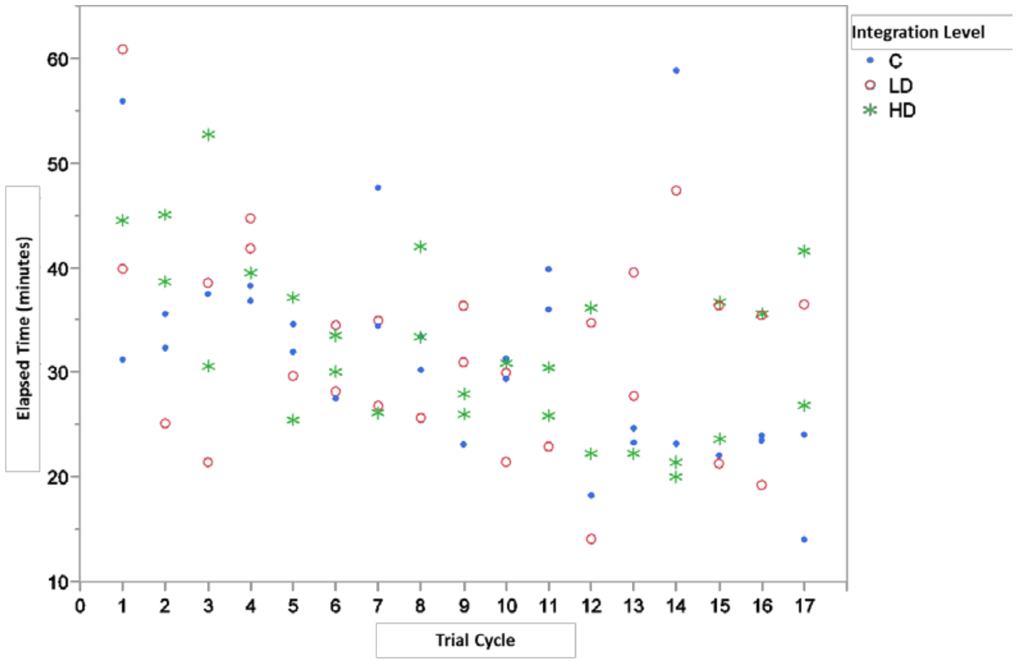
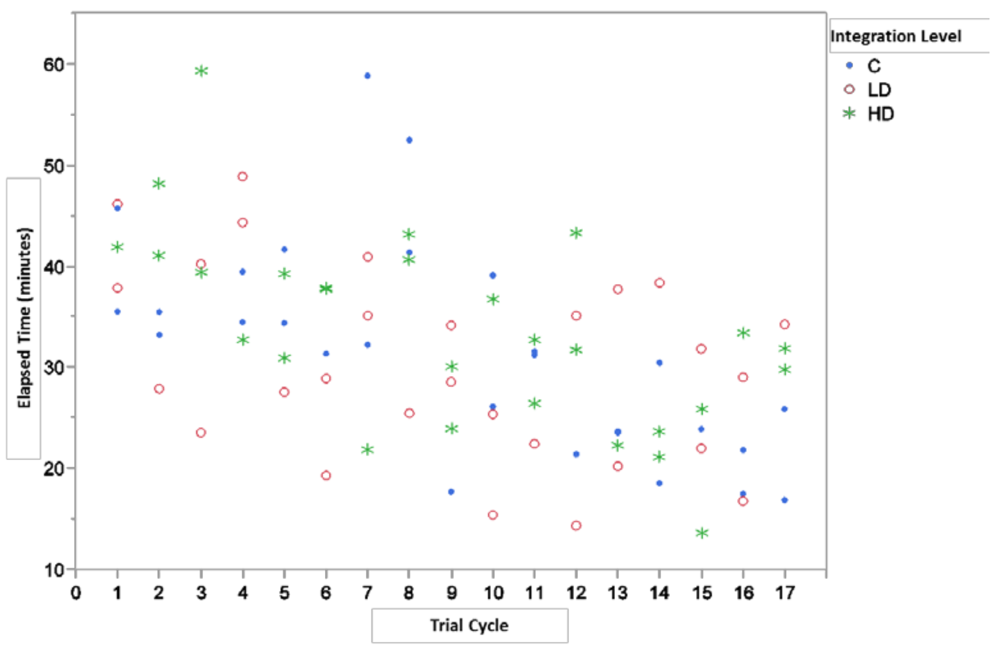


Figure B-56. Mount/Remove Spare Tire; Re-Mounting Original



B.5.12.1.2 Mount/Remove Spare Tire by Integration Level Data Table and Analysis

The tables below summarize the results for the Mount/Remove Spare Tire task. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-44. Mount/Remove Spare Tire

Metric	Integration Level	Sample Size	Mean	SD
Remove/Mount Spare Tire (minutes)	C	26	61.84	18.15
	LD	27	61.93	18.36
	HD	28	65.95	16.58
Remove/Mount Spare Tire; Mounting Spare (minutes)	C	27	30.80	9.28
	LD	27	31.83	9.63
	HD	28	32.38	8.27
Remove/Mount Spare Tire; Re-Mounting Original (minutes)	C	27	31.04	10.27
	LD	28	30.40	9.40
	HD	28	33.57	9.65

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-45. Mount/Remove Spare Tire ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Remove/Mount Spare Tire (minutes)	0.48 (2, 78)	0.62	LD-C	0.09	0.15%	1.00	-8.33	8.51	-10.04	10.22
			HD-C	4.10	6.64%	0.67	-4.24	12.45	-5.93	14.14
			HD-LD	4.01	6.48%	0.68	-4.25	12.28	-5.93	13.96
Remove/Mount Spare Tire; Mounting Spare (minutes)	0.21 92, 79)	0.81	LD-C	1.03	3.34%	0.91	-3.24	5.30	-4.11	6.17
			HD-C	1.58	5.13%	0.80	-2.65	5.81	-3.51	6.67
			HD-LD	0.55	1.74%	0.97	-3.68	4.79	-4.54	5.65
Remove/Mount Spare Tire; Re-Mounting Original (minutes)	0.82 (2, 80)	0.44	LD-C	-0.64	-2.07%	0.97	-5.20	3.92	-6.13	4.85
			HD-C	2.53	8.14%	0.61	-2.04	7.09	-2.96	8.01
			HD-LD	3.17	10.42%	0.45	-1.35	7.69	-2.27	8.61

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

For all tasks above, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the overall time to remove/mount the spare tire, the C, LD, and HD groups are all normally distributed with Shapiro-Wilk Test p-values of 0.35, 0.92, and 0.23, respectively. For the mount spare portion of the time, the C, LD, and HD groups are all normally distributed with Shapiro-Wilk Test p-values of 0.04, 0.23, and 0.37, respectively. For the re-mount original portion of the time, the C, LD, and HD groups are all normally distributed with Shapiro-Wilk Test p-values of 0.19, 0.81, and 0.79, respectively.

All times are expressed in man-minutes derived from the amount of time the Gunner and Driver required to complete the tire change and the addition of the time when the Vehicle Commander assisted in the process. Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. For the overall time to mount and remove the spare tire, the C group had a mean composite time of 61.84

minutes, which was faster than the LD and HD group mean composite times of 61.93 minutes and 65.95 minutes, respectively. No result is statistically significant. Results show differing levels of variance as shown by the SD column. See Table B-44 and Table B-45 for detailed analytical results.

B.5.12.1.3 Mount/Remove Spare Tire by Integration Level Contextual Comments

The observed output from this task contained a high level of variance, making it difficult to draw a definitive conclusion about this task. As one of the lengthiest subtasks, observed, times of integrated and non-integrated crews were largely the same, with C groups slighter faster when compared to HD groups. The ability to rapidly change a tire has critical implications to ensuring mission accomplishment, especially during long-range movements. Depending on the terrain, it would not be unusual for a vehicle crew to change multiple tires throughout the course of an operation.

B.5.12.1.4 Mount/Remove Spare Tire by Integration Level Additional Insights

Subjective comments and test manager observation reveal that a large portion of high-density crews and many low-density crews required assistance or struggled to remount the spare tire on the side of the vehicle. A large portion of the control-group crews performed a straight lift without assistance from the Vehicle Commander when mounting the spare tire. Vehicle Commander assistance or additional manpower to compensate during such tasks can detract from local security or concurrent actions. Compiled subjective comment results can be found in section 5.6.

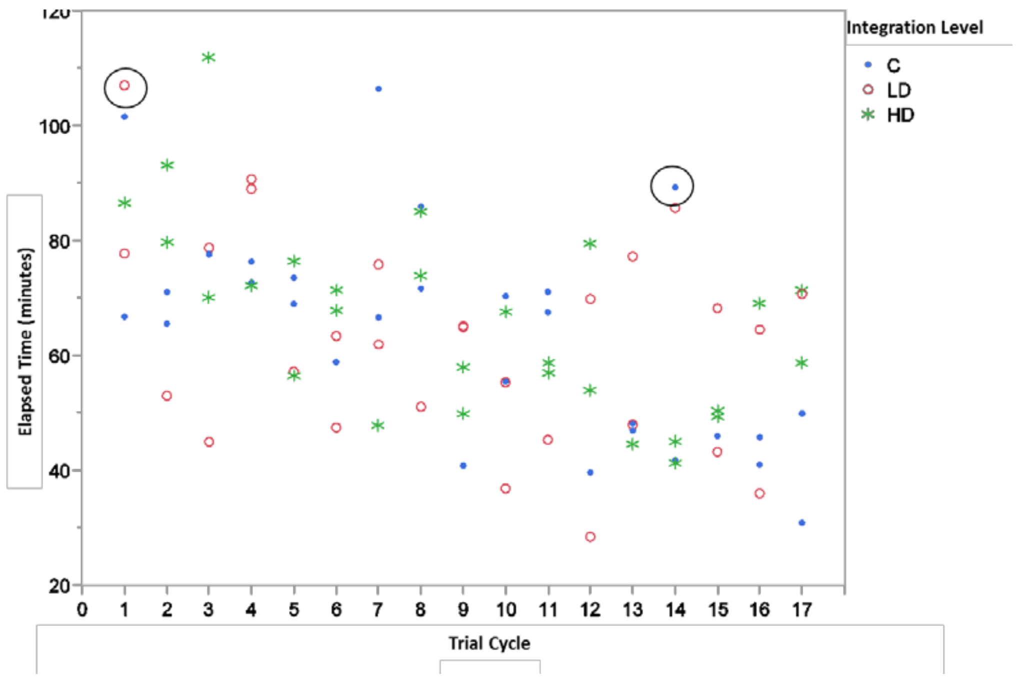
B.5.12.2 Mount/Remove Spare Tire by Integration Level (Excluding Potential Influential Points)

See B.5.18 for task description.

B.5.12.2.1 Mount/Remove Spare Tire by Integration Level (Excluding Potential Influential Points) Scatterplots

The scatter plots below display the results of Mount/Remove Spare Tire by integration level task. No outliers were identified for this task. Influential points are identified by lightly colored black circles.

Figure B-57. Mount/Remove Spare Tire (Excluding)



B.5.12.2.2 Mount/Remove Spare Tire by Integration Level (Excluding Potential Influential Points) Data and Analysis

The tables below summarize the results for the Mount/Remove Spare Tire task. The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table B-46. Mount/Remove Spare Tire (Excluding)

Remove/Mount Spare Tire **[excluding potential influential points]** (minutes)	C	25	60.74	17.62
	LD	26	60.20	16.32
	HD	28	65.95	16.58

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table B-47. Mount/Remove are Tire (Excluding) ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2- side P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Remove/Mount Spare Tire **[excluding infuential points]** (minutes)	0.97 (2, 76)	0.39	LD-C	-0.55	-0.90%	0.99	-8.71	7.62	-10.37	9.28
			HD-C	5.20	8.57%	0.50	-2.82	13.22	-4.45	14.85
			HD-LD	5.75	9.55%	0.43	-2.19	13.69	-3.80	15.30

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

For the overall time to remove/mount the spare tire, the C, LD, and HD groups are all normally distributed with Shapiro-Wilk Test p-values of 0.23, 0.90, and 0.23, respectively. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

When excluding potential influential results to accomplish the task, the C group mean time of 60.74 minutes was slower than the LD group time of 60.20 minutes, but faster than HD group composite time of 65.95 minutes. None of the results were statistically significant. There were no potential influential points identified for the mounting spare or remounting original metrics and they were thus omitted. Results show greater variance in the HD and C groups as shown by the SD column. See Table B-46 and Table B-47 for detailed analytical results.

B.5.12.2.3 Mount/Remove Spare Tire by Integration Level (Excluding Potential Influential Points) Additional Insights

Observations after excluding potential influential results remain largely the same.

B.6 Statistical Modeling Results

B.6.1 Statistical Modeling Results Overview

The previous section presented results in terms of integration levels and critical billets. This section describes statistical modeling, which applied here, is to estimate, simultaneously, the effect of gender-integration levels and other relevant variables on LAV crew performance. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For the selected tasks described in the previous section, this section presents an overview of the analysis and results, and then presents the modeling results for each task.

For each task, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates the increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time.

B.6.2 0313 Selected Tasks Method of Analysis

A linear mixed model was not suitable for these data because some volunteers only participated in a particular billet one to two times. Therefore, the 0313 selected tasks were modeled using Ordinary Least Squares regression. In addition to the Gunners' and Drivers' physical characteristics used as variables in the model, we also included a fixed effect for the VC to evaluate the leadership impact on task completion.

The statistical model included integration level and personnel data for each crewmember. Age, height, and other variables for each member of the crew and integration level were modeled with the result (response time) as the response variable. For tasks that had critical billets, i.e., only one or two Marines performing the task without help from the others, only their personnel variables and gender were included in the model. Where possible, a backward stepwise regression, using AIC, determined which variables were optimal in the model. In case of missing values, backward stepwise could not run and significant variables were automatically reported based on p-values from the overall model. Variables reported as significant were significant based on at least a one-sided test.

B.6.3 Selected Tasks Overall Modeling Results

There were no personnel data variables that were statistically significant and had a practical impact to the model. Each time personnel data variables are statistically significant in a model, their effects are practically negligible, conflicting, and/or incomplete for the crew; i.e., there are no tasks for which a variable is significant for all, or even most, members of the crew.

Refer to Section B.4 for the ANOVA summary for each below-mentioned 0313 task.

B.6.3.1 Manually Traverse and Elevate the LAV-25 Turret (Critical Billet)

We modeled elapsed time for the manually manipulating the turret as a function of the personnel variables and gender of the Gunner performing the task. The covariates in each model are the values of each personnel variable for each volunteer member in the LAV crew. We report statistically significant positive and negative correlations and whether we observe any patterns.

There was one potentially influential point for these results. However, investigation showed no effect on regression results so it was included in the analysis. The critical billet variable was eliminated with the AIC criterion during variable selection, meaning it wasn't a good predictor of elapsed time given the other variables in the model.

The models for the following variables do not run due to missing values:

- None.

The integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with manually manipulating the turret:

- Height
- AFTQ
- CFT MUF.

The following personnel variables are significant in their respective models and are negatively correlated with manually manipulating the turret:

- Age
- GT Score
- Weight
- CFT MTC
- PFT Crunch Score
- PFT Run Time.

The model identified a small effect of the VC with respect to leadership impact on task completion.

There are no patterns for any personnel variables for manually manipulating the turret. See section B.5.5.1.2 for the ANOVA summary of this task.

B.6.3.2 Prep M242 Main Gun (Critical Billet)

We model elapsed time for prepping the M242 Main Gun as a function of the personnel variables and gender of the Marine performing the task. The covariates in each model are the values of each personnel variable for the Gunner and Driver. We report statistically significant positive and negative correlations and whether we observe any patterns.

There were three potentially influential points in the analysis of this task. Running the models with and without these points gives slightly different results on the Gunner's and Driver's covariates. However, both models, when put through the AIC variable selection process, exclude the gender of the Gunner as a predictor. Because we consider the potentially influential points as valid data, and because their inclusion makes no difference on our variable of interest, we present the results of the model that include them in the analysis.

The models for the following variables do not run due to missing values:

- None.

The integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following Gunner personnel variables are significant in their respective models and are positively correlated with prepping the M242 Main Gun:

- Age
- GT
- CFT MTC
- PFT Crunch Score.

The following Gunner personnel variables are significant in their respective models and are negatively correlated with prepping the M242 Main Gun:

- Height
- AFQT
- Rifle Range Score.

There are no patterns for any personnel variables for prepping the M242 Main Gun. See section B.5.6 for the ANOVA summary of this task.

The VC's and Driver's personnel variables had no significant relationship with time to Prep M242 Main Gun. The opposing relationships between Gunner's GT and AFQT scores poses difficulties establishing any correlation between mental performance and task effectiveness. Anaerobic capability can be correlated to faster times to complete this task, as displayed by the correlation of the Gunner's Movement to Contact time and crunch score. The negative correlation with height may also be a determining factor due to the difficulty of lifting heavy objects in a confined space.

B.6.3.3 Conduct Casualty Evacuation

We model elapsed time for evacuating a casualty from the LAV as a function of each personnel variable for each crewmember and integration level. The covariates in each model are the values of each personnel variable for each volunteer member in the LAV crew. We report statistically significant positive and negative correlations and whether we observe any patterns.

The AIC criterion chooses a model in which gender integration is a significant variable. Specifically, the HD crews took, on average, 1 minute longer to complete the task than all-male crews, holding other variables constant. LD crews did not show a significant difference from C group crews.

The models for the following variables do not run due to missing values:

- None.

The integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following Gunner personnel variables are significant in their respective models and are positively correlated with evacuating a casualty from the LAV:

- Age
- CFT MTC.

The following Gunner personnel variables are significant in their respective models and are negatively correlated with evacuating a casualty from the LAV:

- GT
- Weight
- PFT Run Time
- PFT Crunch Score.

There are no patterns for any personnel variables for evacuating a casualty from the LAV time. See section B.5.2.2 for the ANOVA summary of this task

The VC's and Driver's personnel variables had no significant relationship with time to evacuate casualty from the LAV. Due to the opposing correlation between Gunner's CFT MTC and PFT run time, it is difficult to establish a correlation between physical performance variables and task effectiveness. The Gunner's weight could also be a determining factor on elapsed time to complete task since a heavier crewman could utilize more leverage when extracting a casualty from the turret.

B.6.3.4 Perform Main Gun Remedial Action (with assistance times)

We model elapsed time for performing remedial action on the main gun as a function of each personnel variable for each crewmember and integration level. The covariates in each model are the values of each personnel variable for each volunteer member in the LAV crew. We report statistically significant positive and negative correlations and whether we observe any patterns.

The task was examined by critical billet and integration level. In both models, the gender variables were eliminated with the AIC criterion during variable selection. The remaining results were similar between the two.

The models for the following variables do not run due to missing values:

- None.

The integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following Driver personnel variables are significant in their respective models and are positively correlated with performing remedial action on the main gun:

- Weight
- PFT Run Time.

The following personnel variables are significant in their respective models and are negatively correlated with performing remedial action on the main gun:

- Gunner Weight
- Gunner AFQT
- Gunner PFT Run Time
- Driver Age
- Driver CFT MTC
- Driver PFT Crunch Score.

There are no patterns for any personnel variables for performing remedial action on the main gun. See section B.5.9 for the ANOVA summary of this task.

The model identified a large effect of the VC with respect to leadership impact on task completion. Proficient VC’s with higher experience levels will have a positive effect on the overall time to complete this task, while the least experienced VC will yield a detrimental effect. The Gunner’s weight can be a determining factor as a heavier crewman could utilize more leverage when manipulating heavy objects in a confined space. Due to the opposing correlation between Driver's Movement to Contact time and PFT run time, it is difficult to establish a correlation between Driver's physical performance variables and task effectiveness.

B.6.3.5 Remove / Mount Spare Tire; with assistance time

We model elapsed time for removing/mounting a spare tire as a function of each personnel variable for each crewmember and integration level. The covariates in each model are the values of each personnel variable for each volunteer member in the LAV crew. We report statistically significant positive and negative correlations and whether we observe any patterns.

There were two potentially influential points for analysis in this task. Upon investigation, they did not make a sizeable difference on the model, so we report results with the points included. The AIC variable selection process yields a model that does not include integration level, meaning that in presence of additional variables in did not have a strong linear relationship with the time to complete this task.

The models for the following variables do not run due to missing values:

- None.

The integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with removing/mounting a spare tire:

- CFT MTC of the Gunner
- PFT Crunch Score of the Gunner
- Age of the Driver
- PFT Crunch Score of the Driver.

The following personnel variables are significant in their respective models and are negatively correlated with removing/mounting a spare tire:

- Weight of the Gunner
- CFT MUF of the Gunner,

- Rifle Range Score of the Gunner
- Height of the Driver
- Rifle Range Score of the Driver.

Due to the opposing correlation between Gunner's MTC and MANUF, it is difficult to establish a correlation between the Gunner's physical performance variables and task effectiveness. A Vehicle Commander's experience could correlate to faster times. Crunch score could correlate to better core strength when lifting and mounting the heavy tires. Additionally, height could be correlated to faster times when lifting the spare tire to mount on the side of the vehicle

There are no patterns for any personnel variables for removing/mounting a spare tire. See section B.5.12 for the ANOVA summary of this task.

B.6.3.6 Load & Stow Additional Ammunition

We model elapsed time for loading and stowing additional ammunition as a function of each personnel variable for each crew member and integration level. The covariates in each model are the values of each personnel variable for each volunteer member in the LAV crew. We report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with loading and stowing additional ammunition:

- AFQT score of the Driver

The following personnel variables are significant in their respective models and are negatively correlated with loading and stowing additional ammunition:

- GT of the Gunner
- Weight of the Gunner

- PFT Run Time of the Gunner
- Age of the Driver
- GT of the Driver
- CFT MTC of the Driver.

Due to the opposing correlation between Driver's GT and AFQT scores, it is difficult to establish a correlation between mental performance and task effectiveness. Anaerobic capability could be correlated to faster times during the execution of this task as displayed by the correlation of the Gunner's PFT run time and the Driver's CFT MTC.

There are no patterns for any personnel variables for loading and stowing additional ammunition. See Section B.5.7 for the ANOVA summary of this task.

B.6.3.7 Engage Defensive Targets (#2) – Manual (Critical Billet)

We model elapsed time for engaging defensive targets as a function of the personnel variables and gender of the Marine performing the task. The covariates in each model are the values of each personnel variable for each volunteer member in the LAV crew. We report statistically significant positive and negative correlations and whether we observe any patterns.

For this task, the measurement of interest is time to kill. The data are right censored because crews exceeded the maximum time of 80 seconds to complete the task. Therefore, we modeled time to kill using a Cox Proportional Hazards Model. The result of the model shows that gender of the Gunner has no significant effect.

The models for the following variables do not run due to missing values:

- None.

The integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with engaging defensive targets:

- Age of the Gunner
- Rifle Score of the Gunner,

- Height of the Driver
- CFT MTC of the Driver
- AFQT of the Driver.

The following personnel variables are significant in their respective models and are negatively correlated with engaging defensive targets:

- Weight of the Gunner
- Age of the Driver
- CFT MTC of the Driver
- PFT Crunch Score of the Driver.

There were also no statistically significant differences in first hit success rates when accounting for the gender of the Marine in the Gunner billet

There are no patterns for any personnel variables for engaging defensive targets. See section B.5.8 for the ANOVA summary of this task.

B.6.3.8 Prep M240 Coaxial Machine Gun (Critical Billets)

We model elapsed time for preparing M240 Coaxial Machine Gun as a function of the personnel variables and gender of the Marine performing the task. The covariates in each model are the values of each personnel variable for the Gunner. We report statistically significant positive and negative correlations and whether we observe any patterns.

Including the influential points into analysis yields a statistically significant effect of the gender of the Marine in the Gunner billet (women are almost 4 minutes slower, on average).

Excluding the influential points from analysis yields a smaller but still significant effect of a female Gunner (1.56 minutes slower). The signs of the remaining effect stay the same as in the previous model, except the sign on the PFT run time becomes negative.

The models for the following variables do not run due to missing values:

- None.

The integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following Gunner personnel variables are significant in their respective models and are positively correlated with preparing M240 Coaxial Machine Gun:

- Age
- GT
- CFT MTC
- PFT Crunch Score.

The following Gunner personnel variables are significant in their respective models and are negatively correlated with preparing M240 Coaxial Machine Gun:

- AFQT
- Rifle Range Score.

There are no patterns for any personnel variables for preparing M240 Coaxial Machine Gun. See section B.5.4 for the ANOVA summary of this task.

Appendix to Annex B
0313 Supplemental Information

This appendix provides supplemental information for the 0313 portion of the GCEITF experiment. It provides information regarding the GCEITF subjective comments and additional descriptive and basic inferential statistics not described in Annex B.

Section 1: Subjective Comments

Subjective comments were gathered and reported from daily observation and interaction with volunteers during the execution of the experiment. Table B A displays a summary of these comments broken down by task, integration level, gender, and type of comment. Comments were both positive and negative depending on the observation. For example, the table shows that 2 males in a C group and 14 females in an HD group executing the Mount/Remove Spare Tire task required extra assistance to complete the task.

Table B A - Summary Subjective Comments

Task Description	Gender	Cohesion				Extra Assistance				Failed to do task properly				Marine avoiding task				Needs No Assistance				Stamina and Strength				Grand Total
		CG	LD	HD	Total	CG	HD	LD	Total	CG	HD	LD	Total	CG	HD	LD	Total	CG	HD	LD	Total	CG	LD	HD	Total	
Casevac	Female	2	0	0	2	1	2	2	5	3	2	0	5	0	1	0	1	1	1	0	2	1	0	0	1	16
	Male	0	0	0	0	1	0	0	1	1	1	2	4	0	0	0	0	13	0	2	15	0	0	0	0	20
	M/F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Live Fire Sub-Tasks	Female	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	M/F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Load and Stow Additional Ammo	Female	0	0	0	0	0	1	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	Male	0	0	0	0	2	0	0	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3
	M/F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maintenance Day SubTasks	Female	0	0	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	Male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	2
	M/F	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Mount/Remove Spare Tire	Female	0	0	0	0	2	14	2	18	0	0	0	0	0	0	0	0	0	9	0	9	0	0	0	0	27
	Male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	11	0	0	0	0	11
	M/F	0	0	0	0	0	1	3	4	0	0	0	0	0	0	0	0	0	0	14	14	0	0	0	0	18
Prep For Combat	Female	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2
	Male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	M/F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Prep M242 Main Gun	Female	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
	Male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	M/F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rig for Recovery	Female	0	0	1	1	0	8	1	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
	Male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1
	M/F	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Rig For Tow	Female	0	0	0	0	0	5	2	7	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	8
	Male	0	0	0	0	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	M/F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Grand Total	2	0	1	3	6	37	16	59	5	3	2	10	0	1	2	3	25	10	17	52	2	2	0	4	131

Section 2: Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for additional 0313 tasks. Annex B contains the descriptive statistics for the remainder of the 0313 tasks. The words “metric” and “task” are used interchangeably throughout this Appendix. They both refer to the experimental task.

The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Section 3: Additional Task Results

Table B B – 0313 Selected Task Results

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Evac Operator Stations (seconds)*	C	29	23.17	4.44	3.72%	14.20%	10.10%
	LD	28	24.04	4.98			
	HD	28	26.46	6.56			
Upload Main Gun (minutes)	C	29	1.52	1.00	3.74%	-13.10%	-16.24%
	LD	25	1.58	0.78			
	HD	27	1.32	0.57			
Upload M240 (minutes)	C	29	1.59	0.73	8.81%	-14.15%	-21.10%
	LD	27	1.73	0.80			
	HD	28	1.36	0.47			
Engage Defensive Targets (#1) (seconds)	C	27	37.67	16.98	-4.11%	1.31%	5.65%
	LD	25	36.12	16.99			
	HD	25	38.16	22.20			
Engage Offensive Targets (#1) (seconds)	C	27	15.26	4.39	-3.53%	5.91%	9.79%
	LD	25	14.72	4.16			
	HD	25	16.16	7.80			
Engage Offensive Targets (#2) (seconds)	C	27	14.30	5.74	3.79%	15.55%	11.32%
	LD	25	14.84	7.79			
	HD	25	16.52	6.21			

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean or percent hits values for the Integration Level

Table B C - 0313 Selected Task ANOVA and T-test Results

Metric	F statistic (df)	F Test P-value	Comparison	Difference	% Difference	2-sided p-Value	80% LCB	80% UCB	90% LCB	90% UCB
Evac Operator Stations (seconds)*	2.84 (2, 82)	0.06*	LD-C	0.86	3.72%	0.82	-1.61	3.34	-2.11	3.84
			HD-C	3.29	14.20%	0.06*	0.82	5.76	0.32	6.27
			HD-LD	2.43	10.10%	0.22	-0.07	4.92	-0.57	5.43
Engage Defensive Targets (#1) (seconds)	0.08 (2, 74)	0.92	LD-C	-1.55	-4.11%	0.95	-10.60	7.51	-12.45	9.35
			HD-C	0.49	1.31%	0.99	-8.56	9.55	-10.41	11.39
			HD-LD	2.04	5.65%	0.92	-7.19	11.27	-9.07	13.15
Engage Offensive Targets (#1) (seconds)	0.41 (2, 74)	0.66	LD-C	-0.54	-3.53%	0.94	-3.26	2.19	-3.82	2.74
			HD-C	0.90	5.91%	0.83	-1.82	3.63	-2.38	4.18
			HD-LD	1.44	9.79%	0.64	-1.34	4.22	-1.90	4.78
Engage Offensive Targets (#2) (seconds)	0.79 (2, 74)	0.46	LD-C	0.54	3.79%	0.95	-2.64	3.72	-3.28	4.37
			HD-C	2.22	15.55%	0.45	-0.96	5.40	-1.60	6.05
			HD-LD	1.68	11.32%	0.64	-1.56	4.92	-2.22	5.58

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean or percent hits values for the Integration Level

Evacuate Operator Stations. All crewmen (VC, Gunner, and Driver) were required to exit the vehicle safely and move as a crew to the rally point 25 m away. Marines wore fighting loads and carried individual weapons, as would be expected when evacuating a vehicle in a tactical environment. Although the crew evacuation subtask was not as physically demanding as extracting a casualty from an LAV-25 turret, the ability to evacuate a vehicle quickly is important in the event of a vehicle fire or other life-threatening issue.

For this task, the C and HD groups are normally distributed as evidenced by a Shapiro-Wilk Test p-value of 0.16 and 0.01 respectively, while the LD group is not normally distributed with Shapiro-Wilk p-value of less than 0.01. We proceed with presenting the ANOVA results shown above which are confirmed by a Kruskal-Wallis Test (p-value = 0.10). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

For the overall time required to evacuate the vehicle, the C group had a mean time of 23.17 seconds, which was statistically significant and faster when compared to the HD group time of 26.46 seconds. The C group's 14.20% faster mean time when compared to the HD group was statistically significant. Additionally, the LD and HD groups show greater variability than the C group as shown by the Standard Deviation (SD) column.

- **Contextual Comments**. In the event of a vehicle casualty, egress from the vehicle is essential. Longer times to egress could lead to greater exposure to various threats, such as a vehicle fire.
- **Additional Insights.** The time differences between groups in this task are very small. When viewed as a standalone event, the differences are minor. When viewed in conjunction with additional requirements or tasks, such differences have the potential to accumulate and cause greater mission effects.

Upload Main Gun. This subtask required the LAV-25 crew to upload the HE and AP rounds from their respective ready-boxes into the M242 main gun. This was accomplished by ratcheting the rounds from the ready box, through the gun's feed chutes, and into the feeder assembly on the gun. The AP ready-box is located near the Driver, while the HE ready-box is located next to the Gunner. Without assistance, a crew must upload their weapons systems as part of normal combat preparation.

The C group had a mean time of 1.52 minutes, which was faster than the LD group time of 1.58 minutes but slower than the HD group of 1.32 minutes. None of the results were statistically significant. Additionally, the C group shows greater variability than the LD and HD groups as shown by the Standard Deviation (SD) column.

- **Additional Insights.** The observed output from this task contained a high level of variance, and because of this it is difficult to draw a definitive conclusion. After it has been loaded into the vehicle, uploading ammunition requires little physical exertion but remains a key component of combat preparation.

Upload M240 Coax. This subtask required the LAV-25 Gunner, without assistance, to upload a belt of 200 7.62-mm rounds from coax ready box into the M240 coaxially mounted machine gun. This was accomplished by feeding the rounds through the feed chute that extends across the turret and onto the feed tray of the M240 coax. Without assistance, a crew must upload their weapons systems as part of normal combat preparation.

The C group had a mean time of 1.59 minutes, which was faster than the LD group time of 1.73 minutes but slower than the HD group time of 1.36 minutes. None of the results were statistically significant. Additionally, groups showed differing levels of variability as shown by the Standard Deviation (SD) column.

- **Additional Insights.** Like the Upload Main Gun task, these results show a very high level of variability and make it difficult to draw any definitive conclusions. As stated above, uploading ammunition requires little physical exertion but remains a key component of combat preparation.

Defensive Engagement #1. The first live-fire engagement was conducted from a defensive battle position. The crews engaged two vehicle targets, one with AP and one with HE ammunition. The time to engage each target and number of bursts required to kill each target was recorded. Crews had 80 seconds to acquire, identify, engage, and destroy the target. Live fire engagements were adapted from MCWP 3-14.1 Table VI gunnery qualification engagements, which all vehicle crews are required to fire to achieve qualified status.

For this task, the HD group is normally distributed as evidenced by a Shapiro-Wilk Test p-value of 0.06, while the C and LD groups are not normally distributed with Shapiro-Wilk p-values of less than 0.01. We proceed with presenting the ANOVA results shown above which are confirmed by a Kruskal-Wallis Test (p-value = 0.92). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The C group had a mean time of 37.67 seconds, which was slower than the LD group time of 36.12 seconds but faster than the HD group time of 38.16 seconds. None of the results were statistically significant. Additionally, the HD group showed greater variability than the C and LD groups as shown by the Standard Deviation (SD) column.

- **Additional Insights.** The results above indicate the time required to initially acquire and begin to engage identified targets. These results show a very high level of variability and make it difficult to draw any definitive conclusions. A high pass rate and a proficient Gunner is desired to ensure a vehicle crew's effectiveness.

Offensive Engagement #1. The first live-fire offensive engagement was conducted on the move through a maneuver box. The crews engaged one vehicle target with the LAV-25 main gun and a bank of troop targets with the coaxially mounted machine gun. The time to engage each target and number of bursts required to kill the main gun target was recorded. Live fire engagements were adapted from MCWP 3-14.1 Table VI gunnery qualification engagements, which all vehicle crews are required to fire to achieve qualified status.

For this task, the C and LD groups are normally distributed as evidenced by Shapiro-Wilk Test p-values of 0.02 and 0.44, respectively, while the HD group is not normally distributed with Shapiro-Wilk p-value of less than 0.01. We proceed with presenting the ANOVA results shown above which are confirmed by a Kruskal-Wallis Test (p-value = 0.93). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The C group had a mean time of 15.26 seconds, which was slower than the LD group time of 14.72 seconds but faster than the HD group time of 16.16 seconds. None of the results were statistically significant. Additionally, the HD group showed greater variability than the C and LD groups as shown by the Standard Deviation (SD) column.

- **Additional Insights.** The results above indicate the time required to initially acquire and begin to engage identified targets. These results show a very high level of variability and make it difficult to draw any definitive conclusions. A high pass rate and a proficient Gunner is desired to ensure a vehicle crew's effectiveness.

Offensive Engagement #2. The second live-fire offensive engagement was conducted on the move through a maneuver box as a retrograde engagement. With the turret oriented over the rear end of the vehicle, the crews engaged one vehicle target with the LAV-25 main gun and a bank of troop targets with the coaxially mounted machine gun. The time to engage each target and number of bursts required to kill the main gun target was recorded. Live fire engagements were adapted from MCWP 3-14.1 Table VI gunnery qualification engagements, which all vehicle crews are required to fire to achieve qualified status.

For this task, the C, LD, and HD groups are all not normally distributed as evidenced by Shapiro-Wilk Test p-values of less than 0.01. We proceed with presenting the ANOVA results shown above which are confirmed by a Kruskal-Wallis Test (p-value = 0.12). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The C group had a mean time of 14.30 seconds which was faster than the LD group time of 14.84 seconds and the HD group time of 16.52 seconds. None of the results were statistically significant. Additionally, groups showed differing levels of variability as shown by the Standard Deviation (SD) column.

- **Additional Insights.** The results above indicate the time required to initially acquire and begin to engage identified targets. These results show a very high level of variability and make it difficult to draw any definitive conclusions. A high pass rate and a proficient Gunner is desired to ensure a vehicle crew's effectiveness.

Vehicle Recovery and Tow Operations Overview. Recovery operations are an integral part of an LAR unit's mobility and versatility when operating independently. Should a vehicle become mired or lose power, crews must quickly and efficiently work together to perform this critical skill. All recovery tasks were conducted as crew events.

Rig for Recovery. This sub-task required the LAV-25 crew to utilize the winch mounted on the front of the vehicle and the snatch blocks mounted on the side of the LAV-25 and the LAV Logistics Variant (LAV-L). With the two vehicles spaced ten meters apart, the crewmen would pay out the winch cable and utilize two snatch blocks, each weighing 75 lbs., to create a three-point recovery system with the shackles and tow-points of the vehicles. The ability to quickly recover a mired vehicle is an important skill that all crewmen must possess. A three point recovery system is often used as it only requires two vehicles to establish and can be accomplished fairly quickly. Crews were evaluated on the time required to set-up the three point recovery as well as the time to break down and stow all of the tools and equipment used in the process.

Rig for Recovery by Integration Level. The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Table B D - Rig for Recovery by Integration Level

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Rig for Recovery; Rig and Stow (minutes)	C	29	13.76	3.87	-6.07%	7.89%	14.86%
	LD	27	12.92	2.56			
	HD	28	14.84	3.82			
Rig for Recovery; Rig Only (with remediation) (minutes)	C	28	5.62	1.92	7.60%	18.04%	9.71%
	LD	28	6.05	2.57			
	HD	28	6.63	1.82			
Rig for Recovery; Stow Only (minutes)	C	27	7.23	2.27	-7.98%	7.44%	16.76%
	LD	27	6.65	1.26			
	HD	28	7.77	2.57			

Table B E - Rig for Recovery by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-sided p-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Rig for Recovery; Rig and Stow (minutes)	2.11 (2, 81)	0.13	LD-C	-0.83	-6.07%	0.64	-2.45	0.78	-2.77	1.10
			HD-C	1.09	7.89%	0.47	-0.51	2.68	-0.84	3.01
			HD-LD	1.92	14.86%	0.11	0.29	3.55	-0.04	3.88

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests.

For the overall time to rig for recovery, the C, LD, and HD groups are all normally distributed as evidenced by Shapiro-Wilk Test p-values of 0.53, 0.12, and 0.12, respectively. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the overall time to rig for recovery the C group had a mean time of 13.76 minutes, which was slower than the LD group mean time of 12.92 minutes but faster than the HD group mean time of 14.84 minutes. None of the results are statistically significant. Results show differing levels of variance as shown by the Standard Deviation (SD) column.

- **Contextual Comments.** Crews will almost always conduct a vehicle recovery under some sort of duress, as a mired vehicle has the potential to severely impact timelines and hinder mission accomplishment. Thus, vehicle crews must be able to execute recoveries as quickly as possible. While not necessarily retaining the urgency of a casualty evacuation or live fire task, recovery is an essential skill each crewman must possess. MCWP 3-14, *Employment of the Light Armored Reconnaissance Battalion*, describes the importance of maintenance and self-recovery, and ensuring vehicles are properly configured to conduct recovery operations.
- **Additional Insights.** During this task, Vehicle Commanders had a large influence on the performance of their crew, regardless of gender composition. The performance of vehicle crews was also highly variable for this task, leading to our failure to reject the null hypothesis. Observed times of integrated and non-integrated crews were largely the same, with C groups performing slightly faster than HD groups.

Rig for Recovery by Integration Level (Excluding Potential Influential Points). The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Table B F - Rig for Recovery by Integration Level (Excluding)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Rig for Recovery; Rig Only (with remediation) **[excluding potential influential points]** (minutes)*	C	27	5.4	1.56	11.92%	22.79%	9.71%
	LD	28	6.05	2.57			
	HD	28	6.63	1.82			
Rig for Recovery; Stow Only **[excluding potential influential points]** (minutes)	C	27	7.23	2.27	-7.98%	4.64%	13.72%
	LD	27	6.65	1.26			
	HD	27	7.56	2.38			

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. When excluding potential influential results for rig time, the C group had a mean time of 5.40 minutes, which was faster than the LD group and HD group mean times of 6.05 minutes and 6.63 minutes, respectively. The 22.79% difference between the C group and HD group during disassembly was the only statistically significant result. There were no potential influential points identified for the overall time metric and it was thus omitted. Results show differing levels of variance as shown by the Standard Deviation (SD) column.

- **Additional Insights.** Observations after excluding potential influential results remain largely the same, with the exception of statistically significant results for the time to rig metric comparison between C and HD groups. C groups performed slightly faster than HD groups.

Rig for Tow Overview. This sub-task required the LAV-25 crew to utilize the LAV tow bar, weighing approximately 175 lbs., located on the side of the LAV-L. With the two vehicles spaced ten meters apart, the crewmen would detach the tow bar from the LAV-L and connect it to the tow points on the LAV-25. They would then ground guide the LAV-L and, with one crewman holding the tow bar, attach the eye of the tow bar to the tow pintle of the LAV-L. The subtask was completed when all tools and equipment were returned to their proper locations. Vehicles will often break down, either due to mechanical issues or enemy action, and the ability to quickly rig a vehicle for towing directly affects combat effectiveness. Crews were evaluated on the time required to set-up the tow configuration as well as the time to break down and stow all of the tools and equipment used in the process

Rig for Tow by Integration Level. The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Table B G - Rig for Tow by Integration Level

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Rig for Tow; Rig and Stow (minutes)	C	29	9.99	2.13	6.89%	6.27%	-0.58%
	LD	28	10.68	2.61			
	HD	28	10.62	2.72			
Rig for Tow; Rig Only (minutes)	C	29	5.92	1.33	7.10%	5.81%	-1.2%
	LD	28	6.34	1.49			
	HD	28	6.26	1.73			
Rig for Tow; Stow Only (minutes)	C	28	3.8	0.92	6.99%	8.36%	1.27%
	LD	28	4.07	1.35			
	HD	28	4.12	1.17			

Table B H - Rig for Tow by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-sided p-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Rig for Tow; Rig and Stow (minutes)	0.67 (2, 82)	0.52	LD-C	0.69	6.89%	0.55	-0.46	1.83	-0.69	2.06
			HD-C	0.63	6.27%	0.61	-0.52	1.77	-0.75	2.00
			HD-LD	-0.06	-0.58%	1.00	-1.22	1.09	-1.45	1.33

Statistical analysis included ANOVA and multiple comparisons using Tukey tests.

For the overall time to rig for tow, the C, LD, and HD groups are all normally distributed as evidenced by Shapiro-Wilk Test p-values of 0.78, 0.23, and 0.07, respectively. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the overall time to rig for tow the C group had a mean time of 9.99 minutes, which was faster than the LD and HD group mean times of 10.68 minutes and 10.62 minutes, respectively. None of the results are statistically significant. Additionally, the HD and LD groups show greater variability when compared to the C group as shown by the Standard Deviation (SD) column.

- **Contextual Comments.** The context for this sub-task is very similar to the rig for recovery sub-task discussed above. The ability to rapidly rig a downed vehicle for tow is a crucial skill that a vehicle crew must retain.
- **Additional Insights.** During this task, Vehicle Commanders had a large influence on the performance of their crew, regardless of gender composition. The performance of vehicle crews was also highly variable for this task, leading to our failure to reject the null hypothesis. Observed times of integrated and non-integrated crews were largely the same, with C groups performing slightly faster than HD groups.

Rig for Tow by Integration Level (Excluding Potential Influential Points). The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Table B I - Rig for Tow by Integration Level (Excluding)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Rig for Tow; Rig and Stow **[excluding potential influential points]** (minutes)	C	29	9.99	2.13	6.89%	3.13%	-3.52%
	LD	28	10.68	2.61			
	HD	27	10.31	2.19			
Rig for Tow; Rig Only (with remediation) **[excluding potential influential points]** (minutes)	C	29	5.92	1.33	7.10%	3.09%	-3.74%
	LD	28	6.34	1.49			
	HD	27	6.1	1.53			
Rig for Tow; Stow Only (with remediation) **[excluding potential influential points]** (minutes)	C	27	3.7	0.76	6.96%	7.43%	0.44%
	LD	26	3.96	1.2			
	HD	27	3.98	0.91			

Table B J - Rig for Tow by Integration Level (Excluding) ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided p-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Rig for Tow; Rig and Stow **[excluding potential influential points]** (minutes)	0.63 (2, 81)	0.54	LD-C	0.69	6.89%	0.50	-0.38	1.75	-0.59	1.97
			HD-C	0.31	3.13%	0.87	-0.76	1.39	-0.98	1.60
			HD-LD	-0.38	-3.52%	0.82	-1.46	0.71	-1.68	0.93

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests.

For the overall time to rig for tow (excluding influential points), the C, LD, and HD groups are all normally distributed as evidenced by Shapiro-Wilk Test p-values of 0.78, 0.23, and 0.34, respectively. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. When excluding potential influential results for the overall time to rig for tow, the C group had a mean time of 9.99 minutes, which was faster than the LD group and HD group mean times of 10.68 minutes and 10.31 minutes, respectively. None of the results were statistically significant. Additionally, the HD and LD groups show greater variability when compared to the C group as shown by the Standard Deviation (SD) column.

- **Additional Insights.** Observations after excluding potential influential results remain largely the same, with the C group performing slightly faster than LD and HD groups.

Conduct Maintenance Actions Overview. While a maintenance task does not always convey the same sense of urgency typically found in a casualty evacuation or live fire task, they remain no less important. The ability to effectively perform continuing maintenance actions is a vital necessity of sustained operations. Vehicles require extensive preventative maintenance even in garrison, and austere terrain and expeditionary operating conditions only magnify these demands. This work is often physically demanding because the parts, tools, and equipment organic to LAR units are large and heavy. The ability of a vehicle crew to work with their maintainers to keep their vehicles in the fight is an everyday necessity. The maintenance actions portion of the non-live fire trial day consisted of three separate subtasks; remove scout hatch armor, remove side panel armor, and a tire change. Each of these tasks was evaluated on the time to reach the mid-point of the task and the overall time to completion.

The vehicle crew used the tools and equipment (SL-3) associated with the LAV-25 to remove each of the armor panels mounted on the back hatches of the vehicle. Each panel is a solid piece of composite armor weighing approximately 125 lbs. The crew would drop and remount first one hatch panel, followed by the second. Crews were evaluated on the time required to mount and remove each of the scout hatch panels.

Mount/Remove Scout Hatch Armor by Integration Level. The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Table B K - Mount/Remove Scout Hatch Armor by Integration Level

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Scout Hatch Panels (minutes)*	C	29	6.56	2.56	2.23%	28.91%	26.1%
	LD	27	6.71	2.25			
	HD	28	8.46	2.6			
Scout Hatch Panels; 1st Panel removal & remount (minutes)*	C	29	3.28	1.26	9.88%	33.23%	21.25%
	LD	27	3.6	1.41			
	HD	28	4.36	1.57			
Scout Hatch Panels; 2nd Panel removal & remount (minutes)*	C	29	3.29	1.5	-5.38%	24.62%	31.7%
	LD	27	3.11	1.11			
	HD	28	4.1	1.38			

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level

Table B L - Mount/Remove Scout Hatch Armor by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided p-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Scout Hatch Panels (minutes)	5.1 (2, 81)	< 0.01*	LD-C	0.15	2.23%	0.97	-1.00	1.29	-1.23	1.53
			HD-C	1.90	28.91%	0.01*	0.76	3.03	0.53	3.26
			HD-LD	1.75	26.10%	0.03*	0.59	2.91	0.36	3.14

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level

Statistical analysis included ANOVA and multiple comparisons using Tukey tests.

For the overall time to mount and remove scout hatch panels, the C, LD, and HD groups are all normally distributed as evidenced by Shapiro-Wilk Test p-values of 0.18, 0.04, and 0.28, respectively. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For the overall time required to mount and remove the scout hatch panels the C group had a mean time of 6.56 minutes, which was statistically significantly faster than the HD groups that recorded a mean time of 8.46 minutes. The C group's 28.91% faster mean time when compared to the HD group was statistically significant. Results show differing levels of variance as shown by the Standard Deviation (SD) column.

- **Contextual Comments.** Removing armor panels, while not conducted as regularly as other maintenance actions, is required should the panels become damaged or during wash-downs and cleaning after extended operations. Longer times to conduct regular or routine maintenance could lead to lower readiness levels.
- **Additional Insights.** The statistical significance of the large performance gap between low density and high density crews suggests that the other crew members, specifically males, are less able to compensate for their counterpart's performance at higher integration levels. In real world operations, this difference

can translate into increased times required to prepare for a mission, perform repairs in the midst of an operation, or perform maintenance after an operation.

Mount/Remove Scout Hatch Armor by Integration Level (Excluding Potential Influential Points). The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Table B M - Mount/Remove Scout Hatch Armor by Integration Level (Excluding)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Scout Hatch Panels2nd Panel removal & remount **[excluding potential influential points]** (minutes) *	C	29	3.29	1.5	-9.23%	20.16%	32.38%
	LD	26	2.98	0.91			
	HD	27	3.95	1.17			

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests. When excluding potential influential results for 2nd scout hatch time, the C group had a mean time of 3.29 minutes, which was faster than the HD group time of 3.95 minutes but slower than the LD group mean time of 2.98 minutes. The 32.38% difference between the LD group and HD group was the only statistically significant result.

- **Additional Insights.** Observations after excluding potential influential results remain largely the same, although it places the C-HD group comparison just outside the threshold for statistical significance.

Mount/Remove Side-Panel Armor Overview. The vehicle crew used the SL-3 associated with the LAV-25 to remove each of the front-most armor panels mounted on the left and right sides of the vehicle. Each panel is a solid piece of composite armor weighing approximately 60 lbs. The crew would drop and remount first one panel, followed by the second. Crews were evaluated on the time required to mount and remove each of the armor panels.

Mount/Remove Side-Panel Armor by Integration Level. The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Table B N - Mount/Remove Side Panel Armor by Integration Level

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Front Side Armor Panels (minutes)	C	29	3.88	1.78	4.80%	11.86%	6.74%
	LD	28	4.07	1.57			
	HD	28	4.34	1.16			
Front Side Armor Panels; 1st panel removal and remount (minutes)	C	29	2.19	1.33	0.34%	0.66%	0.33%
	LD	28	2.2	1.06			
	HD	28	2.2	0.58			
Front Side Armor Panels; 2nd panel removal and remount (minutes)	C	29	1.62	0.71	5.96%	23.1%	16.18%
	LD	28	1.72	0.71			
	HD	28	2	0.91			

Table B O - Mount/Remove Side Panel Armor by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided p-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Front Side Armor Panels (minutes)	0.65 (2, 82)	0.52	LD-C	0.19	4.80%	0.89	-0.52	0.89	-0.66	1.03
			HD-C	0.46	11.86%	0.49	-0.24	1.16	-0.38	1.30
			HD-LD	0.27	6.74%	0.78	-0.43	0.98	-0.58	1.13

Statistical analysis included analysis of variance (ANOVA) and multiple comparisons using Tukey tests.

For the overall time to mount and remove side panels, the C and LD groups are not normally distributed as evidenced by Shapiro-Wilk Test p-values of less than 0.01, while the HD is normally distributed with Shapiro-Wilk p-value of 0.16. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.10). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

- **Contextual Comments.** Removing armor panels, while not conducted as regularly as other maintenance actions, is required should the panels become damaged or during wash-downs and cleaning after extended operations.

- **Additional Insights.** The observed output from this task contained a high level of variance, and because of this it is difficult to draw a definitive conclusion. As discussed during the scout hatch armor sub-task, longer times to conduct regular or routine maintenance could lead to lower readiness levels. In real world operations, this difference can translate into increased times required to prepare for a mission, perform repairs in the midst of an operation, or perform maintenance after an operation.

Mount/Remove Side-Panel Armor by Integration Level (Excluding Potential Influential Points). The tables below display results, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Also ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Table B P - Mount/Remove Side Panel Armor (Excluding)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Front Side Armor Panels **[excluding potential influential points]** (minutes)	C	28	3.67	1.37	5.47%	18.45%	12.31%
	LD	27	3.87	1.17			
	HD	26	4.34	1.16			
Front Side Armor Panels; 1st panel removal and remount **[excluding potential influential points]** (minutes)	C	28	1.91	0.67	6.77%	15.7%	8.37%
	LD	27	2.03	0.62			
	HD	26	2.2	0.58			
Front Side Armor Panels; 2nd panel removal and remount **[excluding potential influential points]** (minutes)	C	28	1.55	0.6	6.94%	24.69%	16.6%
	LD	27	1.65	0.64			
	HD	26	1.93	0.87			

Table B Q - Mount/Remove Side Panel Armor (Excluding)ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided p-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Front Side Armor Panels **[excluding potential influential points]** (minutes)	2.19 (2, 80)	0.12	LD-C	0.20	5.47%	0.56†	-0.24†	0.65	-0.38	0.78
			HD-C	0.68	18.45%	0.05†	0.24	1.12	0.11	1.25
			HD-LD	0.48	12.31%	0.14†	0.07	0.89	-0.05	1.00

†Due to lack of normality, p-values and confidence intervals have been replaced with Mann-Whitney Test results.

For the overall time to mount and remove side panels (excluding influential points), the C and LD groups are not normally distributed as evidenced by Shapiro-Wilk Test p-values of less than 0.01, while the HD is normally distributed with Shapiro-Wilk p-value of 0.16. Group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA, however, we recommend using Kruskal-Wallis results shown above due to lack of normality and conflicting results with the standard ANOVA. When excluding potential influential results for the overall time, the C group had a mean time of 3.67 minutes, which was faster than the LD and HD group times of 3.87 minutes and 4.34 minutes, respectively. None of the results comparisons were statistically significant. Results show differing levels of variance as shown by the Standard Deviation (SD) column.

- **Additional Insight.** Observations after excluding potential influential results remain largely the same.

THIS PAGE INTENTIONALLY LEFT BLANK

Annex C.
Infantry Machine Gunner (MOS 0331)

This annex details the Infantry Machine Gunner (MOS 0331) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 2 March – 26 April 2015 at Range 107 and Range 110, aboard the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, CA. The sections outline the Infantry Machine gunner Scheme of Maneuver (SOM), Limitations, Deviations, Data Set Description, Descriptive and Basic Inferential Statistics, and Modeling Results.

C.1 Scheme of Maneuver

C.1.1 Experimental Cycle Overview

The Infantry Machine Gunner (MOS 0331) assessment of the GCEITF took place in a field environment aboard MCAGCC, Twentynine Palms, CA. The assessment consisted of 21 trial cycles, each of which was a 2-day test cycle, conducted over the course of 55 days. the Marines spent 1 recovery day at Camp Wilson After every 4 days of trials. Every machinegun squad consisted of three volunteers and a direct-assignment (non-volunteer) squad leader. Each member of the squad was trained to fill each billet within the squad: gunner, assistant gunner, and ammo man. For the sake of consistency throughout the report, the term “squad” will be used for the M240B medium machinegun squad. The assessment was executed under the supervision of MCOTEA functional test managers and a range Officer in Charge (OIC)/Range Safety Officer (RSO) from the GCEITF.

C.1.2 Experimental Details

The 2-day 0331 assessment replicated offensive and defensive tasks. The 0331s began each cycle on the Day 1/Offensive task. Three 0331 squads executed each trial cycle: a control (C) non-integrated group, a low-density (LD) integrated group with one female, and a high-density (HD) integrated group with two females.

Day 1 of the trial cycle was executed on Range 107 and consisted of supporting a squad attack from the support-by-fire (SBF) position then displaced to the limit of advance (LOA). The Machine gunners repelled an enemy counterattack by fire for 90 seconds. Finally, each squad conducted a 100-meter CASEVAC of a 220-lb dummy.

Day 2 of the trial cycle was executed on Range 110 and consisted of defensive actions. The day started with a 7-km forced march from Range 107 to Range 110 wearing an approach load and carrying personal weapons, crew-served weapons, and ammunition. Each heavy machinegun squad prepared a M2 heavy machine gun for employment. The squads engaged three targets based on a prescribed course of fire with 400 rounds of .50-cal ammunition. Immediately upon the gun going out of action, the machinegun squad displaced the machine gun to a designated location. The trials concluded with a mount and dismount of the M2 machine gun to a HMMWV. At the conclusion of Day 2, the Marines reorganized into new squads for the next experimental cycle.

C.1.3 Additional Context

Throughout the assessment, Marines bivouacked at Range 107, sleeping in 2-man tents. Prior to the experiment, each machine gun was zeroed to maximize accuracy. During trial execution, Marines wore/carried prescribed loads for each task. Weighing packs each day prior to the 7-km forced march ensured consistency. After each trial day, the Marines operated under the guidance of their Company leadership, performing minimal physically demanding tasks. The Marines who were not part of an assessed squad conducted the same experimental subtasks after the assessed squads to ensure equity between individuals participating in a trial cycle and those not chosen for that particular cycle. These tasks will be discussed in detail in the loading section below.

Fatigue surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. It allows for outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) to be accounted for in the analysis of the performance data. Workload surveys collected the volunteers' perceived average and maximal level of exertion during the performance specified tasks. Cohesion surveys provided a method of collecting subjective data relating to each squad's ability to work as a team and their overall perspective on the cohesiveness of the squad.

C.1.4 Experimental Tasks

C.1.4.1 1-km Movement

Assaulting an enemy position never starts from a static position; first, a movement must be conducted to the assault position (AP). The distance from the line of departure to the AP is dependent upon myriad factors. Based on time and space constraints, this distance was set at just under 1 km for the experimental event. Each machinegun squad moved this distance as quickly as possible while carrying an assault load in addition to the M240B machine gun, tripod, spare barrel, and ammunition (600 rounds of 7.62 mm) divided among the members.

C.1.4.2 Occupy and Engagement from SBF Position

Prior to commencing an assault, it is common for the machinegun section to occupy a position of overwatch and provide SBF for the attacking unit. Movement to a SBF position varies in distance based on the terrain, which is often challenging. From the SBF position, the machinegun squad must rapidly get their gun into action, acquire targets, and accurately engage the enemy. During the assessment, each machinegun squad moved approximately 100 meters from an AP to the SBF position and emplaced the M240B. Once the squad leader confirmed "gun up," they engaged targets presented in a predetermined course of fire. Three targets and 400 rounds were allocated during this course of fire. This task determined how quickly the squad could move to a SBF position and get the gun into action, as well as determined the squad's accuracy while engaging targets.

C.1.4.3 Displace to the LOA

After providing initial suppressive fires from a SBF position to support an assault, machinegun squads generally move to another position of advantage. Given the command, they must break down their weapon system and displace to a follow-on firing position, quickly. During the assessment, each machinegun squad displaced from their initial SBF position approximately 300 meters to a LOA/secondary SBF position. From this new position, they prepared to repel the enemy counterattack. This task determined how much time it took for a machinegun squad to displace and prepare for an enemy counterattack.

C.1.4.4 Repel Counterattack

At the conclusion of every assault, the attacking unit must be prepared for the enemy to regroup and organize a counterattack. Upon consolidation at the LOA, the squad oriented its weapons in the direction of the enemy's retreat. Upon targets being presented, the counterattack commenced for 90 seconds. Two targets and 200 rounds were allocated during the machinegun counterattack course of fire.

C.1.4.5 CASEVAC

Casualties are an inevitable part of conducting combat operations. Units train to become proficient in triaging, handling, and transporting a casualty. When a casualty is sustained, it is essential to move with a sense of urgency to get the injured Marine to the appropriate level of care. At the conclusion of the live-fire and counterattack, each machinegun squad moved a 220-lb dummy 100 meters from a position of cover to a casualty collection point (CCP) while also transporting their assault packs and crew-served weapon system (M240B, tripod, and spare barrel). The machinegun squads could use a variety of techniques for transport but had to carry the dummy off the ground. This task determined the machinegun squad's proficiency in moving a simulated casualty to a CCP. After the CASEVAC, each Marine took a fatigue and workload survey to assess overall fatigue and workload of the entire offensive task (see GCEITF Experimental Assessment Plan [EAP], Annex D).

C.1.4.6 7-km Hike

Infantry units must move through all sorts of terrain on foot. Units train by conducting a forced march with an approach load at a sustained rate of march. For the assessment, each machinegun squad moved a distance of 7.2 km as quickly as possible while carrying an approach load, and the M240B machine gun, tripod, spare barrel, and ammunition spread-loaded across all three members. This task determined the squad's rate of movement over a 7.2-km route while carrying the approach load and their crew-served weapon. Each Marine took a fatigue and workload survey after completion of the 7.2-km hike.

C.1.4.7 M2 Emplacement, Engagement, and Displacement

Providing defensive fires with the M2 heavy machine gun entails moving the system to a position of advantage, engaging the enemy, and conducting a rapid displacement. Oftentimes, the M2 is employed from the tripod. During the assessment, each machinegun squad emplaced the M2 on a tripod at a specified firing location. The assessment began with targets exposed and engaged by the squad. Three targets and 400 rounds were allocated for this course of fire. Immediately upon going out of action, the squad displaced from the firing line to a designated location, moving a heavy load a short distance, manipulating a weapon while fatigued, and accurately engaging targets. This task determined accuracy and displacement times.

C.1.4.8 M2 Mount and Dismount

The M2 heavy machine gun is often employed from a vehicular platform, such as a HMMWV. To mount this system, the squad must lift all components from the ground to the turret and assemble the system. Similarly, to dismount the system, the squad must manually lower each component to the ground. During the assessment, each machinegun squad worked together to mount and dismount an M2 from a HMMWV. This task determined the time for a squad of three Marines to fully mount and dismount the M2 from a tactical vehicle and required the strength to lift, manipulate, and lower heavy components. At completion of the mount and dismount tasks, Marines took a fatigue and workload survey to assess their fatigue and workload during execution (see GCEITF EAP, Annex D).

At the completion of the 2-day cycle, Marines took a cohesion survey to record their cohesion during execution of the 2-day trial cycle.

C.1.5 Loading Plan

Due to the number of volunteers, several Marines were not part of an assessed squad each 2-day cycle. The loading plan ensured, to the greatest extent possible, equity of physical activity amongst all volunteers throughout the duration of the experimental assessment. Collaboration with the Company leadership determined the best method of loading non-assessed Marines was to have them perform the same tasks as an assessed squad to experience the same conditions and physical strain. Minor modifications were permitted due to the reduced size of the squad, such as conducting a trial as the fourth Marine in a 3-Marine element when not enough individuals were available to form another squad. Every trial and task was conducted in the same manner and sequence to ensure consistency.

C.1.6 Scheme of Maneuver Summary

The 0331 experiment consisted of a 2-day trial cycle comprising an offensive and defensive day. The offensive day involved five subtasks based around supporting a squad attack: 1-km movement, movement to and occupy a SBF position, displacement, repel a counterattack, and CASEVAC. The defensive day involved four subtasks: a 7-km forced march, M2 engagement, displacement, and mount/dismount drill. During the course of the experiment, the 0331 machinegun squad executed 2 pilot trial cycles and 21 record trial cycles. During trial execution, Marines rotated through every billet within the machinegun squad, carrying components of the crew-served load.

C.2 Limitations

C.2.1 0331 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a similar manner to those in an operational environment. However, under certain situations, artificial limitations or interruptions were introduced that changed or altered the way a task would normally be performed. While these limitations represent a degree of artificiality, they do not detract significantly from our abilities to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for 0331 assessment.

C.2.2 Relative Difficulty of Record Test

The 0331 GCEITF assessment was designed to gather data associated with some of the most physically demanding tasks within the 0331 MOS. These tasks in isolation did not fully replicate life experienced by a Marine during a typical field exercise (FEX) or a combat environment and did not obtain the maximum cumulative load that could be placed on an 0331 Marine. With limited time available, only selected 0331 tasks were assessed. Other tasks/duties outside of the assessment were minimized due to specific experimental constraints and human factors. During a typical FEX, it is common for Marines to conduct 24-hour operations that include day and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. The offensive day SOM took squads approximately 1 hour to complete, and the defensive day SOM took approximately 3 hours to complete. Outside the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

Another concern in designing the 0331 assessment was making it was achievable and sustainable for a 60-day period. The 7-km forced march distance was selected based on the training time available prior to the assessment. However, many of the loads carried were decreased; the crew-served load was altered from the M2 heavy machine gun to the M240B medium machinegun. Marines were authorized 1 day off after every 4 days of training; this artificial recovery period is not achievable when conducting training or combat operations.

A final factor affecting the relative difficulty of the record test pertains to the intangible physiological impact of the volunteers' ability to DOR at any point during a trial. Any time a volunteer dropped during a trial cycle, that squad performed the following subtasks with fewer personnel. This factor could have affected the cohesion of each squad and influenced its performance.

C.2.3 Geometries of Fire and Conditions Set

Several artificialities were present during the M240B live-fire portion of the assessment. Although the SBF position was realistic, the geometries of fire for the initial and secondary position were offset to prevent fires from interfering with the rifle-squad assessment. The machinegun course of fire began when they reached the SBF position, rather than waiting for the tactical conditions to be established. The loss of tactical realism in basing fires off a maneuver element resulted in this task being less challenging than in training or combat, in which the squad leader and gunner must make intuitive decisions regarding rates of fire, target precedence, and shifting and ceasing criteria.

C.2.4 Number of Volunteer Participants

For the 0331 experiment, six male and six female volunteers began the experiment, but by the end five males and six females completed the assessment. The results presented in this annex are based on the performance of 11 to 12 Marines.

C.2.5 Limitations Summary

The 0331 assessment was designed to replicate realistic training in a field environment. The end-state was to create an experiment in which the volunteers felt they were conducting realistic and operationally relevant tasks, but unavoidable limitations to the assessed tasks and non-assessed operating environment introduced a level of artificiality not normally present in a field training or combat environment.

C.3 Deviations

C.3.1 M2 Heavy Machine Gun Employment

The EAP stated that just prior to the M2 .50-cal engagement, the Marines would be assessed moving the M2 and ammunition approximately 100 meters to the firing line. Discussion with Company leadership about the relevancy of this subtask informed the decision not to make the emplacement a timed event, but to administratively move the M2 to the firing line. This subtask was not assessed as a standalone measurement.

C.4 Data Set Description

C.4.1 Data Set Overview

The 0331 portion of the experiment consisted of 2 pilot trial cycles and 21 record trial cycles. The pilot trial cycles were conducted from 2 March 2015 to 6 March 2015. Pilot trial data are not used in analysis due to variations in the conduct of the test. We based all analysis on the 21 record trial cycles executed from 7 March 2015 to 26 April 2015.

C.4.2 Record Test Volunteer Participants

At the beginning of the first record trial cycle, there were six male 0331 volunteers and six female volunteers. There was one male DOR.

C.4.3 Planned, Executed and Analyzed Trial Cycles.

Table C-1 displays the number of trial cycles planned, executed and analyzed by task. The planned number of trials for the 0331 MOS per Section 7.5.3 of the GCEITF EAP was 120 trial cycles or 40 per integration level (C, LD, and HD). The original plan called for 6 squads per day (2 per integration level) over the 20 trial cycles. However, due to the number of Marines who voluntarily withdrew or were involuntarily withdrawn from the experiment prior to the execution of the first record trial cycle, only one squad of the C and HD integration levels remained. The planned number of trial cycles in Table C-1 reflects 21 planned trial cycles for each integration level.

Of note, there are several occurrences of missing data for the 7-km hike by individual kilometer. The individual kilometer times were derived from GPS data. Early in the experiment, the Garmin GPS's were set to record a volunteer's position every second. Due to the storage space on the GPS and length of the trial, when volunteers executed the 7-km hike and then follow-on tasks, the GPS could not hold all of the data and overwrote the hike data. Once the problem was found, the GPS's were corrected to record location every 2 seconds.

Table C-1. 0331 Planned, Executed, and Analyzed Trial Cycles[1]

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
7-km Hike	C	21	21	21	
	LD	21	18	18	
	HD	21	21	21	
M240B Movement & Emplacement	C	21	20	20	Mar 19 did not execute due to range fire
	LD	21	18	18	
	HD	21	20	20	Mar 19 did not execute due to range fire
M240B Engagement	C	21	20	18	Missing Mar 27 (no TRACR); TIR Mar 12; Did not execute Mar 19
	LD	21	18	16	Missing Mar 27 (no TRACR); TIR Apr 11
	HD	21	20	19	Missing Apr 18
CASEVAC to CPP	C	21	20	20	Mar 19 did not execute due to range fire
	LD	21	18	18	
	HD	21	20	19	TIR Mar 29; Mar 19 did not execute due to range fire
Mount/Dismount Total Elapsed Time	C	21	21	21	
	LD	21	18	18	
	HD	21	21	21	
Displace to LOA; Elapsed Time	C	21	20	20	Mar 19 did not execute due to range fire
	LD	21	18	18	
	HD	21	20	20	Mar 19 did not execute due to range fire
M2 Displacement	C	21	21	21	
	LD	21	18	18	
	HD	21	21	21	
1-km Hike; Elapsed Time	C	21	20	20	
	LD	21	18	18	
	HD	21	20	20	
7-km Hike; 1km Time	C	21	20	18	Missing Mar 13 and Mar 15
	LD	21	18	15	Missing Mar 10, Mar 15, Mar 18
	HD	21	21	16	Missing Mar 8, Mar 10, Mar 13, Mar 15, Mar 18 data
7-km Hike; 2km Time	C	21	20	18	Missing Mar 13 and Mar 15
	LD	21	18	15	Missing Mar 10, Mar 15, Mar 18
	HD	21	21	16	Missing Mar 8, Mar 10, Mar 13, Mar 15, Mar 18 data
7-km Hike; 3km Time	C	21	20	18	Missing Mar 13 and Mar 15
	LD	21	18	16	Missing Mar 15 and Mar 18
	HD	21	21	17	Missing Mar 8, Mar 10, Mar 15, Mar 18 data
7-km Hike; 4km Time	C	21	20	20	Missing Mar 18 data
	LD	21	18	17	Missing Mar 18 data
	HD	21	21	18	Missing Mar 8, Mar 15 and Mar 18 data
7-km Hike; 5km Time	C	21	20	20	Missing Mar 18
	LD	21	18	17	Missing Mar 18
	HD	21	21	19	Missing Mar 8 and Mar 18 data
7-km Hike; 6km Time	C	21	20	20	Missing Mar 18
	LD	21	18	17	Missing Mar 18 data
	HD	21	21	20	Missing Mar 18 data
7-km Hike;	C	21	20	19	Missing Mar 18 data; Mar 8 likely GPS data error

[1] A TIR in this table refers to a Test Incident Report, which is a report the test team or direct assignment leaders completed when an incident occurred that affected the natural execution of a trial. If a data point is removed due to a TIR, it is because the TIR affected the data in such a way that it is not comparable to the rest of the data set.

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
7km Time	LD	21	18	17	Missing Mar 18
	HD	21	20	20	Missing Mar 18 data
M240B Movement	C	21	20	20	Mar 19 did not execute due to range fire
	LD	21	18	18	
	HD	21	20	20	Mar 19 did not execute due to range fire
M240B Emplacement	C	21	20	20	Mar 19 did not execute due to range fire
	LD	21	18	18	
	HD	21	20	20	Mar 19 did not execute due to range fire
M240B Engagement; Attack Hits on Target	C	21	20	20	Did not execute Mar 19
	LD	21	18	17	Missing Apr 11
	HD	21	20	20	Did not execute Mar 19
M240B Engagement; Counterattack hits on Target	C	21	20	18	Missing Mar 12 (TIR) and Mar 27 (no TRACR); Did not execute Mar 19
	LD	21	18	16	Missing Mar 27 (no TRACR) and Apr 11
	HD	21	20	18	Missing Apr 11 and Apr 18 run 3; Did not execute Mar 19
Mount M2; Elapsed Time	C	21	21	21	
	LD	21	18	18	
	HD	21	21	21	
Dismount M2; Elapsed Time	C	21	21	21	
	LD	21	18	18	
	HD	21	21	21	
M2 Percent Hits	C	21	21	21	
	LD	21	18	17	Mar 13 TRACR recorded greater than 400 hits
	HD	21	21	19	Apr 2 and Apr 9 TRACR recorded greater than 400 hits

C.5 Descriptive and Basic Inferential Statistics

C.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common infantry tasks and are indicative of unit-level proficiency during either field exercises or combat operations. This section presents the descriptive statistics results for 7 tasks out of 21 tasks. The Appendix to this Annex contains the descriptive statistics for the remainder of the 0331 tasks. The words “metric” and “task” are used interchangeably throughout this Annex; both refer to the experimental task.

Each machinegun squad consisted of three volunteer Marines: the gunner, assist-gunner, and ammo man. A direct assignment (non-volunteer) squad leader led each squad. There were three integration levels for all tasks. A C group was all male, a LD group had one female, and a HD group had two females.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA), Tukey Tests (or non-parametric tests as necessary), and scatter plots. The subsequent sections will cover each task in detail. Lastly, contextual comments, additional insights, and subjective comments (as applicable) tying back to each experimental task are incorporated.

Special caution should be taken when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks, such as distances, techniques, leadership, load carried, group size, and group composition.

C.5.2 0331 Selected Tasks Descriptive Statistics Results

The two tables below display the results for the seven selected 0331 metrics. Table C-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations. Table C-3 displays ANOVA and Tukey Test results, including metrics and integration levels, p-values suggesting statistical significance, integration level elapsed time differences, and percentage differences between integration levels. For each task, an ANOVA was conducted to compare the three groups and Tukey Tests were conducted to compare each pair of two groups. If non-parametric tests were needed, Table C-3 displays these results instead of ANOVA and Tukey Test results. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response was not found to be the same across all three groups.

Table C-2. 0331 Selected Task Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD
7km Hike (minutes)*	C	21	86.44	6.76
	LD	18	113.29	14.75
	HD	21	117.30	14.65
M240B Movement & Emplacement (minutes)*	C	20	2.55	0.31
	LD	18	2.90	0.37
	HD	20	3.32	0.46
M240B Engagement (% hits)†	C	18	33.08%	17.23%
	LD	16	38.07%	15.07%
	HD	19	27.34%	12.75%
CASEVAC (minutes)†	C	20	2.23	0.51
	LD	18	3.11	1.72
	HD	19	3.95	1.88
CASEVAC **[excluding potential influential points]** (minutes)*	C	20	2.23	0.51
	LD	15	2.41	0.65
	HD	17	3.38	0.86
M2 Mount/Dismount (minutes)	C	21	3.95	0.81
	LD	18	3.94	0.64
	HD	21	4.35	1.03
M240B Displace to LOA (minutes)†	C	20	3.75	0.36
	LD	18	4.58	0.68
	HD	20	4.74	0.69

Metric	Integration Level	Sample Size	Mean	SD
M2 Displacement (minutes)*	C	21	1.11	0.24
	LD	18	1.18	0.26
	HD	21	1.42	0.36
M2 Displacement **[excluding potential influential points]** (minutes)*	C	21	1.11	0.24
	LD	18	1.18	0.26
	HD	20	1.34	0.16

*Indicates there is a statistically significant difference a two-sided hypothesis test between integration levels according to ANOVA or a non-parametric equivalent test.
†Indicates contradicting statistical significance results between ANOVA and a non-parametric equivalent test.

Table C-3. 0331 Selected Task ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80% LCB	80% UCB	90% LCB	90% UCB
7km Hike (minutes)††	113.77	< 0.01*	LD-C	26.85	31.06%	< 0.01*	21.86	31.83	20.37	33.32
			HD-C	30.86	35.70%	< 0.01*	26.24	35.48	24.87	36.85
			HD-LD	4.01	3.54%	0.40	-2.15	10.18	-3.96	11.97
M240B Movement & Emplacement (minutes)*	20.07 (2, 55)	< 0.01*	LD-C	0.35	13.68%	0.02*	0.13	0.57	0.09	0.61
			HD-C	0.76	29.92%	< 0.01*	0.55	0.97	0.51	1.02
			HD-LD	0.41	14.28%	< 0.01*	0.20	0.63	0.15	0.67
M240B Engagement (% hits)*	2.22 (2, 50)	0.12	LD-C	0.05	15.08%	0.45†	-0.02†	0.14†	-0.03†	0.15†
			HD-C	-0.06	-17.35%	0.25†	-0.10†	0.01†	-0.13†	0.03†
			HD-LD	-0.11	-28.18%	0.02†	-0.16†	-0.04†	-0.19†	-0.03†
CASEVAC (minutes)*	20.56‡	< 0.01‡	LD-C	0.87	39.10%	0.14†	0.03†	0.63†	-0.02†	0.80†
			HD-C	1.71	76.67%	< 0.01†	0.95†	1.55†	0.82†	1.72†
			HD-LD	0.84	27.01%	0.02†	0.48†	1.25†	0.27†	1.37†
CASEVAC **[excluding potential influential points]** (minutes)*	14.65 (2, 49)	< 0.01*	LD-C	0.17	7.76%	0.74	-0.23	0.58	-0.31	0.66
			HD-C	1.15	51.54%	< 0.01*	0.76	1.54	0.68	1.62
			HD-LD	0.98	40.62%	< 0.01*	0.56	1.40	0.47	1.48
M2 Mount/Dismount (minutes)	1.54 (2, 57)	0.22	LD-C	-0.02	-0.48%	1.00	-0.49	0.45	-0.59	0.55
			HD-C	0.39	9.98%	0.30	-0.06	0.85	-0.15	0.94
			HD-LD	0.41	10.50%	0.29	-0.06	0.89	-0.16	0.98
M240B Displace to LOA (minutes)*	16.01 (2, 55)	< 0.01*	LD-C	0.83	22.23%	< 0.01†	0.52†	0.87†	0.48†	0.92†
			HD-C	1.00	26.58%	< 0.01†	0.88†	1.22†	0.82†	1.27†
			HD-LD	0.16	3.56%	0.10†	0.07†	0.55†	0.00†	0.58†
M2 Displacement (minutes)*	6.28 (2, 57)	< 0.01*	LD-C	0.07	5.98%	0.40†	-0.03†	0.15†	-0.05†	0.17†
			HD-C	0.31	27.61%	< 0.01†	0.22†	0.35†	0.18†	0.38†
			HD-LD	0.24	20.40%	0.01†	0.13†	0.30†	0.10†	0.33†
M2 Displacement **[excluding potential influential points]** (minutes)*	6.06 (2, 56)	< 0.01*	LD-C	0.07	5.98%	0.62	-0.06	0.19	-0.08	0.22
			HD-C	0.23	21.15%	< 0.01*	0.11	0.35	0.09	0.38
			HD-LD	0.17	14.31%	0.06*	0.04	0.29	0.02	0.32

*Indicates there is a statistically significant difference in a two-sided hypothesis test between integration levels according to ANOVA or a non-parametric equivalent test.
**Tukey intervals have familywise confidence of the indicated percentage, each interval is not of the given confidence level on its own.
†Results presented are from Kruskal-Wallis and Mann-Whitney non-parametric tests due to non-normality.
††Indicates results presented are from Robust ANOVA and Welch's t-tests with p-values compared to 0.033 for Bonferroni adjustment due to unequal variances. The reported F-statistic is a Chi-square statistic from Robust ANOVA, and the F-test p-value is the Robust ANOVA p-value. The p-values in columns labeled "2-sided P-value" and "1-sided P-value" are p-values from Welch's t-tests, and the confidence intervals are from Welch's t-tests.
‡Indicates results presented are from Robust ANOVA due to unequal variances. The reported F-statistic is a Chi-square statistic from Robust ANOVA, and the F-test p-value is the Robust ANOVA p-value.

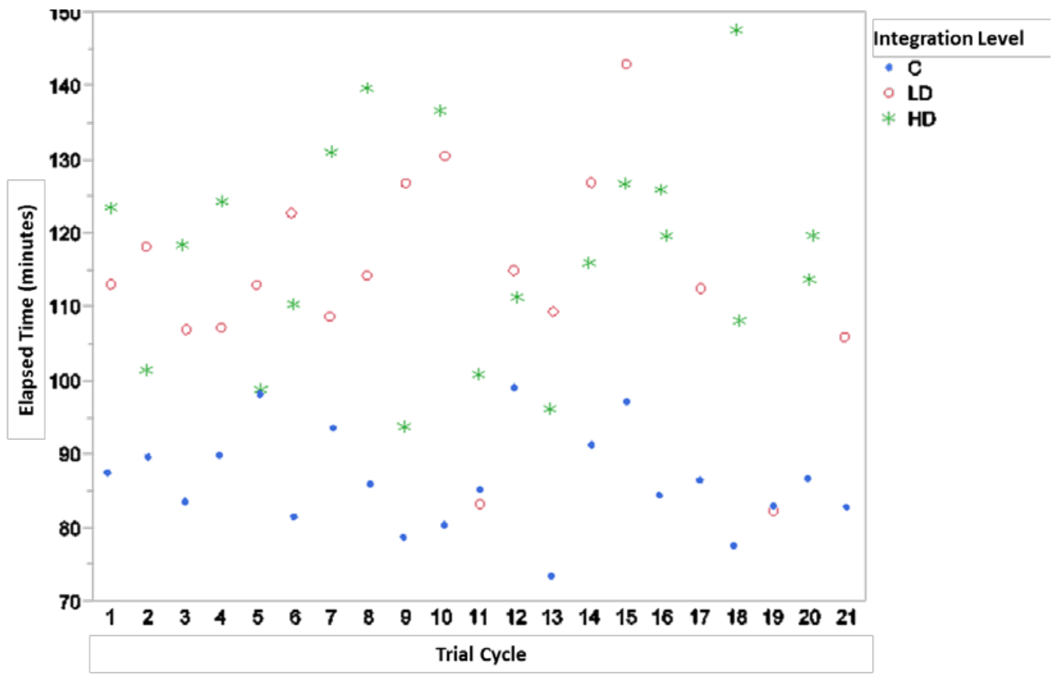
C.5.2.1 7-km Hike Results

C.5.2.1.1 7-km Hike Overview

This experimental task assessed a squad of three Marines moving 7.2 km while each Marine carried an approach load, and individual weapon (M-4), and a portion of the crew-served weapon load. The crew-served load consisted of an M240B medium machinegun, tripod, A-bag with spare barrel, and four cans of ammo, resulting in a cumulative load of 118-130 lb per Marine. The recorded time for this task started when the squad departed the Range 107 start point and stopped when the squad arrived at the Range 110 stop point. The squads moved as fast as the slowest person and could take as many breaks as necessary.

Figure C-1 displays all 0331 7-km hike data. All data on the scatter plot are valid for analysis.

Figure C-1. 7-km Hike



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.82 for the C group, 0.18 for the LD group, and 0.89 for the HD group.

The C group had a mean time of 86.44 minutes. This time is statistically significantly faster than the LD mean time of 113.29 minutes and the HD mean time of 117.30 minutes. These differences result in 31.06% (26.85 minutes) and 35.70% (30.86 minutes) degradations in time for the LD and HD groups, respectively. Additionally, the LD and HD groups had greater variability as shown by the 7.99- and 7.89-minute, respective, increases in standard deviation (SD) (6.76 minutes for the C group, 14.75 minutes for the LD group, and 14.65 minutes for the HD group). The LD group was faster, on average, than the HD group. There was a 3.54% (4.01 minutes) degradation in hike time from the LD to HD group, but this difference is not statistically significant. See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.1.2 7-km Hike Contextual Comments

C.5.2.1.2.1 USMC Hike Standards

The 7-km hike is a task that the following MOSs completed during the experiment: 0311, 0331, 0341, 0351, and 0352, Provisional Infantry, Provisional Machine Gunners, and Combat Engineers. There are varying standards to which we can compare this result. The following sections define those standards as well as the one we choose as a comparison.

The Infantry T&R Manual (30 Aug 2013) identifies minimum standards that Marines must be able to perform in combat, to include standards for tactical marches. In

Chapter 8 of the Infantry T&R Manual in the MOS 0300 Individual Events section, task "0300-COND-1001: March under an approach load" is applicable to all 03XX MOSs, ranks PVT – LtCol. The condition and standard established by this task is: "Given an assignment as a member of a squad, individual weapon, and an approach load, complete a 20 kilometer march in under 5 hours." The march pace required by this standard is 4 kilometers per hour (km/h). In Chapter 9 of the Infantry T&R Manual in the MOS 0302/0369 Individual Events section, tasks "0302-OPS-2001: Lead an approach march" and "0369-OPS-2501: Lead an approach march" are applicable to MOS 0302 and 0369, ranks SSgt – MGySgt and 2ndLt – LtCol. The condition and standard established by these tasks is: "Given a mission, time constraints, an approach march load, organic weapons, and a route, move 24.8 miles (40 km) in a time limit of 8 hours." The march pace required by this standard is 5 km/h. Appendix E of the Infantry T&R Manual (Load Terms and Definitions) states: "The approach march load will be such that the average infantry Marine will be able to conduct a 20-mile hike in 8 hours with the reasonable expectation of maintaining 90% combat effectiveness." The march pace required by this definition is 4.02 km/h.

Chapter 3 of Fleet Marine Force Reference Publication (FMFRP) 3-02A, Marine Physical Readiness Training for Combat (16 Jun 2004) states: "The normal pace is 30 inches. A pace of 30 inches and a cadence of 106 steps per minute result in a speed of 4.8 km/h or 3 mi/h and a rate of 4 km/h or 2.5 mi/h if a 10-minute rest halt per hour is taken." Common Infantry practice today is to hike for 50 minutes and take a 10-minute break, while maintaining an overall pace of 4 km/h (resulting in a hiking pace of 4.8 km/h).

Driven by the need to pick an evaluative reference standard, this report follows the T&R Manual's intent to establish minimum standards and uses the 4 km/h march pace for a 20-km march established by task 0300-COND-1001 and supported by the definition of an approach load. Further, while an established reference standard is required to anchor observed performance to the T&R program, more important information is provided by performance differences observed between gender integrated and non-gender integrated units.

C.5.2.1.2.2 0331 7-km Hike Pace

This result is relevant to both the training and combat environment as it will take integrated squads more time to conduct foot movements. Per the tactical march standards noted above, the Marine Corps standard of hiking is 4.0 km/h. The LD and HD groups failed to meet this standard. The C group average pace was 5.00 km/h; the LD group average pace was 3.81 km/h; and the HD group average pace was 3.68 km/h, finishing 26.85 and 30.86 minutes behind the C group, respectively. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), it would take the C group 4.0 hr, the LD group

5.25 hr, and the HD group 5.43 hr to complete the 20-km movement, meaning the LD and HD groups would finish 75 and 85.8 minutes, respectively, behind the C group.

Furthermore, on any given day (under the same environmental conditions), the C group was faster than the LD group 90.5% of the time (19 of 21 trial cycles) and faster than the HD group 100% of the time. With the exception of 5 out of 39 total integrated trials, the slowest C group was faster than the fastest HD group. Based on the standard deviations, the variation in performance of the LD and HD group is greater than twice as much as the variation in performance of the C group. This inconsistency in the performance of the integrated squads leads to greater uncertainty and less confidence in their future performance around their average time.

C.5.2.1.3 7-km Hike Additional Insights

The high degree of variability within the HD group can be explained by a variety of factors: the weather, Day 2 vs. Day 4 hike execution, or most notably the crew-served load that varied from 118-130 lb.

Based on the USMC standard of a 4 km/h pace over a 7.2-km route (which would result in a 108-minute hike completion time over the 7.2 km), the LD and HD groups were 5.29 and 9.3 minutes slower, respectively, than that standard. In a battlefield situation, in which speed is essential, this delay is advantageous for the enemy. An enemy maneuvering at 4 km/h would have the time to move between 353-620 meters, shift indirect fires from preplanned targets, commit the employment of their least engaged unit, or conduct a spoiling attack.

Marine Corps Doctrinal Publications (MCDP) consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to “Being Faster” and states, “Physical speed, moving more miles per hour, is a powerful weapon in itself.” MCDP-6 Command and Control also speaks to speed relative to the enemy and states, “The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results.” Further insights may be gleaned from the Appendix, which shows the difference in speed by kilometer. In general, the difference in performance increased as the movement got longer.

C.5.2.1.4 7-km Hike Subjective Comments

For subjective comments relating to this task, see the Appendix.

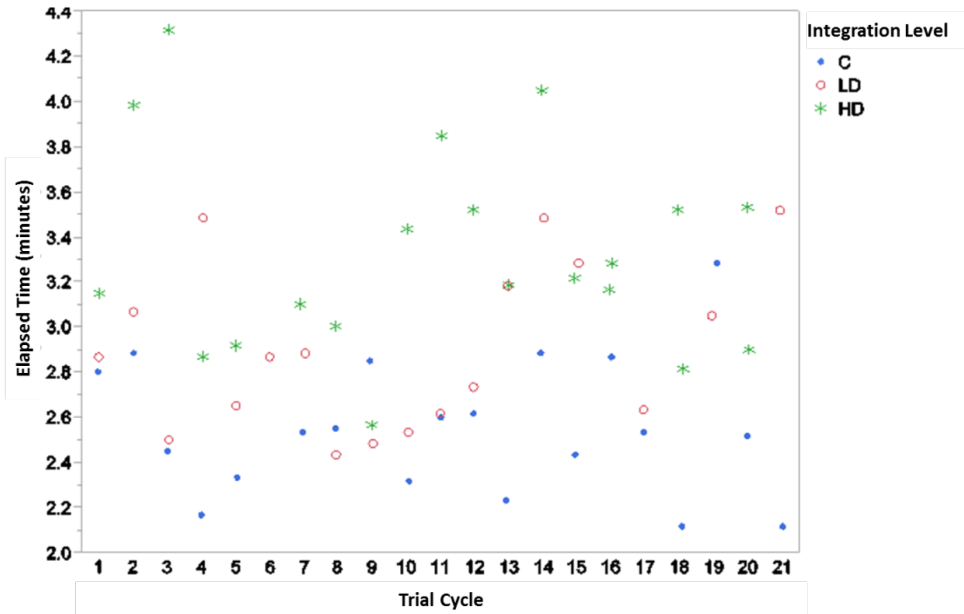
C.5.2.2 M240B Movement & Emplacement

C.5.2.2.1 M240B Movement & Emplacement Overview

This experimental task assessed the time for a three Marine squad to move approximately 100 meters to a SBF position and prepare the M240B medium machinegun for firing. This task was conducted immediately after completing a 1-km movement. The recorded time started immediately upon completing the 1-km movement, and stopped when the gunner yelled “Gun up,” indicting that it was ready to fire. During the emplacement process, the squad assembled the M240B, checked headspace and timing, and loaded the source of ammunition. A direct-assignment squad leader verified all procedures.

Figure C-2 displays all 0331 M240B Movement & Emplacement data. All data on the scatter plot are valid for analysis.

Figure C-2. M240B Movement & Emplacement



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.38 for the C group, 0.10 for the LD group, and 0.44 for the HD group.

The C group had a mean time of 2.55 minutes. This time is statistically significantly faster than the LD mean time of 2.90 minutes and the HD mean time of 3.32 minutes. These differences result in 3.68% (0.35 minute) and 29.92% (0.76 minute) degradations in time for the LD and HD groups, respectively. Additionally, the LD and HD groups had greater variability as shown by the 0.06- and the 0.15-minute, respective, increases in SD (0.31 minutes for the C group, 0.37 minutes for the LD group, and 0.46 minutes for the HD group). There was a 14.28% (0.41 minute) statistically significant degradation from the LD to the HD group. See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.2.2 M240B Movement and Emplacement Contextual Comments

This result is relevant to the combat environment as it will take integrated squads more time to move and initiate supporting fires. On average, it took the LD group 27 seconds longer and the HD group 46 seconds longer to conduct this movement and emplacement. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the LD group 82.4% of the time (14 of 17 trial cycles) and faster than the HD group 95% of the time (19 of 20 trial cycles).

C.5.2.2.3 M240B Movement and Emplacement Additional Insights

A purely objective evaluation of 27-46 seconds is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated squad. Considering the 650 rounds-per-minute sustained rate of fire for the Russian RPK-47 machinegun, a single enemy machinegun squad would have the opportunity fire 292-498 rounds against the pinned down squad prior to the initiation of friendly suppression by an integrated squad. The resultant trade in casualty exchange could be significant.

C.5.2.2.4 M240B Movement and Emplacement Subjective Comments

For subjective comments relating to this task, see the Appendix.

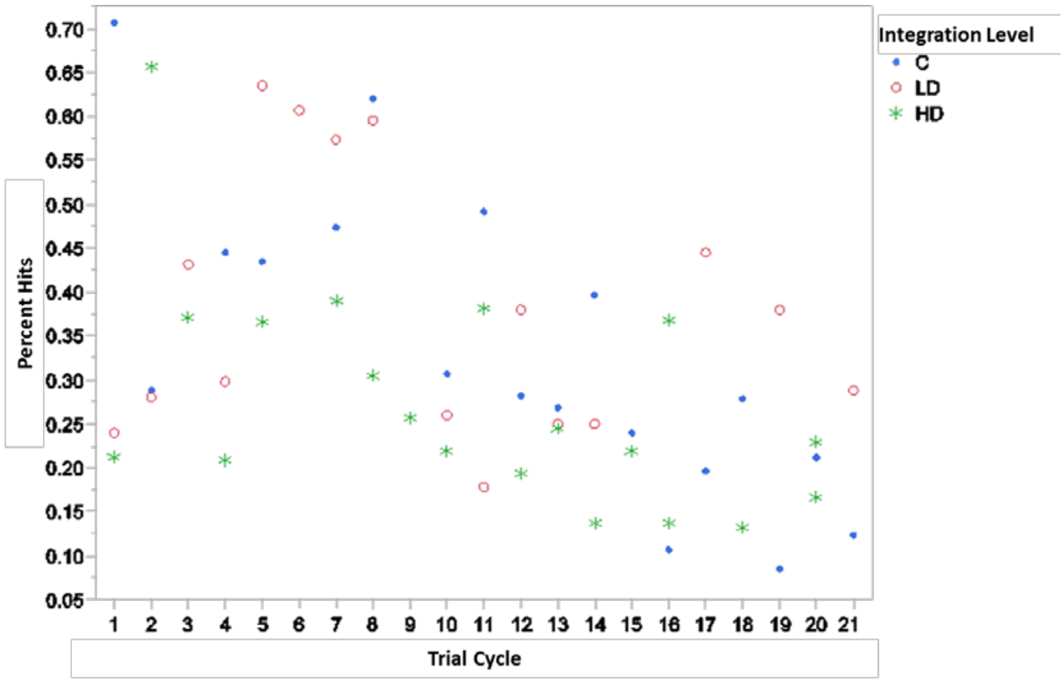
C.5.2.3 M240B Engagement

C.5.2.3.1 M240B Engagement Overview

This experimental task assessed the accuracy of a medium machinegun squad engaging GTSs with 400 rounds during an attack and 200 rounds during a counterattack. Machine gun squads engaged three GTSs during the attack and two during the counterattack. Each GTS captured the precise location of a round that passes within 3 meters of the location of hit and miss (LOMAH) sensor. A significant limitation of the GTS was that it did not capture any data that impacted the low (the berm of the target). The accuracy was determined by dividing the number of rounds detected on each target by the total amount of ammunition expended by each squad.

Figure C-3 displays all 0331 M240B Engagement data. All data on the scatter plot are valid for analysis.

Figure C-3. M240B Engagement



The data for the C and LD groups are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in p-values of 0.46 and 0.05, respectively, but not normally distributed for the HD group, which had a p-value of 0.01. The Kruskal-Wallis Test does not run due to unequal samples between integration levels.

The C group had an average percent hit of 33.08%; the LD, 38.07%; and the HD, 27.34%. These differences from the C group result in a 15.08% (4.99 percentage points) degradation from the LD group and a 17.35% (5.74 percentage points) improvement from the HD group. Additionally, the LD group had less variability as shown by the 2.16 percentage point increase in SD from the C group, and the HD had less variability as shown by the 4.48 percentage point decrease in SD from the C group (17.23% for the C group, 15.07% for the LD group, and 12.75% for the HD group). There was a 28.18% (10.73 percentage point) degradation from the LD to the HD group.

The difference in percent hits was not statistically significant between the LD and C groups and the HD and C groups in a two-sided Mann-Whitney Test. The difference in percentage hits was statistically significant between the HD and LD groups in a Mann-Whitney Test. This difference was not statistically significant between the HD and LD groups in a Tukey Test. Because of a lack of normality, we recommend using the Mann-Whitney test results (reported in Table C-3). See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.3.2 M240B Engagement Contextual Comments

In combat operations, accuracy is highly desirable in destroying or effectively suppressing an enemy position. The LD squads had the highest accuracy of 38.07% hits, which also relates to the best conservation of ammunition.

C.5.2.3.3 M240B Engagement Additional Insights

Although the assistance gunner has a role to play in adjusting fire, machinegun accuracy primarily rests on the skill of the gunner. It is important to note that during the course of the experiment, there were a total of 16 LD trials. Of these LD trials, a female served as the gunner on only 2 occasions (12.5%), both times during which the accuracy was below the LD mean. Conversely, of 19 total HD trials a female served as the gunner on 14 occasions (73.7%). It is also interesting to note that the overall accuracy of each group had a downward (negative) trend over the course of the experiment. One would have expected the accuracy to improve overtime. This result can primarily be attributed to a loss in interest due to the repetitive nature of the task and a lack of feedback to the volunteers.

C.5.2.3.4 M240B Engagement Subjective Comments

For subjective comments relating to this task, see the Appendix.

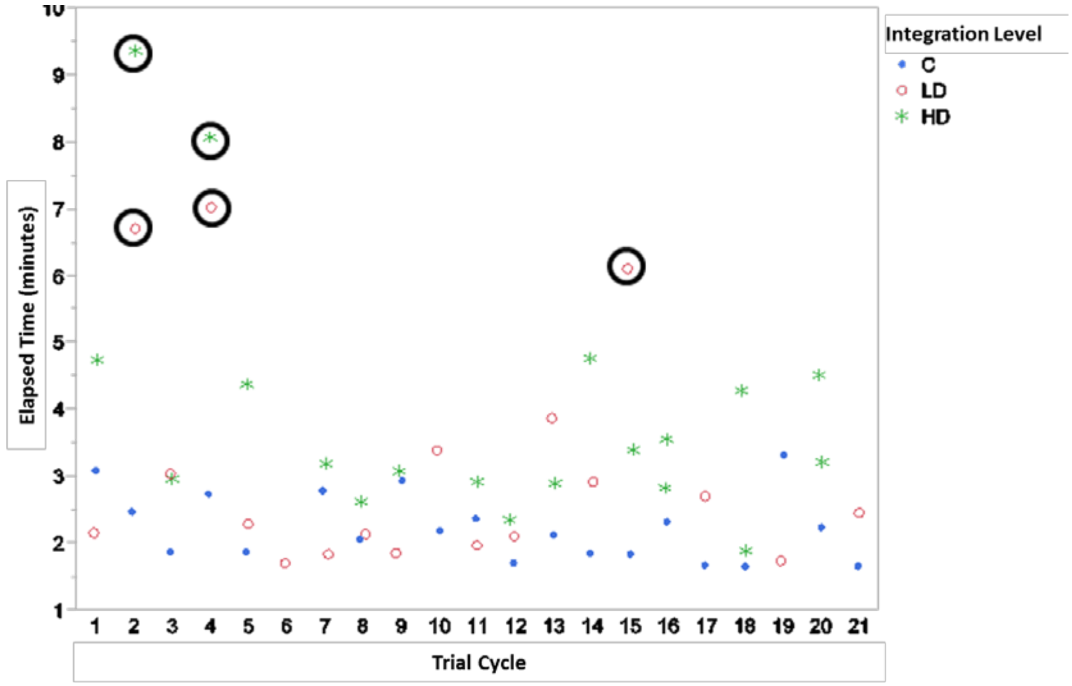
C.5.2.4 CASEVAC

C.5.2.4.1 CASEVAC Overview

This experimental task assessed the machinegun squad's ability to move a 210-lb dummy a distance of 100 meters to a casualty collection point (CCP) while wearing an assault load, individual weapon (M4), and the M240B crew-served load. Squads could use a variety of techniques, but they had to move all personnel and gear the entire distance, as well as carry the dummy off the ground. The machinegun squad conducted this task at the conclusion of the counterattack engagement. Techniques varied from the 1-Marine (fireman) to 2-Marine carry. The recorded time started when Marines touched the dummy and it stopped when the dummy and all members of the squad arrived at the CCP.

Figure C-4 displays all 0331 M240B CASEVAC data. There are five potential influential points: the LD and HD on trial cycle 2, the LD and HD on trial cycle 4, and the LD on trial cycle 15. Because the impact of these points is unknown, we perform all analysis with and without these points. All data on the scatter plot are valid for analysis with potential influential points circled.

Figure C-4. CASEVAC with Potential Influential Points Circled



The inclusion of the potential influential points does not change the statistical significance between groups. It does, however, change the SD and percent differences between the integration levels. Once we remove the potential influential points, the percent differences between the C group and LD group, as well as the C group and HD group, decrease. The percent difference between the LD and HD group increases. The SD for both the LD and HD groups decrease without the potential influential points. The following sections discuss results with and with the potential influential points. The removal of the potential influential points results in the data for the LD and HD groups being normally distributed.

C.5.2.4.1.1 CASEVAC Descriptive Statistics with Potential Influential Points

The data for the C group is normally distributed as evidenced by the Shapiro-Wilk Test that resulted in p-value of 0.08 but not normally distributed for the LD and HD groups, which both had a p-values of <0.01. The Kruskal-Wallis Test does not run due to unequal samples between integration levels.

The C group had a mean time of 2.23 minutes; the LD, 3.11 minutes; and the HD, 3.95 minutes. The difference from the C group results in 39.10% (0.87 minutes) and 76.67% (1.71 minutes) degradations in time for the LD and HD groups, respectively. Additionally, the LD and HD groups had greater variability, as shown by the 1.21-minute and 1.37-minute, respective, increases in SD (0.51 minutes for the C group, 1.72 minutes for the LD group, and 1.88 minutes for the HD group). There was a 27.01% (0.84 minute) degradation in time from the LD to the HD group.

The difference in CASEVAC time was not statistically significant between the LD and C groups in a two-sided Mann-Whitney Test but is statistically significant in a one-sided Mann-Whitney Test. The difference in CASEVAC time was statistically significant between the HD and C group as well as the HD and LD group in a Mann-Whitney Test. The difference in CASEVAC time was not statistically significant in for the LD and C group in a Tukey Test. Because of a lack of normality, we recommend using the Mann-Whitney test results (reported in Table C-3). However, the one-sided significance (between the LD and C groups) of the non-parametric results suggest that further study of this task is warranted. See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.4.1.2 CASEVAC Descriptive Statistics without Potential Influential Points

The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in p-values of 0.08 for the C group, 0.10 for the LD group, and 0.22 for the HD group.

The C group had a mean time of 2.23 minutes. This time is not statistically significant but faster than the LD mean time of 2.41 minutes, and it is statistically significantly faster the HD mean time of 3.38 minutes. These differences result in a 7.76% (0.17 minutes) and 51.54% (1.15 minutes) degradation in times for the LD and HD groups, respectively. The LD and HD groups had greater variability, as shown by the 0.14- and 0.35-minute, respective, increases in SD (0.51 minutes for the C group, 0.65 minutes for the LD group, and 0.86 minutes for the HD group). There was a 40.62% (0.98 minutes) statistically significant degradation from the LD to the HD group. See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.4.2 CASEVAC Contextual Comments

The implications of this task contain relevance to both the training and combat environment as a casualty must be moved expediently to a higher echelon of medical care. The data demonstrate that the LD group took 0.88 minutes longer than the C group, and the HD group took 1.72 minutes longer than the C group. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the LD group 75% of the time (15 of 20 trial cycles), and faster than the HD group 100% of the time (19 of 19 trial cycles). Based on the standard deviations, the variation in performance of the LD and HD group is greater than twice as much as the variation in performance of the C group. This inconsistency in the performance of the integrated squads leads to greater uncertainty and less confidence in their future performance from the mean.

C.5.2.4.3 CASEVAC Additional Insights

While the “Golden Hour” is a common medical planning construct for C2 and logistical support, medical literature supports a “Platinum Ten” philosophy of first response. The U.S. National Library of Medicine references a French article that espouses, “on the battlefield, the majority of casualties die within ten minutes of the trauma*.” (Wounded in*

Action: The Platinum Ten Minutes and the Golden Hour, Daban) The fundamental principle is that a patient needs to be correctly triaged and moved to medical care as fast as possible. Any time degradation will reduce the probability of survival.

C.5.2.4.4 CASEVAC Subjective Comments

There are seven instances in the leadership subjective logs that indicate integrated squads required extra breaks, compared to zero comments for the C group squads. The implication is that a C group could have maintained their CASEVAC pace for a longer distance, while the integrated groups could not have maintained their CASEVAC pace over a longer distance. Of the 21 instances stating that a Marine fireman-carried the dummy during integrated trials, three instances (14.3%) applied to a female, and 17 instances (85.7%) applied to male, implying that the males were doing more of the workload and contributing to better performance. This is confirmed by the fact that six of the seven fastest LD group times were instances of a male carrying the casualty the entire distance.

For more subjective comments relating to this task, see the Appendix.

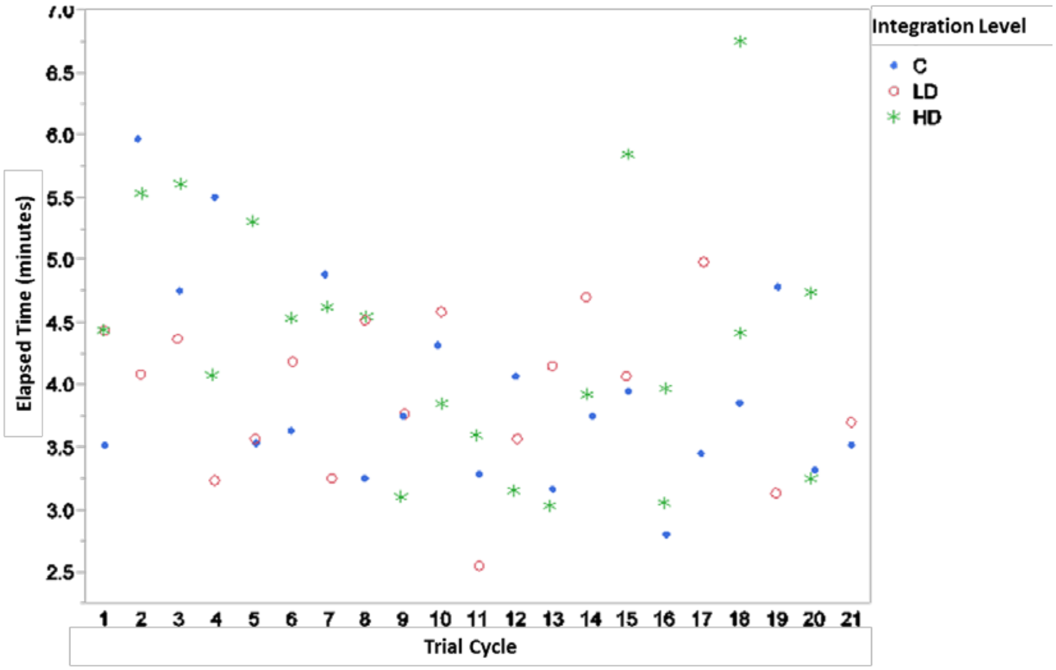
C.5.2.5 M2 Mount/Dismount

C.5.2.5.1 M2 Mount/Dismount Overview

This experimental task assessed the machinegun squad's ability to mount an M2 heavy machinegun onto a tactical vehicle (HMMWV) and then dismount it. A brief, non-assessed administrative pause was conducted between the mounting and dismounting sub-tasks. This task was conducted immediately following the M2 live-fire portion of the experiment. The recorded time started when Marines touched a component of the M2 and it stopped when the last component of the M2 was back on the deck.

Figure C-5 displays all 0331 M2 Mount/Dismount data. All data on the scatter plot are valid for analysis.

Figure C-5. M2 Mount/Dismount



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in p-values of 0.04 for the C group, 0.87 for the LD group, and 0.25 for the HD group.

The C group had mean time of 3.95 minutes. This time is slower (but not statistically significantly) than the LD mean time of 3.94 minutes but faster (but not statistically significantly) than the HD mean time of 4.35 minutes. These differences result in a 0.48% (0.02 minutes) degradation from the LD group and a 9.98% (0.39 minutes) improvement from the HD group. Additionally, the LD group had less variability as shown by the 0.17-minute increase in SD from the C group, the HD had more variability as shown by the 0.22-minute increase in SD from the C group (0.81 minutes for the C group, 0.64 minutes for the LD group, and 1.03 minutes for the HD group). There was a 10.50% (0.41 minutes) degradation from the LD to the HD group that was not statistically significant. See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.5.2 M2 Mount/Dismount Contextual Comments

The challenging aspect of mounting and dismounting the M2 heavy machinegun is the weight of the components. On average, it took the LD group less than 1 second less, and the HD group 24 seconds more than the C group.

C.5.2.5.3 M2 Mount/Dismount Additional Insights

This task began with all the equipment prestaged on the deck within 2 meters of the HMMWV. The Marines were not required to move the components to/from a secured location (armory). For an indication of the performance when moving heavy objects, see the TOW Engagement data, which involved moving a 50-lb missile a distance of 100 meters.

C.5.2.5.4 M2 Mount/Dismount Subjective Comments

For subjective comments relating to this task, see the Appendix.

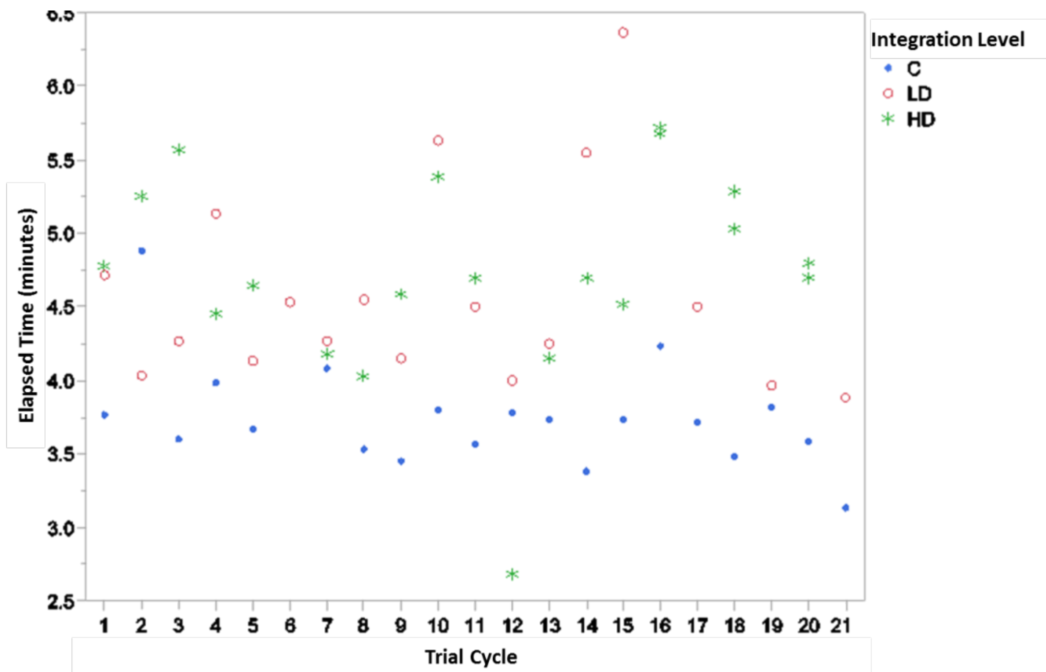
C.5.2.6 M240B Displace to LOA

C.5.2.6.1 M240B Displace to LOA Overview

This experimental task assessed a medium machinegun squad moving approximately 300 meters to a LOA immediately after engaging targets from a SBF position. The recorded time started when the squad was told to displace and stopped when the tripod was down at the LOA.

Figure C-6 displays all 0331 M240B Displace to LOA data. All data on the scatter plot are valid for analysis.

Figure C-6. M240B Displace to LOA



The data for the C and HD are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.02 and 0.05, respectively, but not normally distributed for the LD group, which had a p-value of <0.01. The Kruskal-Wallis Test does not run due to unequal samples between integration levels.

The C group had a mean time of 3.75 minutes; the LD, 4.58 minutes; and the HD. 4.74 minutes. The difference from the C group result in a 22.23% (0.83 minutes) and 26.58% (1.00 minute) degradation in times for the LD and HD groups, respectively. Additionally, the LD and HD groups had greater variability as shown by the 0.32- and

0.33-minute, respective, increase in SD (0.36 minutes for the C group, 0.68 minutes for the LD group, and 0.69 minutes for the HD group). The LD group was faster than the HD group. There was a 3.56% (0.16-minute) degradation in hike time from the LD to HD group.

The difference in time was not statistically significant between the HD and LD groups in a two-sided Mann-Whitney Test but was statistically significant in a one-sided Mann-Whitney Test. This difference was not statistically significant in a Tukey Test. The difference in time was statistically significant between the LD and C groups and HD and C groups in a Mann-Whitney Test. Because of a lack of normality, we recommend using the Mann-Whitney test results (reported in Table C-3). See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.6.2 M240B Displace to LOA Contextual Comments

The ability to close with the objective after having conducted an attack is a crucial aspect to maintaining the momentum during offensive operations. On average, the LD group took 50 seconds longer and the HD group took 59 seconds longer, which results in that much less time to prepare for a counterattack. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the LD group 94.1% of the time (16 of 17 trials) and faster than the HD group 95% of the time (19 of 20 trials). Based on the standard deviations, the variation in performance of the LD and HD group is nearly than twice as much as the variation in performance of the C group. This inconsistency in the performance of the integrated squads leads to greater uncertainty and less confidence in their future performance from the mean.

C.5.2.6.3 M240B Displace to LOA Additional Insights

A purely objective evaluation of 50-59 seconds is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated squad. Considering the 650 rounds-per-minute sustained rate of fire for the Russian RPK-47 machinegun, a single enemy machinegun squad would have the opportunity to fire 541-639 rounds against an integrated squad while it moved to the LOA and was exposed to enemy fire. Similarly, the integrated squad would have that much less time to provide reinforcing fires against an enemy counterattack. The resultant trade in casualty exchange could be significant.

C.5.2.6.4 M240B Displace to LOA Subjective Comments

For subjective comments relating to this task, see the Appendix.

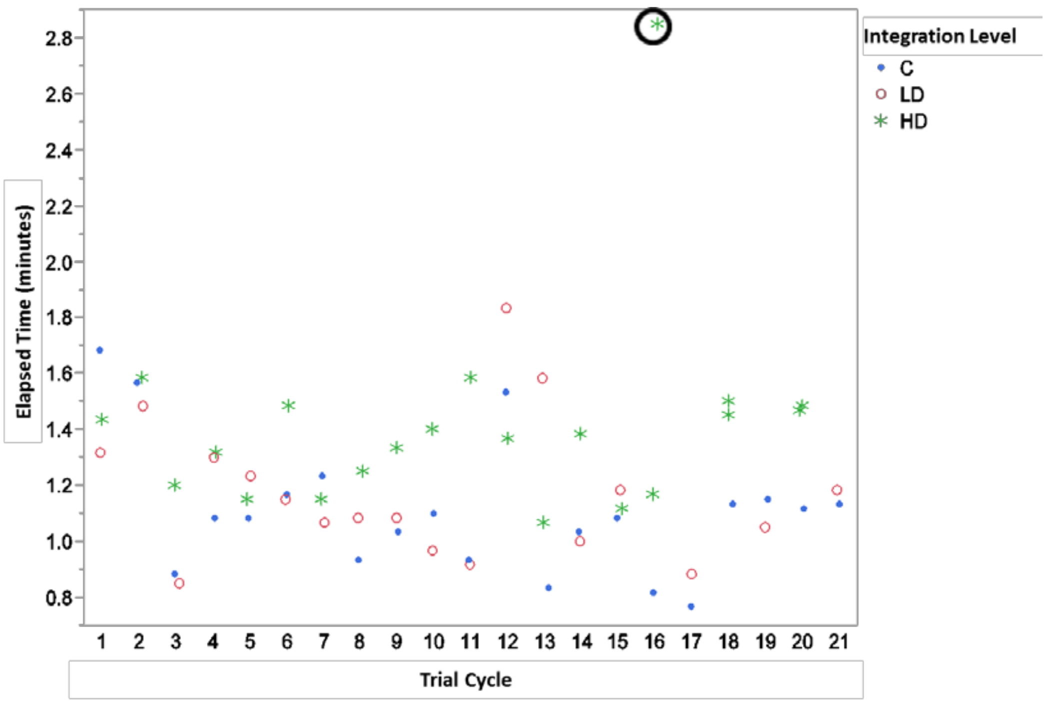
C.5.2.7 M2 Displacement

C.5.2.7.1 M2 Displacement Overview

This experimental task assessed the machinegun squad's ability to move an M2 heavy machinegun approximately 100 meters from a firing position to a rally point. The recorded time started when the machinegun squad yelled, "Out of action" and the time stopped when the entire squad arrived at the rally point.

Figure C-7 displays all 0331 M2 displacement data. There was one influential point: the HD on trial cycle 16. Because the impact of these points is unknown, we perform all analysis with and without this point. All data on the scatter plot are valid for analysis with the potential influential point circled.

Figure C-7. M2 Displacement with Potential Influential Point Circled



The inclusion of the potential influential points does not change the statistical significance between groups. It does change the SD and percent differences between the integration levels. Once we remove the potential influential points, the percent difference between both the C group and HD and the LD and HD group decreases. The SD for the HD groups decreases without the potential influential point. Additionally, the removal of the potential influential points results in the data for HD group being normally distributed. The following sections discuss results with and with the potential influential points.

C.5.2.7.1.1 M2 Displacement Descriptive Statistics with Potential Influential Points

The data for the C and LD groups are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in p-values of 0.03 and 0.11, respectively, but not normally distributed for the HD group, which had a p-value of <0.01. The Kruskal-Wallis Test does not run due to unequal samples between integration levels. Because of a lack of normality, we recommend using the Mann-Whitney test results (reported in Table C-3). We proceed with presenting ANOVA results because they were confirmed by Mann-Whitney Tests.

The C group had a mean time of 1.11 minutes. This time is faster (but not statistically significantly) than the LD mean time of 1.18 minutes, and it is statistically significantly faster the HD mean time of 1.42 minutes. These differences result in 5.98% (0.07 minutes) and 27.61% (0.31 minutes) degradations in time for the LD and HD groups, respectively. There was a 20.40% (0.04 minutes) statistically significant degradation from the LD to the HD group. See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.7.1.2 M2 Displacement Descriptive Statistics without Potential Influential Points

The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in p-values of 0.03 for the C group, 0.11 for the LD group, and 0.26 for the HD group.

The C group had a mean time of 1.11 minutes. There is no change between the C and LD groups. The C group is still statistically significantly faster than the HD group as the HD group mean is 1.34 minutes, which results in a 21.15% (0.23-minutes) degradation. The HD group is still less variable than the C group (0.16 minutes for the HD group). There was a 14.31% (0.17 minutes) statistically significant degradation from the LD to the HD group. See Table C-2 and Table C-3 for detailed analytical results.

C.5.2.7.2 M2 Displacement Contextual Comments

After an engagement, a machinegun squad must be able to breakdown their weapon system and move to a position of cover as rapidly as possible. On average, the LD group took 4 seconds longer and the HD group took 19 seconds longer, which translates to longer enemy exposure. On any given day (under the same environmental conditions), the C group was faster than the LD group 50% of the time (9 of 18 trials) and faster than the HD group 85% of the time (17 of 20 trials).

C.5.2.7.3 M2 Displacement Additional Comments

One interesting comparison can be made between this task and the M240B Displacement. The M2 Displacement involved a shorter movement of 100 meters with a heavier load versus the M240B Displacement that involved a longer movement of 300 meters with a lighter load. The data that were statistically significant (HD-C comparison) reveals roughly the exact same result: a 27.61% difference and 26.58% difference, respectively, between the two different tasks.

C.5.2.7.4 M2 Displacement Subjective Comments

For subjective comments relating to this task, see the Appendix.

C.6 Statistical Modeling Results

C.6.1 Statistical Modeling Overview

The previous section discussed results only as they pertain to differences due to integration level alone. The goal of statistical modeling as applied here is to estimate, simultaneously, the effect of gender integration levels and other relevant variables on machinegun squad performance. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For the same seven selected tasks described in the previous section, this section presents an overview of the analysis and results and then presents the modeling results for each of the tasks.

For each task, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates an increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time but not a desired outcome for the percent hits outcome. The results report where certain patch numbers are significant for a given variable. The experiment tracked Marines within the machinegun squad by a patch number that associated their random position within the squad to a specific billet. Table C-4 displays the patch numbers and associated billet titles for the machinegun squad.

Table C-4. Patch Numbers and Billet Titles for the Machinegun Squad

Patch Number	Billet Title
1	Gunner
2	Assistant Gunner
3	Ammo Man

C.6.2 0331 Selected Tasks Method of Analysis

Due to the small number of trials, a mixed effects model with all machinegun squad members and all types of personnel data does not work for the 0331 data set. Thus, we model each personnel variable with integration level separately with a random effect for who filled each position within the machinegun squad. For example, age for each member of the machinegun squad (three variables), a random effect for who filled each billet, and integration level are modeled with the result (response time or percentage hits) as the response variable. Where maximum likelihood estimation converged, AIC was used for variable selection. Otherwise, we comment on the significance of individual variables in the full model. Variables reported as significant are concluded to be significant based on at least a one-sided test.

C.6.3 0331 Selected Tasks Overall Modeling Results

There are no personnel data variables that are both statistically significant and have a practical impact to the model. Each time personnel data variables are statistically significant in a model, their effects are practically negligible, conflicting, and/or incomplete for the machinegun squad; i.e., there are no tasks for which a variable is significant for all members of the machinegun squad.

The HD and LD integration levels are significant for all models for the 7-km hike, M240B Movement & Emplacement, and M240B Displace to the LOA tasks. Only the HD integration level was significant for both the M2 Mount/Dismount task and the M2 Displacement tasks. For each task, modeling the random effects for individuals participating in the task resulted in changes from the initial results in the descriptive statistics. These changes are described in the respective task paragraphs.

The CASEVAC and M2 Displacement tasks had potential influential points, and we model these tasks with and without the influential points. When modeling the CASEVAC with influential points, the final model with integration level has HD and LD integration levels significant but only the LD integration level is significant without influential points. Additionally, not all of the same variables remain significant when modeled with and without influential points. Analysis with and without potential influential points for the M2 Displacement task does not change the overall result of the effect of the integration levels.

The M240B engagement task did not have a final model with any significant variables. Refer to the Section C.5.2.3 for this task for the ANOVA results.

C.6.3.1 7-km Hike

We model elapsed time for the 7-km hike as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position within the machinegun squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

Both the HD and the LD integration levels are significant and positively correlated with the response for the models that include the following personnel variables:

- All.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 7-km hike time:

- Height of patch 2,
- Rifle score of patches 2 and 3.

The following variables are significant in their respective models and are negatively correlated with the 7-km hike time:

- Squad leader,
- Weight of patches 1 and 3,
- AFQT score of patch 1,
- GT score of patch 3,
- CFT MTC of patch 2,
- CFT MANUF time of patch 2,
- PFT crunches of patch 2,
- PFT three-mile run time of patch 2.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes integration level only where HD has a difference of 31.48 minutes when compared to a C group. The comparison yields a statistically significant p-value of <0.01. This difference is an increase from the 30.86-minute difference identified in the descriptive statistics, which is a 2.01% change. The LD integration level has a statistically significant difference of 25.13 minutes when compared to a C group and a p-value of <0.01. This difference is a decrease from the 26.85-minute difference identified in the descriptive statistics, which is a 6.41% change.

C.6.3.2 M240B Movement & Emplacement

We model elapsed time for the M240B Movement & Emplacement as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position within the machinegun squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

Both the HD and the LD integration levels are significant and positively correlated with the response for the models that include the following personnel variables:

- All.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following variables are significant in their respective models and are positively correlated with the M240B movement and emplacement time:

- Age of patches 1 and 2,
- Height of patch 3,
- Weight of patch 3,
- CFT MANUF time of patches 1 and 2.

The following variables are significant in their respective models and are negatively correlated with the M240B movement and emplacement time:

- AFQT score of patch 2,
- GT score of patch 2,
- CFT MTC of patch 3.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes integration level only where HD has a difference of 0.80 minutes when compared to a C group and a statistically significant p-value of <0.01. This difference is an increase from the 0.76-minute difference identified in the descriptive statistics, which is a 5.26% change. The LD integration level has a difference of 0.38 minutes when compared to a C group with a statistically significant p-value of <0.01. This difference is an increase from the 0.35-minute difference identified in the descriptive statistics, which is an 8.57% change.

C.6.3.3 M240B Engagement

We model percent hits for the M240B engagement as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position within the machinegun squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the model that includes the following variables:

- None.

The LD integration level is significant and positively correlated with the response for the model that includes the following variables:

- AFQT score,

Both the HD and LD integration levels are significant and negatively correlated with the response for the model that includes the following variable:

- PFT crunches.

The HD integration level is significant and negatively correlated with the response for the model that includes the following variable:

- PFT 3-mile run.

The following variables are significant in their respective models and are positively correlated with the M240B engagement percent hits:

- AFQT score of patches 1 and 2,
- GT score of patch 1.

The following variable is significant in its respective models and is negatively correlated with the M240B engagement percent hits:

- Age of patch 1.

The models for the following variables have no significant variables in the model:

- Squad leader,
- Height,
- Weight,
- CFT MTC,
- CFT MANUF,
- Rifle score.

Because integration level is not significant in the final model and there are no variables that are significant for the whole squad, there is no final mixed effects model for this task. Refer to Section C.5.2.3 for this task to see the ANOVA results for differences between integration levels.

C.6.3.4 CASEVAC

We model elapsed time for the CASEVAC as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position within the machinegun squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

There were five potential influential points for this task. Therefore, we model this task with and without the influential points to determine if the influential points affect the outcome.

C.6.3.4.1 CASEVAC with potential influential points

The models for the following variables do not run due to missing values:

- None.

Both the HD and LD integration levels are significant and positively correlated with the response for the models that include the following variables:

- Squad leader,
- Age,
- Height,
- CFT MANUF,
- PFT crunches,
- Rifle score.

The HD integration level is significant and positively correlated with the response for the model that includes the following variable:

- GT score.

The LD integration level is significant and positively correlated with the response for the model that includes the following variable:

- PFT 3-mile run.

Both the HD and LD integration levels are significant and negatively correlated with the response for the models that include the following variables:

- None.

The following variables are significant in their respective models and are positively correlated with the CASEVAC time:

- Age of patches 1 and 3,
- Height of patch 3,
- CFT MTC of patches 1 and 2.

The following variables are significant in their respective models and are negatively correlated with the CASEVAC time:

- Height of patch 1,
- Weight of patch 1
- AFQT score of patch 1.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes integration level only where HD has a statistically significant difference of 1.43 minutes when compared to a C group and a p-value of 0.05. This difference is a decrease from the 1.71-minute difference identified in the descriptive statistics, which is a 16.37% change. The LD integration level has a difference of 0.79 minutes when compared to a C group and a one-sided statistically significant p-value of 0.05. This difference is a decrease from the 0.87-minute difference identified in the descriptive statistics, which is a 9.20% change.

C.6.3.4.2 CASEVAC without potential influential points

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following variables:

- All.

The LD integration level is significant and positively correlated with the response for the models that include the following variables:

- CFT MTC,
- PFT 3-mile run.

Both the HD and LD integration levels are significant and negatively correlated with the response for the models that include the following variables:

- None.

The following variables are significant in their respective models and are positively correlated with the CASEVAC time:

- Age of patches 2 and 3,
- CFT MANUF time of patch 2.

The following variables are significant in their respective models and are negatively correlated with the CASEVAC time:

- AFQT score of patches 2 and 3,
- GT score of patches 2 and 3,
- CFT MTC of patches 1 and 3,
- PFT three-mile run time of patch 1.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, the final model includes integration level only where HD has a statistically significant difference of 1.17 minutes when compared to a C group and a p-value of <0.01. This difference is an increase from the 1.15-minute difference identified in the descriptive statistics, which is a 1.74% change. The LD integration level was not significant in the final model.

C.6.3.5 M2 Mount/Dismount

We model elapsed time for the M2 Mount/Dismount as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position within the machinegun squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following variables:

- GT score,
- CFT MTC.

The LD integration level is significant and positively correlated with the response for the models that include the following variables:

- None.

Both the HD and LD integration levels are significant and negatively correlated with the response for the models that include the following variables:

- None.

The following variables are significant in their respective models and are positively correlated with the M2 mount/dismount time:

- Rifle score of patch 2.

The following variables are significant in their respective models and are negatively correlated with the M2 mount/dismount time:

- CFT MTC of patch 3,
- PFT three-mile run of patch 3.

The models for the following variables have no significant variables in the model:

- Squad leader,
- Age,
- Height,
- Weight,
- AFQT score,
- CFT MANUF.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, the final model includes integration level only where HD has a statistically significant difference of 0.41 minutes when compared to a C group and a one-sided p-value of 0.07. This difference is an increase from the 0.39-minute difference identified in the descriptive statistics, which is a 5.13% change. The LD integration level was not significant in the final model.

C.6.3.6 M240B Displace to the LOA

We model elapsed time for M240B displace to the LOA as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position within the machinegun squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

Both the HD and LD integration levels are significant and positively correlated with the response for the following models:

- All.

Both the HD and LD integration levels are significant and negatively correlated with the response for the models that include the following variables:

- None.

The following variables are significant in their respective models and are positively correlated with the M240B displace to the LOA time:

- Age of patches 1 and 2,
- Height of patch 2,
- Weight of patch 2,
- CFT MANUF time of patches 1 and 3,
- PFT three-mile run time of patch 1.

The following variables are significant in their respective models and are negatively correlated with the M240B displace to the LOA time:

- AFQT score of patch 3,
- GT score of patch 3,
- CFT MTC of patch 2,
- PFT crunches of patch 3.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes integration level only where HD has a statistically significant difference of 1.14 minutes when compared to a C group and a p-value of <0.01. This difference is an increase from the 1.00-minute difference identified in the descriptive statistics, which is a 14.00% change. The LD integration level has a statistically significant difference of 0.90 minutes when compared to a C group and a p-value of <0.01. This difference is an increase from the 0.83-minute difference identified in the descriptive statistics, which is an 8.43% change.

C.6.3.7 M2 Displacement

We model elapsed time for the M2 displacement as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position within the machinegun squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

There was one potential influential point for this task. Therefore, we model this task with and without the influential point to determine if the influential points affect the outcome.

C.6.3.7.1 M2 Displacement with potential influential points

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following variables:

- Squad leader,
- Age,
- Height,
- Weight,
- AFQT score,
- GT score,
- CFT MANUF,
- PFT crunches,
- Rifle score.

The LD integration level is significant and positively correlated with the response for the models that include the following variables:

- None.

Both the HD and LD integration levels are significant and negatively correlated with the response for the models that include the following variables:

- None.

The following variables are significant in their respective models and are positively correlated with the M2 displacement time:

- Age of patch 3,
- CFT MTC of patches 1 and 2.

The following variables are significant in their respective models and are negatively correlated with the M2 displacement time:

- AFQT score of patch 3,
- GT score of patch 3,
- PFT crunches of patch 1.

The model for the following variable has no significant variables in the model:

- PFT 3-mile run.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes integration level only where HD has a statistically significant difference of 0.31 minutes when compared to a C group and a p-value of <0.01. This difference is the same as the 0.31 minute difference identified in the descriptive statistics. The LD integration level was not significant in the final model.

C.6.3.7.2 M2 Displacement without potential influential points

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following variables:

- Squad leader,
- Age,
- Height,
- Weight,
- AFQT score,
- GT score,
- CFT MANUF,
- PFT crunches,
- Rifle score.

The LD integration level is significant and positively correlated with the response for the models that include the following variables:

- Weight,
- AFQT score.

Both the HD and LD integration levels are significant and negatively correlated with the response for the models that include the following variables:

- None.

The following variables are significant in their respective models and are positively correlated with the displacement from M2 firing position time:

- Age of patch 3,
- CFT MTC of patches 1 and 2;
- CFT MANUF time of patches 1 and 2.

The following variables are significant in their respective models and are negatively correlated with the displacement from M2 firing position time:

- Weight of patch 3,
- PFT crunches of patch 1.

The model for the following variable has no significant variables in the model:

- PFT 3-mile run.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes integration level only where HD has a statistically significant difference of 0.24 minutes when compared to a C group and a p-value of <0.01. This difference is an increase from the 0.23-minute difference identified in the descriptive statistics, which is a 4.35% change. The LD integration level was not significant in the final model.

Appendix to Annex C
0331 Supplemental Information

This appendix provides supplemental information for the 0331 portion of the GCEITF experiment. It provides information regarding the GCEITF leadership subjective comments and additional descriptive and basic inferential statistics not described in Annex C.

Section 1: GCEITF Leadership Subjective Comments

The GCEITF leadership provided comments on their observations of the experiment throughout its execution. Table C A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table C A – Summary of GCEITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement				Requesting extra breaks				Requires extra assistance				Needs no assistance				Compensating for another Marine				Gear pass off				Other				No category				Total
		C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	
7-km Hike	M	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	2	1	0	0	1	0	0	0	0	0	0	0	0	4
	F	0	1	12	13	0	6	5	11	0	1	0	1	0	0	0	0	0	0	0	0	0	1	3	4	0	0	0	0	0	0	0	0	29
	Unit	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	2	2	0	0	1	1	0	0	0	0	5
M240B Movement & Emplacement	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	F	0	5	4	9	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	11
	Unit	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
M240B Engagement	M	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	F	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	Unit	0	0	0	0	0	0	0	0	2	0	5	7	11	7	4	22	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2	31
CASEVAC to CPP	M	0	0	0	0	0	0	0	0	0	1	0	1	12	9	5	26	0	1	2	3	0	0	0	0	0	0	0	0	0	0	0	0	30
	F	0	0	0	0	0	0	0	0	0	1	6	7	0	1	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
	Unit	0	0	0	0	0	0	0	0	0	0	1	1	0	1	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
M2 Mount/Dismount	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Displace to LOA	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	0	2	0	2	0	1	0	1	0	0	0	0	0	0	0	0	6
	F	0	7	7	14	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15
	Unit	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	3
M2 Displacment	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1
	F	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Unit	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
1-km Hike	M	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	F	0	5	8	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Section 2. Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences.

This section presents results for fifteen additional 0331 tasks. Annex C contains the descriptive statistics for the remainder of the 0331 tasks. The words “metric” and “task” are used interchangeably throughout this Appendix; they both refer to the experimental task.

The two tables below display the results for fifteen additional 0331 metrics. Table C B displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent differences between integration levels.

Table C C displays ANOVA and Tukey Test results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA was conducted to compare the three groups and Tukey Tests were conducted to compare groups in pairs. If non-parametric tests were needed, Table C C displays these results instead of ANOVA and Tukey Test results. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the result was not found to be the same across all three groups. We present inferential statistics for eight additional tasks.

Table C B – 0331 Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
1-km Hike (minutes)	C	20	8.74	0.51	17.62%	19.57%	1.65%
	LD	18	10.28	0.69			
	HD	20	10.45	0.59			
7-km Hike; first km (minutes)	C	18	9.42	1.25	23.62%	29.69%	2.67%
	LD	15	11.91	1.04			
	HD	16	12.22	1.33			
7-km Hike; second km (minutes)	C	18	10.03	0.77	34.50%	39.65%	3.83%
	LD	15	13.49	1.96			
	HD	16	14.01	2.13			
7-km Hike; third km (minutes)	C	18	10.03	0.90	30.25%	49.46%	14.75%
	LD	16	13.07	2.96			
	HD	17	15.00	3.35			
7-km Hike; fourth km (minutes)	C	20	11.40	2.02	41.56%	34.45%	-5.02%
	LD	17	16.13	3.39			
	HD	18	15.32	2.81			
7-km Hike; fifth km (minutes)	C	20	13.92	2.09	31.46%	39.25%	5.92%
	LD	17	18.30	2.96			
	HD	19	19.39	4.68			
7-km Hike; sixth km (minutes)	C	20	14.36	2.76	28.22%	41.79%	10.59%
	LD	17	18.41	5.44			
	HD	20	20.36	4.30			
7-km Hike; seventh km	C	19	13.31	3.95	29.45%	25.23%	-3.26%
	LD	17	17.23	3.45			

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
(minutes)	HD	20	16.67	2.64			
M240B Movement (minutes)	C	20	1.64	0.21	27.44%	33.11%	4.45%
	LD	18	2.09	0.47			
	HD	20	2.18	0.22			
M240B Emplacement (minutes)	C	20	0.92	0.23	-10.93%	24.20%	39.44%
	LD	18	0.82	0.27			
	HD	20	1.14	0.41			
M240B Engagement Attack (hits on target)	C	20	108.15	64.65	10.79%	-17.20%	-25.27%
	LD	17	119.82	60.35			
	HD	20	89.55	52.88			
M240B Engagement C-Atk (hits on target)	C	18	92.61	43.88	15.60%	-20.46%	-31.19%
	LD	16	107.06	42.38			
	HD	18	73.67	48.96			
M2 Mount (minutes)	C	21	3.09	0.69	2.93%	9.80%	6.68%
	LD	18	3.18	0.63			
	HD	21	3.40	0.77			
M2 Dismount (minutes)	C	21	0.67	0.16	-2.55%	14.80%	17.80%
	LD	18	0.65	0.14			
	HD	21	0.76	0.19			
M2 hits on target (% hits)	C	21	42.42%	21.46%	27.24%	35.93%	6.83%
	LD	17	53.97%	20.70%			
	HD	19	57.66%	21.79%			

Table C C – 0331 ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	P-Value	80 % LCB**	80% UCB**	90% LCB**	90% UCB**
1-km Hike (minutes)	49.09 (2, 55)	< 0.01*	LD-C	< 0.01*	1.20	1.88	1.13	1.95
			HD-C	< 0.01*	1.38	2.04	1.31	2.11
			HD-LD	0.66	-0.17	0.51	-0.24	0.58
7-km Hike; first km (minutes)	27.27 (2, 46)	< 0.01*	LD-C	< 0.01*	1.74	3.22	1.59	3.37
			HD-C	< 0.01*	2.07	3.53	1.92	3.68
			HD-LD	0.75	-0.44	1.08	-0.60	1.24
7-km Hike; second km (minutes)	27.95 (2, 46)	< 0.01*	LD-C	< 0.01*	2.43	4.49	2.21	4.71
			HD-C	< 0.01*	2.96	4.99	2.75	5.20
			HD-LD	0.68	-0.55	1.58	-0.76	1.80
7-km Hike; third km (minutes)	16.25 (2, 48)	< 0.01*	LD-C	< 0.01*	1.48	4.59	1.16	4.91
			HD-C	< 0.01*	3.43	6.50	3.11	6.81
			HD-LD	0.09*	0.35	3.51	0.02	3.83
7-km Hike; fourth km (minutes)	15.97 (2, 52)	< 0.01*	LD-C	< 0.01*	3.15	6.32	2.83	6.65
			HD-C	< 0.01*	2.37	5.49	2.04	5.81

Metric	F Statistic (df)	F Test P-Value	Comparison	P-Value	80 % LCB**	80% UCB**	90% LCB**	90% UCB**
			HD-LD	0.66	-2.43	0.81	-2.77	1.15
7-km Hike; fifth km (minutes)	13.99 (2, 53)	< 0.01*	LD-C	< 0.01*	2.42	6.34	2.02	6.74
			HD-C	< 0.01*	3.56	7.37	3.17	7.76
			HD-LD	0.61	-0.90	3.07	-1.31	3.47
7-km Hike; sixth km (minutes)	10.39 (2, 54)	< 0.01*	LD-C	< 0.01*	1.62	6.48	1.12	6.98
			HD-C	< 0.01*	3.67	8.33	3.19	8.81
			HD-LD	0.35	-0.48	4.38	-0.98	4.88
7-km Hike; seventh km (minutes)	7.35 (2, 53)	< 0.01*	LD-C	< 0.01*	1.96	5.88	1.56	6.28
			HD-C	< 0.01*	1.48	5.24	1.09	5.63
			HD-LD	0.87	-2.50	1.37	-2.90	1.77
M240B Movement (minutes)	71.65‡	< 0.01‡	LD-C	< 0.01*	0.27	0.63	0.24	0.66
			HD-C	< 0.01*	0.37	0.72	0.33	0.75
			HD-LD	0.64	-0.08	0.27	-0.12	0.31
M240B Emplacement (minutes)	7.49‡	0.02‡	LD-C	0.59	-0.28	0.08	-0.31	0.11
			HD-C	0.08*	0.05	0.39	0.01	0.43
			HD-LD	0.01*	0.14	0.50	0.11	0.54
M240B Engagement Attack (hits on target)	1.23 (2, 54)	0.30	LD-C	0.82	-22.44	45.79	-29.44	52.79
			HD-C	0.59	-51.30	14.10	-58.02	20.82
			HD-LD	0.28	-64.39	3.84	-71.39	10.84
M240B Engagement C-Atk (hits on target)	2.33 (2, 49)	0.11	LD-C	0.62	-12.65	41.55	-18.24	47.14
			HD-C	0.43	-45.24	7.35	-50.66	12.77
			HD-LD	0.09*	-60.50	-6.29	-66.08	-0.71
M2 Mount (minutes)	1.03 (2, 57)	0.36	LD-C	0.92	-0.30	0.48	-0.38	0.56
			HD-C	0.35	-0.07	0.68	-0.15	0.76
			HD-LD	0.62	-0.18	0.60	-0.26	0.69
M2 Dismount (minutes)	2.78 (2, 57)	0.07*	LD-C	0.09*	-0.11	0.08	-0.13	0.10
			HD-C	0.15	0.01	0.19	-0.01	0.21
			HD-LD	0.09*	0.02	0.21	0.00	0.23
M2 hits on target (% hits)	2.79 (2, 54)	0.07*	LD-C	0.23	-0.01	0.24	-0.03	0.26
			HD-C	0.07*	0.03	0.27	0.01	0.29
			HD-LD	0.86	-0.09	0.16	-0.11	0.19

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA.
**Tukey intervals have familywise confidence of the indicated percentage, each interval is not of the given confidence level on its own.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality.
‡Indicates results presented are from Robust ANOVA due to unequal variances. The reported F-statistic is a Chi-square statistic from Robust ANOVA, and the F-test p-value is the Robust ANOVA p-value.

Additional Task Results:

7-km Hike by km. The general trend for the 7-km hike was that the difference between the LD and C groups, the HD and C groups, and the HD and LD groups increased over the course of the hike.

1-km Hike. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.34 for the C group, 0.28 for the LD group, and 0.67 for the HD group.

The C group had a mean time of 8.74 minutes. This time is statistically significantly faster than the LD group mean time of 10.28 minutes and the HD group mean time of 10.45 minutes. The LD group was 17.62% slower than the C group, and the HD group was 19.57% slower than the C group. The HD group was 1.65% slower, and this difference is not statistically significant.

- **Contextual Comments**. The implications of this task contain relevance to both the training and combat environment, as it will take integrated squads more time to conduct foot marches. The infantry T&R Manual states that, "The approach march load will be such that the average infantry Marine will be able to conduct a 20 mile hike in 8 hours with the reasonable expectation of maintaining 90% combat effectiveness." This pace equated to 2.5 mph or 4.02 km/h and 1 km traveled every 14.92 minutes if conducted with no breaks or stopping. While both groups completed the first kilometer under the required pace, the statistically significant difference in the group's pace is operationally relevant. To extrapolate this pace (not accounting for any further degradation of performance or breaks taken) over a 20-km movement, on average, it would take the C group 2:54 and the HD group 3:48.
- The operational impact of requiring additional time to move a given distance would be most operationally relevant during a movement to contact or patrolling operations. During such operations, additional time spent moving towards an objective area or area of interest has the potential to decreases the unit's element of surprise and increases the unit's chances of being detected prior to reaching their objective. This difference in pace would likely increase as the machine-gun team enhanced their combat load-out by adding ammunition or a heavy machine-gun.

M240B Movement. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.36 for the C group, 0.08 for the LD group, and 0.97 for the HD group.

The C group had a mean time of 1.64 minutes. This time is statistically significantly faster than the LD group mean time of 2.09 minutes and the HD group mean time of 2.18 minutes. The LD group was 27.44% slower than the C group, and the HD group was 33.11% slower than the C group. The HD group was 4.45% slower than the LD group, and the difference is not statistically significant.

- **Contextual Comments**. The loss in mission capability due to additional time required for movement from an ORP to an SBF depends on the maneuver element's scheme of maneuver, as well as the fire support plan of the supported maneuver commander. The enemy situation at the time of deploying a machinegun team from their ORP also greatly affects the importance of quickly reaching the SBF. The infantry T&R Manual does not specify a time standard for the completion of this task. The friendly situation briefed during the conduct of this task was that the maneuver element was waiting for the machinegun team to begin suppression prior to their movement towards the objective. In this case, time is critical and increased time taken to accomplish this task equates to longer friendly element exposure times.
- **Subjective Comments.** Throughout the conduct of this event, degraded performance of female participants was the cause of 10 GCEITF leadership subjective comments, while no comments were made due to degraded male performance. Of these 10 comments, 9 were due to falling behind/ slowing movement and 1 was made due to the participant requesting extra breaks.

M240B Emplacement. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.95 for the C group, 0.24 for the LD group, and 0.34 for the HD group.

The C group had a mean time of 0.92 minutes. This time is slower (but not statistically significantly) than the LD group mean time of 0.82 minutes and statistically significantly faster than the HD group mean time of 1.14 minutes. The LD group was 10.93% faster than the C group, and the HD group was 24.20% slower than the C group. The HD group was 39.44% slower than the LD group, and the difference is statistically significant.

- **Contextual Comments**. The loss in mission capability due to additional time required for emplacement of a machinegun depends on the maneuver element's scheme of maneuver, as well as the fire support plan of the supported maneuver commander. The enemy situation at the time of deploying a machinegun team also greatly affects the importance of quickly reaching the SBF. The infantry T&R Manual does not specify a time standard for the completion of this task. The friendly situation briefed during the conduct of this task was that the maneuver element was waiting for the machinegun team to begin suppression prior to their movement towards the objective. In this case, time is critical and increased time taken to accomplish this task equates to longer friendly element exposure times.

- **Subjective Comments.** No further comments.

Engage Targets (Attack). The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.40 for the C group, 0.17 for the LD group, and 0.03 for the HD group.

The C group had a mean of 108.15 hits on target. This number is lower (but not statistically significantly) than the LD group mean of 119.82 hits and higher (but not statistically significantly) than the HD group mean of 89.55 hits. The LD group produced 10.79% more hits than the C group, and the HD group produced 17.20% fewer hits than the C group. The HD group produced 25.27% fewer hits than the LD group, and the difference is not statistically significant.

Engage Targets (C-Attack). The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.22 for the C group, 0.59 for the LD group, and 0.16 for the HD group.

The C group had a mean of 92.61 hits on target. This number is lower (but not statistically significantly) than the LD group mean of 107.06 hits and higher (but not statistically significantly) than the HD group mean of 73.67 hits. The LD group produced 15.60% more hits than the C group, and the HD group produced 20.46% fewer hits than the C group. The HD group produced 31.19% fewer hits than the LD group, and the difference is statistically significant.

Mount M2 on Vehicle. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.54 for the C group, 0.53 for the LD group, and 0.43 for the HD group.

The C group had a mean time of 3.09 minutes. This time is faster (but not statistically significantly) than the LD group mean time of 3.18 minutes and the HD group mean time of 3.40 minute. The LD group was 2.93% slower than the C group, and the HD group was 9.80% slower than the C group. The HD group was 6.68% slower than the LD group, and the difference is not statistically significant.

Dismount M2 from Vehicle. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.01 for the C group, 0.24 for the LD group, and 0.05 for the HD group.

The C group had a mean time of 0.67 minutes. This time is statistically significantly slower than the LD group mean time of 0.65 minutes; the HD group had a mean time of 0.76 minutes. The LD group was 2.55% faster than the C group, and the HD group was 14.80% slower than the C group. The HD group was 17.80% slower than the LD group, and the difference is statistically significant in a Tukey test.

M2 hits on target. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.25 for the C group, 0.94 for the LD group, and 0.07 for the HD group.

The C group had a mean percentage hit of 42.42%. This percent is lower (but not statistically significantly) than the LD group percentage hit of 53.97% and statistically significantly lower than the HD group mean of 57.60% percentage hit with a SD of 21.79%. The LD group was 27.24% more accurate than the C group, and the HD group was 35.93% more accurate than the C group. The HD group was 6.83% more accurate than the LD group, and the difference is not statistically significant.

Annex D.

Infantry Mortarman (MOS 0341)

This annex details the Infantry Mortarman (MOS 0341) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 2 March – 26 April 2015 at Range 107 and Range 110, at the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, California. The sections outline the Infantry Mortarman Scheme of Maneuver (SOM), Limitations, Deviations, Data Set Description, Descriptive and Basic Inferential Statistics, and Modeling Results.

D.1 Scheme of Maneuver

D.1.1 Experimental Cycle Overview

The Infantry Mortarman (MOS 0341) assessment of the GCEITF took place in a field environment at MCAGCC, Twentynine Palms, California. The assessment consisted of 21 trial cycles, each of which was a 2-day test cycle, conducted over the course of 55 days. After every 4 days of trials, the Marines spent 1 recovery day at Camp Wilson. Every mortar team/squad consisted of four volunteers and a direct-assignment (non-volunteer) squad leader. Each member of the team/squad was trained to fill each billet within the team: gunner, assistant gunner, ammunition man 1, and ammunition man 2. The assessment was executed under the supervision of MCOTEA functional test managers and a range Officer in Charge (OIC)/Range Safety Officer (RSO) from the GCEITF.

D.1.2 Experimental Details

The 2-day 0341 assessment replicated offensive and defensive tasks. The 0341s began each cycle on the Day 1/Offensive task, followed by the Day 2/Defensive task on the subsequent day. Marine volunteers formed up as a 60-mm mortar team on the first (offensive) day and as an 81-mm mortar squad on the second (defensive) day; on both days, two four-Marine squads executed the tasks: a control (C) non-integrated team/squad and a high-density (HD) integrated team/squad with two females.

Day 1 of the trial cycle was executed on Range 107 and consisted of one 60-mm mortar team with two 2-Marine units moving together and firing the 60-mm mortar in the handheld mode from two different Mortar Firing Positions (MFPs). After the engagement, the two 2-Marine units aggregated as a 4-Marine 60-mm mortar team to conduct a 100-meter casualty evacuation (CASEVAC) of a 220-lb dummy.

Day 2 of the trial cycle was executed on Range 110. Marines started the day with a 7-km forced march from Range 107 to Range 110 while wearing an approach load, carrying personal weapons, and the 81-mm mortar crew-served weapon system. After arriving at Range 110, the 81-mm mortar squad engaged two targets from a MFP. After the engagement portion of the trial, which consisted of two 5-round fire missions, the mortar squad moved from their engagement position back to the Objective Rally Point (ORP). This concluded Day 2, and the Marines reorganized into new teams/squads for the next trial cycle.

D.1.3 Additional Context

Throughout the assessment, Marines bivouacked at Range 107, sleeping in 2-man tents. During trial execution, Marines wore/carried prescribed loads for each task. Weighing packs each day prior to the 7-km forced march ensured consistency. After each trial day, Marines operated under Company leadership, performing minimal physically demanding tasks. Marines not part of an assessed squad conducted the same experimental subtasks after the assessed squads to ensure equity between individuals participating in a trial cycle and those not chosen for that particular cycle. These tasks are discussed in detail in the loading section below.

Fatigue surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. It allows for outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) to be accounted for in the analysis of the performance data. Workload surveys collected the volunteers' perceived average and maximal level of exertion during the performance-specified tasks. Cohesion surveys provided a way to collect subjective data relating to each team/squad's ability to work as a team and their overall perspective on the cohesiveness of the team/squad.

D.1.4 Experimental Tasks

D.1.4.1 1-km Movement

Assaulting an enemy location never starts from a static position; first, a movement must be conducted to the assault position (AP). The distance from the line of departure to the AP depends on myriad factors. Based on time and space constraints, this distance was set at just under 1 km for the experimental event. Each mortar team moved this distance as quickly as possible while carrying an assault load in addition to two 60-mm mortar tubes, two M8 auxiliary baseplates, and six 60-mm mortar rounds divided among the members.

D.1.4.2 60-mm Mortar Engagement

Indirect fire is used to suppress the enemy from farther than direct-fire weapon range and to prepare the battlefield to allow the rifle squad to complete its mission of locate, close-with, and destroy the enemy by fire and maneuver. During conduct of an attack, 60-mm mortars are often employed in the handheld mode as an assault weapon and can close distance with the enemy while still ensuring they can engage. Each two-Marine mortar team engaged one target from two different MFPs with three rounds per MFP. This task determined the mortar team's accuracy while engaging from two different positions.

D.1.4.3 Displace to the LOA

After providing indirect fire in support of an assault, mortar teams typically move with the 0311 squad to a limit of advance (LOA) or displace to a position of advantage. During the assessment, each mortar team displaced from their second MFP approximately 300 m to a LOA/CASEVAC position. This task determined the time for a mortar team to displace with all equipment.

D.1.4.4 CASEVAC

Casualties are an inevitable part of conducting combat operations. Units train to become proficient in triaging, handling, and transporting a casualty. When a casualty is sustained, it is essential to move with a sense of urgency to get the injured Marine to the appropriate level of care. At the conclusion of the live-fire and counterattack, each mortar team moved a 220-lb dummy 100 m from a position of cover to a casualty collection point (CCP) while transporting their assault packs and crew-served weapon system (two 60-mm mortar tubes and M8 auxiliary baseplate). The mortar squads could use a variety of techniques for transport but had to carry the dummy off the ground. This task determined the team's proficiency in moving a simulated casualty to a CCP. After the CASEVAC, each Marine took a fatigue and workload survey to assess overall fatigue and workload of the entire offensive task.

D.1.4.5 7-km Hike

Moving under a load is one of the most fundamental tasks of an infantry unit; it is both physically and mentally demanding. Infantry units must move through all sorts of terrain on foot. Units train by conducting a forced march with an approach load at a sustained rate of march. The Infantry Training and Readiness T&R Manual states, "the approach load will be such that the average infantry Marine will be able to conduct a 20-mile hike in 8 hours with the reasonable expectation of maintaining 90% combat effectiveness." During the GCEITF assessment, each mortar squad moved a distance of 7.2 km from Range 107 to Range 110. This route was flat (minimal elevation change) and conducted on an unimproved surface with varying degrees of conditions (compact dirt and loose sand). Each mortar squad moved as fast as the slowest person while carrying an approach load and the 81-mm mortar system, which consisted of the 81-mm mortar tube, bipods, baseplate, and sight unit spread-loaded across all four members. This task determined the squad's rate of movement over a 7.2-km route while carrying the approach load and crew-served weapons. Each Marine took a fatigue and workload survey after completion of the 7.2-km hike.

D.1.4.6 81-mm Mortar Engagement

Indirect fire is used to suppress the enemy from farther than direct-fire weapon range and prepare the battlefield to allow the rifle squad to complete its mission of locate, close-with, and destroy the enemy by fire and maneuver. During the defensive day, the 81-mm mortar squad conducted an emplacement of the crew-served weapon, engaged two different targets with five 81-mm rounds each, and conducted a displacement of the mortar system. The mortar squad engaged the targets using the direct-lay method and visually acquired and adjusted all rounds during each fire mission. This task determined the 81-mm mortar squad's accuracy of engaging targets from a static position. At the conclusion of the 81-mm mortar engagement, the Marines took a fatigue and workload survey to assess their fatigue and workload for the mortar engagement task (see GCEITF EAP, Annex D).

At the completion of the 2-day cycle, the Marines took a cohesion survey to record their cohesion during execution of the 2-day trial cycle (see GCEITF EAP, Annex M).

D.1.5 Loading Plan

The loading plan ensured, to the greatest extent possible, equity of physical activity among all volunteers throughout the duration of the experimental assessment. Every trial and task was conducted in the same manner and sequence to ensure consistency. Due to the number of volunteers, a handful of Marines were not part of an assessed squad each 2-day cycle. Collaboration with Company leadership determined that the best method of loading non-assessed Marines was to form them into a quasi-team/squad and have them perform the same tasks as an assessed mortar team/squad on the offense and defense test days, respectively. This allowed the Marines to experience the same conditions and physical strain. Minor modifications were permitted because of the reduced size of the squad, e.g., conducting a trial as the fifth Marine in a four-Marine element because not enough individuals were available to form another team.

D.1.6 Scheme of Maneuver Summary

The 0341 experiment consisted of a 2-day trial cycle comprising an offensive and defensive day. The offensive day involved three subtasks based around supporting a live-fire squad attack: 1-km movement, 60-mm mortar engagement, and CASEVAC. The defensive day involved two subtasks: a 7-km forced march and 81-mm mortar engagement. During the course of the experiment, the 0341 (60-mm mortar team and 81-mm mortar squad) executed 2 pilot trial cycles and 21 record trial cycles. During trial execution, Marines rotated through every billet within the mortar elements and changed position on the crew-served weapon system.

D.2 Limitations

D.2.1 0341 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a similar manner to those in an operational environment. Under certain situations, artificial limitations or interruptions were introduced that changed or altered the way a task would normally be performed. Although these limitations represent a degree of artificiality, they do not detract significantly from our ability to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for the 0341 assessment.

D.2.2 Relative Difficulty of Record Test

The 0341 GCEITF assessment was designed to gather data associated with some of the most physically demanding tasks within the 0341 MOS. These tasks in isolation do not fully replicate life experienced by a Marine during a typical field exercise (FEX) or a combat environment and did not obtain the maximum cumulative load that could be placed on an Infantry Marine. With limited time available, only selected 0341 tasks were assessed. Other tasks/duties outside the assessment were minimized due to specific experimental constraints and human factors. During a typical FEX, it is common for Marines to conduct 24-hour operations that include day and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. The offensive day SOM took squads approximately 1 hour to complete, and the defensive day SOM took approximately 3 hours to complete. Outside the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

Another concern in designing the 0341 assessment was making it achievable and sustainable for a 60-day period. The 7-km forced march distance was selected based on the training time available prior to the assessment. However, many of the loads carried were decreased; the crew-served load was lightened by dropping the mortar rounds. The Marines were authorized 1 day off after every 4 days of training. This artificial recovery period is not achievable when conducting training or combat operations.

A final factor affecting the relative difficulty of the record test pertains to the intangible physiological impact of the volunteers' ability to drop on request (DOR) at any point during a trial. Any time a volunteer dropped during a trial cycle, that squad/team performed the following subtasks with fewer personnel. This factor could have affected the cohesion of each squad and influenced its performance.

D.2.3 Geometries of Fire

Outside of normal safety limitations, several artificialities were present during the 60-mm mortar live-fire portion of the assessment. Although it is doctrinal for mortarmen to employ the 60-mm mortar in the handheld mode while imbedded with a rifle squad, the assessed mortar squads were separated by approximately 500 meters and fired different directions to prevent any interference between units. The loss of tactical realism in basing fires off a maneuver element resulted in this task being less challenging than in training or combat, in which the team leader and gunner must make intuitive decisions regarding selection of a mortar firing positon, rate of fire, and target precedence.

D.2.4 Number of Volunteer Participants

For the 0341 experiment, nine male and four female volunteers began the experiment; by the end, seven males and four females completed the assessment. The results presented in this annex are based on the performance of 11 to 13 Marines.

D.2.5 Limitations Summary

The 0341 assessment was designed to replicate realistic training in a field environment. The end-state was to create an experiment in which the volunteers felt they were conducting realistic and operationally relevant tasks, but unavoidable limitations to the assessed tasks and non-assessed operating environment introduced a level of artificiality not normally present in a field training or combat environment.

D.3 Deviations

D.3.1 60-mm Mortar Employment

The EAP for the four-Marine 60-mm mortar team called for the employment of a single 60-mm mortar tube. Because the employment of the 60-mm mortar is only a 2-Marine task, Company leadership recommended employing two 60-mm mortar tubes within each team. The entire assessment consisted of a 4-Marine 60-mm mortar team employing two tubes.

D.3.2 81-mm Mortar Ammunition

Ammunition allocation for the 81-mm mortar employment was C875 (practice rounds). These rounds were used from 2 March to 27 March 2015. The practice rounds had a much smaller signature upon impact than high-explosive (HE) rounds, and, on windy days, it proved difficult to mark the exact location of impact with a vector/dagger. A request for 81-mm HE rounds was submitted and approved. From 28 March to 25 April 2015, C868 (HE rounds) were used. An ammunition shortfall required using the C875 practice rounds for the second fire mission for both mortar teams on 26 April 2015 to complete the assessment.

D.4 Data Set Description

D.4.1 Data Set Overview

The 0341 portion of the experiment consisted of 2 pilot trial cycles and 21 record trial cycles. The pilot trial cycle was conducted from 2 March to 6 March 2015. Pilot trial data are not used in the analysis due to variations in the conduct of the experiment. We based all analysis on the 21 record trial cycles executed from 7 March to 26 April 2015.

D.4.2 Record Test Volunteers

At the beginning of the first record trial cycle, there were nine male 0341 volunteers and four female volunteers. There were two male Marines who voluntarily withdrew, or were involuntarily withdrawn, during the execution of the experiment. The final number of volunteers was seven males and four females.

D.4.3 Planned, Executed, and Analyzed Trial Cycles

Table D-1 displays number of trial cycles planned, executed, and analyzed by task. The planned number of trial cycles for the 0341 MOS per Section 7.5.3 of GCEITF EAP was 120, or 40 trial cycles per planned integration level (C, low-density, and HD). The original plan called for six teams/squads per day (two per integration level) over the 20 trial cycles. However, due to the number of Marines who voluntarily withdrew or were involuntarily withdrawn from the experiment prior to the execution of the first record trial cycle, only one team/squad of each of the C and HD integration levels remained. We chose to keep the HD integration level to employ more Marines in the experiment.

Although the number of male and female volunteers did not support the ability to execute another C group during a trial cycle, there were times when another HD group could execute additional whole or partial trials. A partial trial was a trial in which a two-Marine unit could execute the 60-mm mortar engagement on R107 as these tasks were collected at the two-Marine unit level. The planned number of trials in Table D-1 reflects 21 planned trial cycles for each integration level and accounts for the extra trials the HD integration level completed.

There are several occurrences of missing data for the 7-km hike by individual kilometer. The individual kilometer times were derived from GPS data. Early in the experiment, the Garmin GPSs were set to record a volunteer’s position every second. Due to the storage space on the GPS and the length of the trial, when volunteers executed the 7-km hike and then follow-on tasks, the GPS could not hold all the data. Therefore, it overwrote the hike data. Once the problem was found, the GPSs were corrected to record location every 2 seconds.

Table D-1. 0341 Planned, Executed, and Analyzed Trial Cycles[1]

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
7-km Hike	C	21	21	21	
	HD	21	23	23	
60-mm Mortar Engagement	C	21	21	18	Remove Mar 12 due to missing data. Remove Mar 14 due to TIR. Remove Mar 27 high outlier.
	HD	21	23	20	Remove Mar 12 and Mar 14 due to missing data. Remove Mar 9 run 2 due to TIR
Displace to LOA	C	21	21	21	
	HD	21	23	23	
CASEVAC	C	21	21	20	High outlier: Mar 7
	HD	21	23	23	
81-mm Mortar Emplacement and Displacement	C	21	21	21	
	HD	21	23	23	
1-km Hike	C	21	21	21	
	HD	21	23	23	
7-km Hike; 1km Time	C	21	21	20	No data: Mar 18
	HD	21	23	21	No data: run 2 Mar 10, Mar 18
7-km Hike; 2km Time	C	21	21	20	No data: Mar 18
	HD	21	23	21	No data: run 2 Mar 10, Mar 18
7-km Hike; 3km Time	C	21	21	20	No data: Mar 18
	HD	21	23	21	No data: run 2 Mar 10, Mar 18
7-km Hike; 4km Time	C	21	21	20	No data: Mar 18
	HD	21	23	22	No data: Mar 18
7-km Hike; 5km Time	C	21	21	20	No data: Mar 18
	HD	21	23	21	High outlier: Apr 7; No data: Mar 18
7-km Hike; 6km Time	C	21	21	20	No data: Mar 18
	HD	21	23	22	No data: Mar 18
7-km Hike; 7km Time	C	21	21	20	No data: Mar 18
	HD	21	23	22	No data: Mar 18
Fire & Movement to MFP; Movement to MFP #1	C	65	65	65	
	HD	21	32	32	
Fire & Movement to MFP; First Three Rounds	C	21	21	19	Missing shots Mar 12. Remove Mar 14 due to TIR.
	HD	21	23	20	Missing shots Mar 12, Mar 14. Remove Mar 9 run 2 due to TIR.
Fire & Movement to MFP; Movement to MFP #2	C	65	65	65	
	HD	21	32	32	
Fire & Movement to MFP; Second Three Rounds	C	21	21	19	Remove Mar 14 due to TIR. Remove Mar 27 high outlier.
	HD	21	23	22	Remove Mar 9 run 2 due to TIR.

[1] A TIR in this table refers to a Test Incident Report, which is a report the test team or direct assignment leaders completed when an incident occurred that affected the natural execution of a trial. If a data point is removed due to a TIR, it is because the TIR affected the data in such a way that it is not comparable to the rest of the data set.

D.5 Descriptive and Basic Inferential Statistics

D.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common infantry tasks and are indicative of unit-level proficiency during either field exercises or combat operations. This section presents results for 6 out of 17 tasks. The Appendix to this Annex contains the descriptive statistics for the remainder of the 0341 tasks. The words “metric” and “task” are used interchangeably throughout this Annex; they both refer to the experimental task.

Each team/squad consisted of four volunteer Marines and a direct assignment (non-volunteer) squad leader. There were two integration levels for all tasks: a C group was non-gender integrated, and a HD group was gender integrated with two female Marines.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA) (or non-parametric tests as necessary), and scatter plots. The subsequent sections cover each task in detail. Lastly, contextual comments, additional insights, and subjective comments (as applicable) tying to each experimental task are incorporated.

Use caution when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks, such as distances, techniques, leadership, load carried, group size, and group composition.

D.5.2 0341 Selected Task Descriptive Statistics Results

The two tables below display the results for the six selected 0341 metrics. Table D-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations. Table D-3 displays ANOVA results, including metrics and integration levels, p-values suggesting statistical significance, integration level elapsed-time differences, and percentage differences between integration levels. For each task, an ANOVA and t-test were conducted to compare the two groups. If non-parametric tests were needed, Table D-3 displays these results instead of ANOVA and t-test results. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the HD group is different from that in the C group.

Table D-2. 0341 Selected Task Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD
7-km Hike (minutes)*	C	21	93.49	6.95
	HD	23	112.46	10.17
60-mm Mortar Engagement (minutes)*	C	18	6.75	0.70
	HD	20	7.24	0.72

Metric	Integration Level	Sample Size	Mean	SD
Displace to LOA (minutes)*	C	21	2.71	0.30
	HD	23	3.19	0.40
CASEVAC (minutes)	C	20	1.69	0.51
	HD	23	1.77	0.44
81-mm Mortar Emplacement & Displacement (minutes)	C	21	2.56	0.91
	HD	23	2.51	0.92
1-km Hike (minutes)*	C	21	9.51	0.61
	HD	23	10.89	0.63

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table D-3. 0341 Selected Task ANOVA and Welch's T-Test Results

Metric	F statistic (df)	F Test P-value	Comparison	Difference	% Difference	2-sided P-Value	1-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
7-km Hike (minutes)*	51.20 (1, 42)	< 0.01*	HD-C	18.97	20.29%	< 0.01*	< 0.01*	15.57	22.37	14.58	23.36
60-mm Mortar Engagement (minutes)*	4.51 (1, 36)	0.04*	HD-C	0.49	7.30%	0.04*	0.02*	0.19	0.79	0.10	0.88
Displace to LOA (minutes)*	20.37 (1, 42)	< 0.01*	HD-C	0.48	17.83%	< 0.01*	< 0.01*	0.35	0.62	0.31	0.66
CASEVAC (minutes)	0.31 (1, 41)	0.58	HD-C	0.08	4.75%	0.58	0.29	-0.11	0.27	-0.17	0.33
81-mm Mortar Emplacement & Displacement (minutes)	0.04 (1, 42)	0.85	HD-C	-0.05	-2.10%	0.88†	0.44†	-0.33†	0.32†	-0.38†	0.37†
1-km Hike (minutes)*	54.21 (1, 42)	< 0.01*	HD-C	1.48	14.49%	< 0.01*	< 0.01*	1.14	1.62	1.06	1.69

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality.

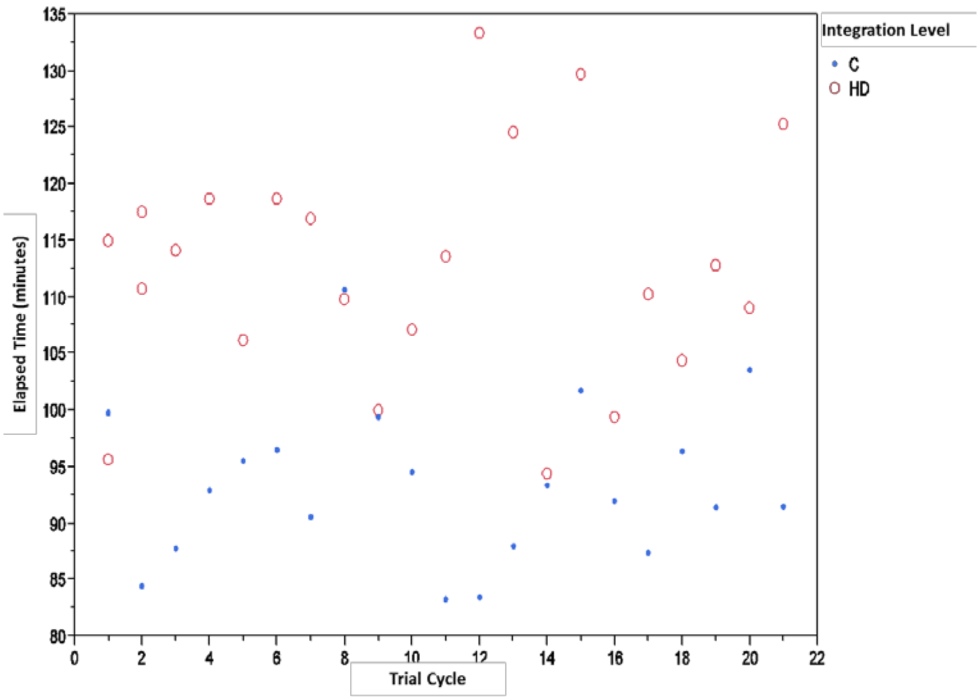
D.5.2.1 7-km Hike

D.5.2.1.1 7-km Hike Overview

This experimental task assessed a squad of four Marines moving 7.2 km while each Marine carried an approach load, an individual weapon (M-4), and a portion of the crew-serve weapon load. The crew-serve load consisted of the 81-mm mortar tube, baseplate, bipod, and sight box, resulting in a cumulative load of 118-125 lb per Marine. The recorded time for this task started when the squad departed the Range 107 start point and stopped when the squad arrived at the Range 110 stop point. The squads moved as fast as the slowest person and could take as many breaks as necessary.

Figure D-1 displays all 0341 7-km hike time data. All data on the scatter plot are valid for analysis.

Figure D-1. 7-km Hike



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.63 for the C group and 0.93 for the HD group.

The C group had a mean time of 93.49 minutes. This time is statistically significantly faster than the HD mean time of 112.46 minutes. This difference results in a 20.29% (18.97-minute) degradation in hike time between the groups. The HD group had greater variability, as shown by the 3.22-minute increase in standard deviation (SD) (6.95 minutes for the C group and 10.17 minutes for the HD group). See Table D-2 and Table D-3 for detailed analytical results.

D.5.2.1.2 7-km Hike Contextual Comments

D.5.2.1.2.1 USMC Hike Standards

The 7-km hike is a task that the following MOSs completed during the experiment: 0311, 0331, 0341, 0351 and 0352, Provisional Infantry, Provisional Machine Gunners, and Combat Engineers. There are varying standards to which we can compare this result. The following sections define those standards, as well as the one we chose as a comparison.

The Infantry T&R Manual (30 Aug 2013) identifies minimum standards that Marines must be able to perform in combat, to include standards for tactical marches. In

Chapter 8 of the Infantry T&R Manual in the MOS 0300 Individual Events section, task "0300-COND-1001: March under an approach load" is applicable to all 03XX MOSs, ranks PVT – LtCol. The condition and standard established by this task is: "Given an assignment as a member of a squad, individual weapon, and an approach load, complete a 20 kilometer march in under 5 hours." The march pace required by this standard is 4 kilometers per hour (km/h). In Chapter 9 of the Infantry T&R Manual in the MOS 0302/0369 Individual Events section, tasks "0302-OPS-2001: Lead an approach march" and "0369-OPS-2501: Lead an approach march" are applicable to MOS 0302 and 0369, ranks SSgt – MGySgt and 2ndLt – LtCol. The condition and standard established by these tasks is: "Given a mission, time constraints, an approach march load, organic weapons, and a route, move 24.8 miles (40 km) in a time limit of 8 hours." The march pace required by this standard is 5 km/h. Appendix E of the Infantry T&R Manual (Load Terms and Definitions) states: "The approach march load will be such that the average infantry Marine will be able to conduct a 20 mile hike in 8-hours with the reasonable expectation of maintaining 90% combat effectiveness." The march pace required by this definition is 4.02 km/h.

Chapter 3 of Fleet Marine Force Reference Publication (FMFRP) 3-02A, Marine Physical Readiness Training for Combat (16 Jun 2004) states: "The normal pace is 30 inches. A pace of 30 inches and a cadence of 106 steps per minute result in a speed of 4.8 km/h or 3 mi/h and a rate of 4 km/h or 2.5 mi/h if a 10-minute rest halt per hour is taken." Common Infantry practice today is to hike for 50 minutes and take a 10-minute break, while maintaining an overall pace of 4 km/h (resulting in a hiking pace of 4.8 km/h).

Driven by the need to pick an evaluative reference standard, this report follows the T&R Manual's intent to establish minimum standards and uses the 4-km/h march pace for a 20-km march established by task 0300-COND-1001 and supported by the definition of an approach load. Further, although an established reference standard is required to anchor observed performance to the T&R program, more important information is provided by performance differences observed between gender-integrated and non-gender-integrated units.

D.5.2.1.2.2 0341 7-km Hike Pace

The difference in 7-km hike times between the C and HD group is relevant to both the training and combat environment as it will take integrated squads more time to conduct foot marches. Per the tactical march standards noted above, the Marine Corps standard of hiking is 4.0 km/h. The HD group failed to meet this standard. The C group average pace was 4.62 km/h, and the HD group average pace was 3.84 km/h, finishing 19.0 minutes behind the C group. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), it would take the C group 4 hours and 20 minutes to complete the hike, and it would take the HD group 5 hours and 13 minutes to complete the hike, finishing 52.76 minutes behind the C group.

Furthermore, on any given day (under the same environmental conditions), the C group was faster than the HD group 90.5% of the time (19 of 21 trial cycles). Based on the standard deviations, the greater variability of the HD group leads to greater uncertainty and less confidence in their future performance from the mean.

D.5.2.1.3 7-km Hike Additional Insights

Based on the USMC standard of a 4-km/h pace over a 7.20-km route (which would result in a 108-minute hike completion time over the 7.20 km), the HD group was 4.46 minutes slower than that standard. In a battlefield situation, in which speed is essential, this delay is advantageous for the enemy. An enemy maneuvering at 4 km/h would have the time to move 297 meters, shift indirect fires from preplanned targets, commit the employment of their least engaged unit, or conduct a spoiling attack.

Marine Corps Doctrinal Publications (MCDPs) consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to “Being Faster” and states, “Physical speed, moving more miles per hour, is a powerful weapon in itself.” MCDP-6 Command and Control also speaks to speed relative to the enemy and states, “The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results.” Further insights may be gleaned from the Appendix, which shows the difference in speed by kilometer. In general, the difference in performance increased as the movement got longer.

D.5.2.1.4 7km-Hike Subjective Comments

For subjective comments relating to this task, see the Appendix.

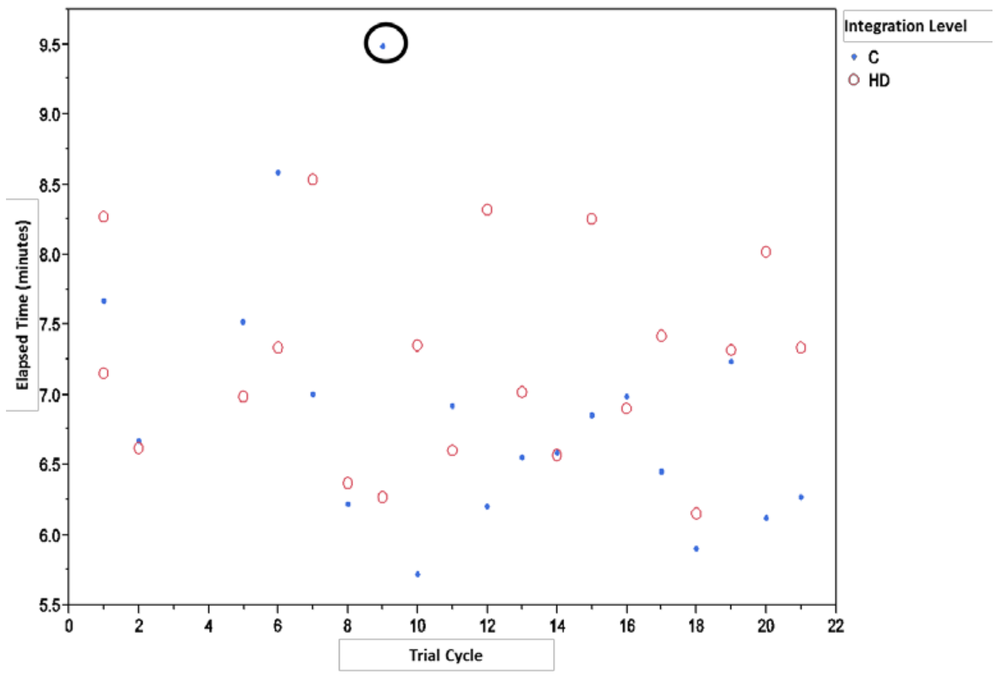
D.5.2.2 60-mm Mortar Engagement

D.5.2.2.1 60-mm Mortar Engagement Overview

This experimental task assessed the time for a four-Marine 60-mm mortar team (two mortar tubes) to move a total distance of approximately 200 meters and engage from two different MFPs. Each tube moved simultaneously, employing the mortar in the handheld method. The team moved approximately 100 meters to MFP #1, where they fired three rounds, and then moved approximately 100 meters to MFP #2, where they fired three rounds. The recorded time started immediately upon completing the 1-km movement and stopped when the last (6th) round was fired at MFP #2.

Figure D-2 displays all 0341 60-mm mortar-engagement time data. The C group on trial-cycle 9 was a high outlier due to a weapons malfunction and removed for analysis. This data point is circled in black. With the exception of this data point, all data on the scatter plot are valid for analysis.

Figure D-2. 60-mm Mortar Engagement



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.33 for the C group and 0.21 for the HD group.

The C group had a mean time of 6.75 minutes. This time is statistically significantly faster than the HD mean time of 7.24 minutes. This difference results in a7.30%, or 0.49-minute, degradation in time between the groups. See Table D-2 and Table D-3 for detailed analytical results.

D.5.2.2.2 60-mm Mortar Engagement Contextual Comments

This result is relevant to the combat environment, as it will take integrated squads more time to move and deliver supporting fires. On average, it took the HD group 29.4 seconds longer, a degradation of 7.30%, to conduct this movement and engagement. One factor slightly masking the results is the time of flight (TOF) for each mortar round. Because the TOF was a constant for every squad, the actual difference in performance occurred during a smaller time window when the Marines were actually performing work. Omitting the TOF further exaggerates the difference between the two groups. Consider, for instance, the average TOF from MFP #1 and MFP #2 was 27 seconds. Assuming each tube fired all six rounds simultaneously (a conservative assumption), the total amount of time the team would be waiting would be 1.67 minutes (waiting on the 1st, 2nd, 4th, and 5th rounds to impact prior to adjusting fire). The resultant engagement time for the C group is 5.08 minutes and for the HD group, 5.57 minutes. Based on this information, the HD group was closer to 9.65% slower than the C group.

D.5.2.2.3 60-mm Mortar Engagement Additional Insights

A purely objective evaluation of a 29.4-second average delay is elusive but may possess some practical significance. Considering the 650 rounds-per-minute sustained rate of fire for the RPK-74 machinegun, a single enemy fighter would have the opportunity to fire 318 rounds against Marines conducting an assault in an integrated squad. The resultant trade in casualty exchange could be significant.

D.5.2.2.4 60-mm Mortar Engagement Subjective Comments

For subjective comments relating to this task, see the Appendix.

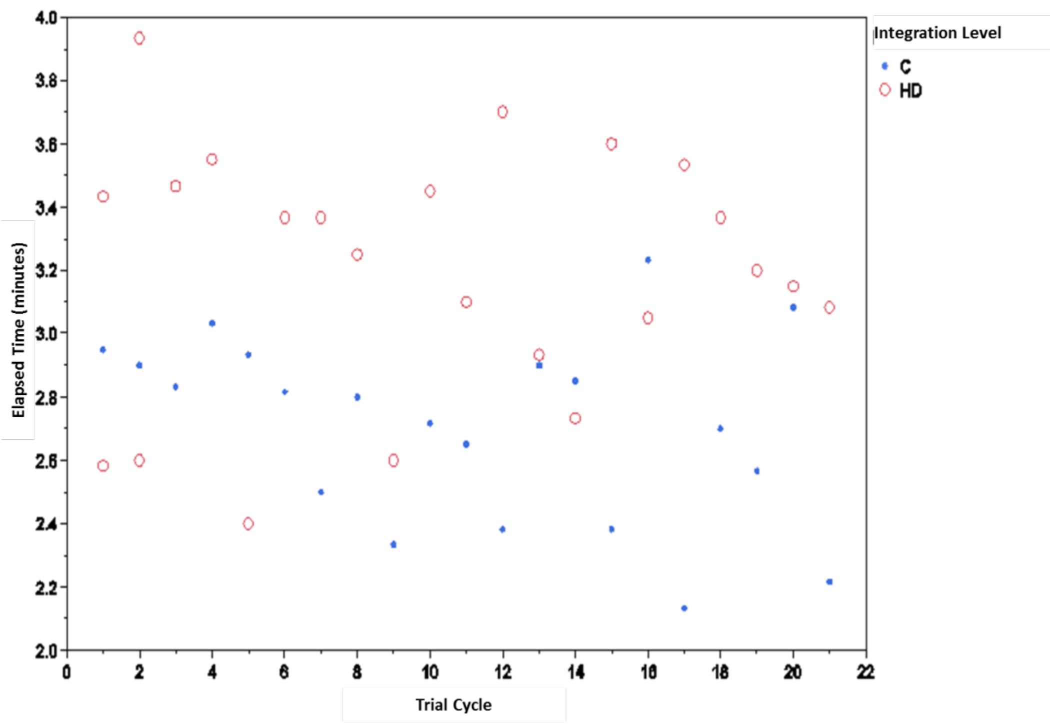
D.5.2.3 Displace to LOA

D.5.2.3.1 Displace to LOA Overview

This experimental task assessed the time for a 60-mm mortar team to displace approximately 300 meters from MFP #2 to an LOA immediately after conducting an assault. The recorded time started after the sixth round was fired by the team and stopped when the last member of the team arrived at the LOA.

Figure D-3 displays all 0341 displace to LOA time data. All data on the scatter plot are valid for analysis.

Figure D-3. Displace to LOA



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.64 for the C group and 0.43 for the HD group.

The C group had a mean time of 2.71 minutes. This time is statistically significantly faster than the HD mean time of 3.19 minutes. This difference results in a17.83%, or 0.48-minute, degradation in hike time between the groups. The HD group had greater variability, as shown by the 0.10-minute increase in SD (0.30 minutes for the C group and 0.40 minutes for the HD group). See Table D-2 and Table D-3 for detailed analytical results.

D.5.2.3.2 Displace to LOA Contextual Comments

The ability to close with the objective after having provided indirect fires is crucial to maintaining the momentum during offensive operations. In this case, the integrated squads would have taken 23.4 seconds longer to begin providing additional indirect fire support from the LOA, a degradation of 17.83%. On any given day (under the same environmental conditions), the C group was faster than the HD group 76.2% of the time (16 of 21 trial cycles).

D.5.2.3.3 Displace to LOA Additional Insights

A purely objective evaluation of 23.4 seconds is elusive but may possess some practical significance when considering a unit in contact. Consider an enemy 82-mm mortar system outside the range of friendly direct fire weapons (1800 m for .50 cal). In lieu of an 82-mm mortar firing table, an 81-mm mortar firing table reveals the maximum range on charge 1 is 1920 meters with a TOF of 20 seconds. Each enemy mortar tube would have the time to fire a round, make an adjustment, and fire another round by the time an integrated 60-mm mortar team is prepared to fire. The resultant trade in casualty exchange could be significant.

D.5.2.3.4 Displace to LOA Subjective Comments

For subjective comments relating to this task, see the Appendix.

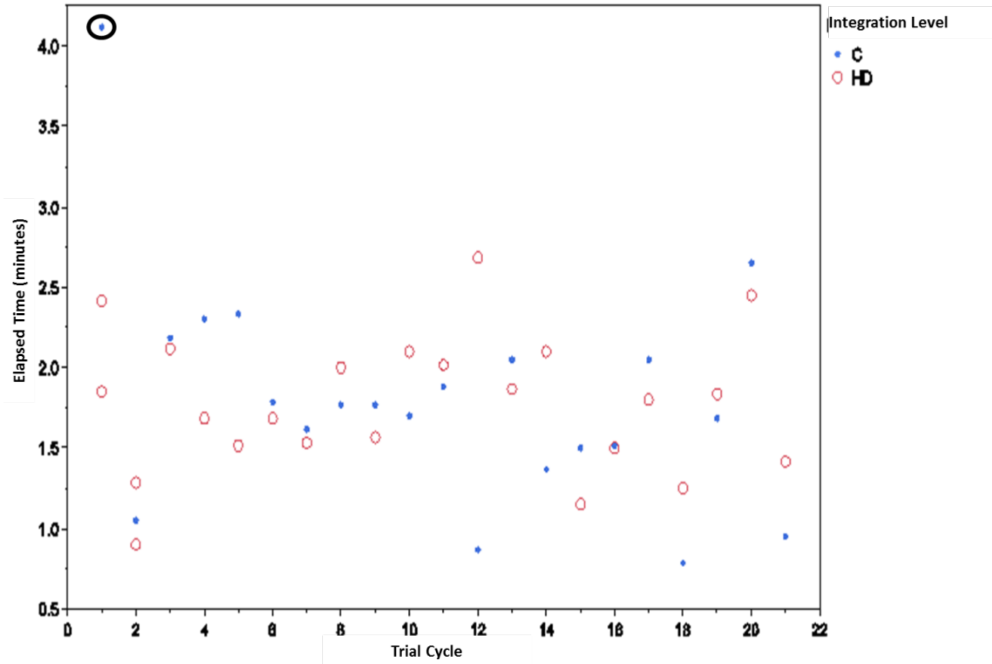
D.5.2.4 CASEVAC

D.5.2.4.1 CASEVAC Overview

This experimental task assessed the 60-mm mortar team's ability to move a 220-lb dummy a distance of 100 meters to a CCP while wearing a fighting load, individual weapon, and two 60-mm mortar tubes. The team conducted this task at the conclusion of the 60-mm mortar engagement and movement to the LOA. Teams could use a variety of techniques, but they had to move all personnel and gear the entire distance and carry the dummy off the ground. Time started when Marines touched the dummy and stopped when the dummy and all members of the fireteam arrived at the CCP.

Figure D-4 displays all 0341 CASEVAC time data. The C group on trial cycle 1 was a high outlier and removed for analysis. This data point is circled in black. With the exception of this data point, all data on the scatter plot are valid for analysis.

Figure D-4. CASEVAC



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.73 for the C group and 0.99 for the HD group.

The C group had a mean time of 1.69 minutes. This time is faster (but not statistically significant) than the HD mean time of 1.77 minutes. This difference results in a 4.75%, or 0.08-minute, degradation in time between the groups. See Table D-2 and Table D-3 for detailed analytical results.

D.5.2.4.2 CASEVAC Contextual Comments

The implications of this task contain relevance to both the training and combat environment as a casualty must be moved expediently to a higher echelon of medical care. The data demonstrates that integrated squads took 4.8 seconds longer, a degradation of 4.75%.

D.5.2.4.3 CASEVAC Additional Insights

Although the “Golden Hour” is a common medical planning construct for C2 and logistical support, medical literature supports a “Platinum Ten” philosophy of first response. The U.S. National Library of Medicine references a French article that espouses, “on the battlefield, the majority of casualties die within 10 minutes of the trauma.” (*Wounded in Action: The Platinum Ten Minutes and the Golden Hour*, Daban) The fundamental principle is that a patient needs to be correctly triaged and moved to medical care as fast as possible. Any time degradation will reduce the probability of survival.

D.5.2.4.4 CASEVAC Subjective Comments

The GCEITF leadership subjective comment log sheds further light on the results as it indicates this task was primarily conducted by an individual Marine using the fireman carry technique. Therefore, the results are highly indicative of the strength/performance of an individual rather than the effort of the team as a whole. A total of 36 comments indicated that an individual Marine carried the casualty for the entire event, 18 of which applied to the C group and 18 that applied to the HD group.

In analyzing frequency of subjective comments of the HD trials, a single male carried the dummy for 16 trials and a single female carried the dummy for 2 trials. On both occasions, it was the same female Marine. This documentation indicates that female Marines were not carrying as much of the cumulative load as their male counterparts during the CASEVAC task. Furthermore, because the C group's average time of 1.69 minutes is reasonably close to the HD group's mean time of 1.77 minutes, it can be assumed that male participation in both groups would have the same time when participating in a HD trial. It is also possible that the two trials in which a female carried the dummy was a large factor in the 4.75% reduction in performance.

For additional subjective comments relating to this task, see the Appendix.

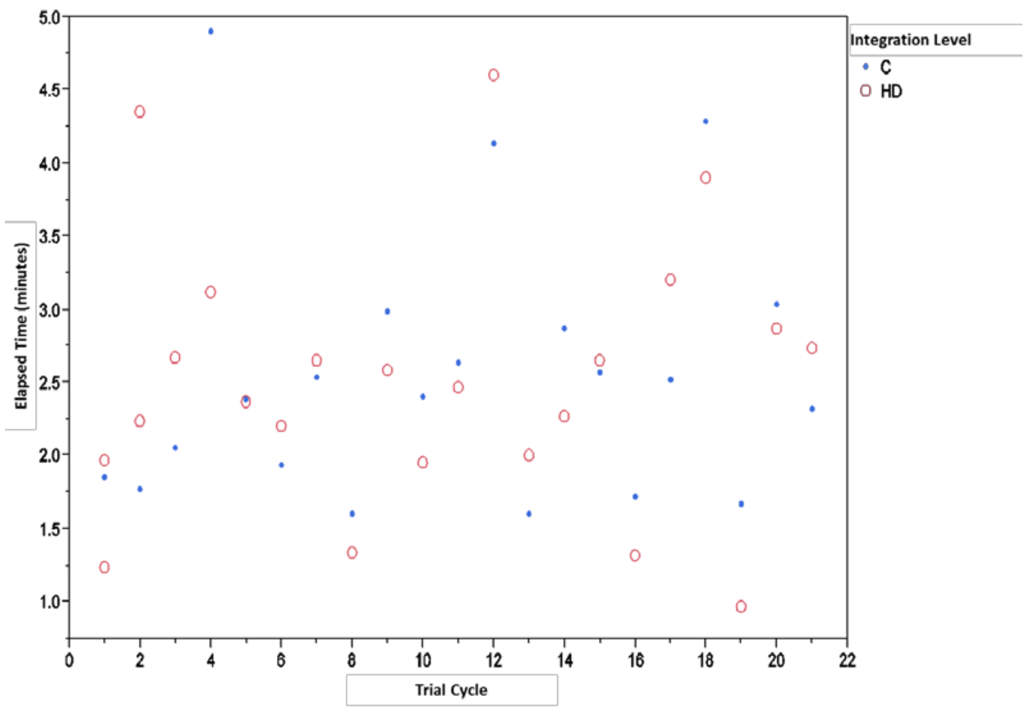
D.5.2.5 81-mm Mortar Emplacement & Displacement

D.5.2.5.1 81-mm Mortar Emplacement & Displacement Overview

This experimental task assessed the mortar squad's ability to move an 81-mm mortar system and 10 mortar rounds approximately 100m from an ORP to a MFP, get the mortar system fire capable (FIRECAP), and conduct a displacement back to the ORP. Although it was not assessed for time, the squad engaged two targets with five rounds each after the mortar system was FIRECAP. The recorded time started when the squad departed the ORP and paused when the first round was fired. The time resumed when the last (10th) round was fired and stopped when the entire squad was back at the ORP.

Figure D-5 displays all 0341 displace to LOA time data. All data on the scatter plot are valid for analysis.

Figure D-5. 81-mm Mortar Emplacement & Displacement



The data for the HD group are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.26 but not normally distributed for the C group, which had a p-value of <0.01. We proceed with presenting ANOVA results because they are confirmed by a Mann-Whitney Test with a p-value of 0.87.

The C group had a mean time of 2.56 minutes. This time is slower (but not statistically significant) than the HD mean time of 2.51 minutes. This difference results in a 2.10%, or 0.05-minute, improvement in movement time between the groups. See Table D-2 and Table D-3 for detailed analytical results.

D.5.2.5.2 81-mm Mortar Emplacement & Displacement Contextual Comments

A mortar squad must be able to move and deliver supporting fires as rapidly as possible on the battlefield. On average, the HD groups were 3 seconds faster than the C group. However, because this difference is so small and did not meet the statistical significance threshold, it can be assumed the C and HD groups performed the same.

D.5.2.5.3 81-mm Mortar Emplacement & Displacement Additional Insights

One source of masking occurred during this task that was not anticipated during the design. The initial movement portion of this task (a physical-based capacity) was masked by the time it took to assemble the mortar and sight in on the target (a skill-based capacity). For instance, it was observed that, when slower members of the squad fell back during the initial movement, their delay was masked (not captured) by the fact that the rest of the team began emplacing the 81-mm mortar system concurrently. By the time the weapon system was FIRECAP, all members had arrived at the MFP. Therefore, a squad that moved quickly versus a squad that got spread out was diffused by the time it took to sight in on the target. If the movement had been longer than 100m, then, based on the results of other similar movement tasks, one could conclude the results of this task would have been different.

D.5.2.5.4 81-mm Mortar Emplacement & Displacement Subjective Comments

For subjective comments relating to this task, see the Appendix.

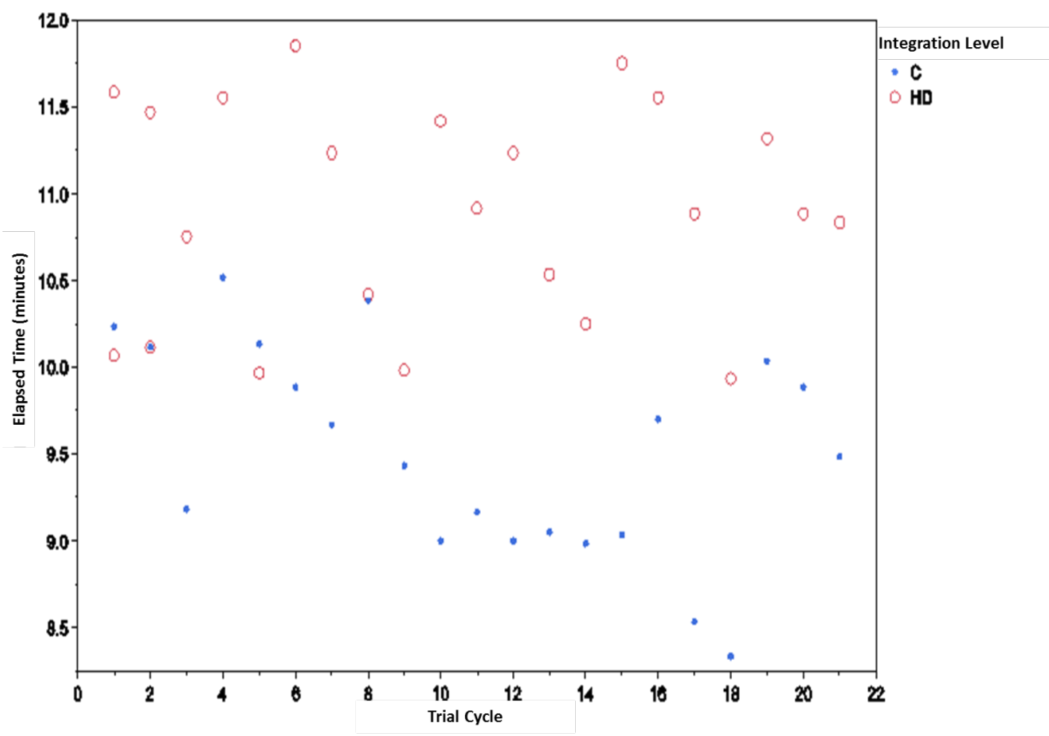
D.5.2.6 1-km Hike

D.5.2.6.1 1-km Hike Overview

This experimental task assessed a mortar team moving approximately 1 km to reinforce a notional friendly rifle squad pinned down by enemy fire. While conducting this movement, the team carried an assault load with 12 mortar rounds, individual weapons (M4), and two 60-mm mortar tubes. The time started when the squad departed the assembly area on Range107 and ended upon reaching a designated location in the down-range area.

Figure D-6 displays all 0341 1-km hike time data. All data on the scatter plot are valid for analysis.

Figure D-6. 1-km Hike



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.54 for the C group and 0.09 for the HD group.

The C group had a mean time of 9.51 minutes. This time is statistically significantly faster than the HD mean time of 10.89 minutes. This difference results in a14.49%, or 1.48-minute, degradation in time between the groups. See Table D-2 and Table D-3 for detailed analytical results.

D.5.2.6.2 1-km Hike Contextual Comments

The Infantry T&R Manual states, "the maximum assault load weight will be such that an average infantry Marine will be able to conduct combat operations indefinitely with minimal degradation in combat effectiveness." While moving 1 km with the assault load, the HD group took 1.38 minutes longer, a degradation of 14.49%. On any given day (under the same environmental conditions), the C group was faster than the HD group 90.5% of the time (19 of 21 trials).

D.5.2.6.3 1-km Hike Additional Insights

A purely objective evaluation of an 83-second delay is elusive but may possess some practical significance when considering a unit in contact. Consider an enemy 82-mm mortar system outside the range of friendly direct fire weapons (1800 m for .50 cal). In lieu of an 82-mm mortar firing table, an 81-mm mortar firing table reveals the maximum range on charge 1 is 1920 meters with a TOF of 20 seconds. Each enemy mortar tube would have the time to fire 3-4 mortar rounds prior to the integrated group being fire capable. The resultant trade in casualty exchange could be significant.

D.5.2.6.4 1-km Hike Subjective Comments

For subjective comments relating to this task, see the Appendix.

D.6 Statistical Modeling Results

D.6.1 Statistical Modeling Overview

The previous section discussed results only as they pertain to differences due to integration level alone. The goal of statistical modeling, as applied here, is to estimate, simultaneously, the effect of gender-integration levels and other relevant variables on squad performance. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For the same six selected tasks described in the previous section, this section presents an overview of the analysis and results and then presents the modeling results for each of the tasks.

For each task, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates an increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time. The results indicate where certain patch numbers are significant for a given variable.

The experiment tracked Marines within the squad by a patch number that associated their random position within the team/squad to a specific billet. Table D-4 displays the 0341 patch numbers and billet titles. The billet titles for the mortar team/squad depended on the day of the trial cycle (day 1 offensive tasks versus day 2 defensive tasks). The 60-mm mortar team had two mortar tubes for the engagement. Therefore, Patch 1 and Patch 3 were the gunners for this task.

Table D-4. 0341 Patch Numbers and Billet Titles

Patch Number	Billet Title Day 1	Billet Title Day 2
1	Gunner	Gunner
2	Assistant Gunner	Assistant Gunner
3	Gunner	Ammunition Man 1
4	Assistant Gunner	Ammunition Man 2

D.6.2 0341 Selected Tasks Method of Analysis

Due to the small number of trials, a mixed-effects model with all mortar team/squad members and all types of personnel data does not work for the 0341 data set. Thus, we model each personnel variable with integration level separately with a random effect for who filled each position within the mortar team/squad. For example, age for each member of the mortar team/squad (four variables), a random effect for who filled each billet, and integration level are modeled with the result (response time) as the response variable. Where maximum likelihood estimation converged, AIC was used for variable selection. Otherwise, we comment on the significance of individual variables in the full model. Variables reported as significant are concluded to be significant based on at least a one-sided test.

D.6.3 0341 Selected Tasks Overall Modeling Results

The displace to LOA task is the only task for which a personnel variable is significant for all members of the mortar team where height is statistically significant for all members of the mortar team and is negatively correlated with the response.

Integration level is significant for the 7-km hike, 60-mm mortar engagement time, and the 1-km hike. For each of these tasks, modeling the random effects for the individuals participating in the task results in changes from the initial results in the descriptive statistics. Each respective task paragraph describes these changes.

The CASEVAC and 81-mm mortar emplacement and displacement tasks did not have a final model with any significant variables. Refer to Section D.5.2.4 and Section D.5.2.5, respectively, for these tasks' ANOVA results.

D.6.3.1 7-km Hike

We model elapsed time for the 7-km hike as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- All.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 7-km hike time:

- CFT MTC of patch 4
- PFT 3-mile run of patch 4.

The following personnel variables are significant in their respective models and are negatively correlated with the 7-km hike time:

- Height of patches 1, 2, and 3
- Weight of patches 1, 2, and 3
- AFQT score of patches 2 and 4
- GT score of patch 2
- PFT crunches of patch 3
- Rifle score of patch 4.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, the final model includes integration level only where HD has a difference of 16.42 minutes when compared to a C group and a p-value of <0.01. This difference is a decrease from the 18.97-minute difference identified in the descriptive statistics, which is a 13.44% change.

D.6.3.2 60-mm Mortar Engagement

We model elapsed time for the 60-mm mortar engagement as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the team. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Age
- AFQT score
- GT score
- CFT MTC
- PFT crunches
- PFT 3-mile run
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 60-mm mortar engagement time:

- CFT MTC for patches 2 and 4.

The following personnel variables are significant in their respective models and are negatively correlated with the 60-mm mortar engagement time:

- Height of patches 1 and 3
- Weight of patches 1, 3, and 4.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patterns and the effects are often negligible, the final model includes integration level only where HD has a difference of 0.45 minutes when compared to a C group and a p-value of <0.01. This difference is a decrease from the 0.49-minute difference identified in the descriptive statistics, which is an 8.16% change

D.6.3.3 Displace to LOA

We model elapsed time for the displacement to LOA as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the team. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Age
- AFQT score
- GT score
- CFT MTC
- CFT MANUF
- PFT crunches
- PFT 3-mile run
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the displacement to LOA time:

- PFT 3-mile run of patch 3.

The following personnel variables are significant in their respective models and are negatively correlated with the displace to LOA time:

- Height of patches 1, 2, 3, and 4
- Weight of patches 2, 3, and 4
- GT score of patch 3
- PFT crunches of patch 3.

The following personnel variables have no significant variables in their respective models:

- None.

The final model includes the height of the mortar team. AIC does not select integration level in this model. Each of the variables is statistically significant and negatively correlated with the displacement to the LOA time. There are no strong correlations between the integration level and height for the mortar participants (less than 0.33). Table D-5 displays the variables and their respective estimates and standard errors. This result implies that, as a Marine's height increases, the displacement to the LOA time will decrease.

Table D-5. Coefficients and Standard Errors for the Final Displace to LOA Model

Effect	Estimate	Std. Error
Height patch 1	-0.04	0.02
Height patch 2	-0.05	0.02
Height patch 3	-0.04	0.02
Height patch 4	-0.04	0.01

D.6.3.4 CASEVAC

We model elapsed time for the CASEVAC as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the team. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the CASEVAC time:

- None.

The following personnel variables are significant in their respective models and are negatively correlated with the CASEVAC time:

- Age of patch 3
- Height of patch 1
- Weight of patch 1
- GT score of patches 3 and 4
- PFT crunches of patch 3.

The following personnel variables have no significant variables in their respective models:

- AFQT score
- CFT MTC
- CFT MANUF
- PFT 3-mile run
- Rifle score.

Because integration level is not significant in the final model and there are no variables that are significant for the whole team, there is no final mixed-effects model for this task. Refer to the Section D.5.2.4 to see the ANOVA results for this task.

D.6.3.5 81-mm Mortar Emplacement and Displacement

We model elapsed time for the 81-mm mortar emplacement and displacement as a function of each personnel variable and integration level in a separate mixed model.

The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 81-mm mortar emplacement and displacement time:

- Squad leader
- CFT MTC of patch 2
- CFT MANUF time of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the 81-mm mortar emplacement and displacement time:

- Height of patch 2
- AFQT score of patch 2
- GT score of patch 2
- PFT crunches of patch 3
- Rifle score of patch 2.

The following personnel variables have no significant variables in their respective models:

- Age
- Weight
- PFT 3-mile run.

Because integration level is not significant in the final model and there are no variables that are significant for the whole squad, there is no final mixed-effects model for this task. Refer to Section D.5.2.5 to see the ANOVA results for this task.

D.6.3.6 1-km Hike

We model elapsed time for the 1-km hike as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position on the team. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- All.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 1-km hike time:

- CFT MTC of patch 2
- PFT crunches of patch 1
- PFT 3-mile run of patch 4.

The following personnel variables are significant in their respective models and are negatively correlated with the 1-km hike time:

- Height of patches 1, 2, and 3
- Weight of patch 3
- PFT crunches of patch 3.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, the final model includes integration level only where HD has a difference of 1.19 minutes when compared to a C group and a p-value of <0.01. This difference is a decrease from the 1.48-minute difference identified in the descriptive statistics, which is a 19.59% change.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix to Annex D
0341 Supplemental Information

This appendix provides supplemental information for the 0341 portion of the GCEITF experiment. It provides information regarding the GCEITF leadership subjective comments and additional descriptive and basic inferential statistics not described in Annex D.

Section 1. GCEITF Leadership Subjective Comments

The GCEITF leadership provided comments on their observations of the experiment throughout it execution. Table D A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table D A – Summary of GCEITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement			Requesting extra breaks			Requires extra assistance			Needs no assistance			Compensating for another Marine			Gear pass off			Other			No category			Total
		C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	
7-km Hike	M	21	4	25	5	0	5	2	0	2	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	34
	F	0	41	41	0	13	13	0	0	0	0	0	0	0	0	0	0	6	6	0	0	0	0	1	1	61
	Unit	3	0	3	0	0	0	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	0	1	1	7
60-mm Mortar Engagement	M	4	3	7	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
	F	0	33	33	0	0	0	0	2	2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	36
	Unit	1	0	1	0	0	0	1	1	2	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	5
Displace to LOA	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
81-mm Mortar Emplacement & Displacement	M	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	F	0	2	2	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	6
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1-km Hike	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASEVAC to CCP	M	0	1	1	0	0	0	0	0	0	17	16	33	0	0	0	0	0	0	0	0	0	1	0	1	35
	F	0	9	9	0	2	2	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	13
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fire & Movement to MFP; Movement to MFP #1	M	2	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	F	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fire & Movement to MFP; First Three Rounds	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Fire & Movement to MFP; Movement to MFP #2	M	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	F	0	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	10
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fire & Movement to MFP; Second Three Rounds	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Section 2. Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for 11 additional 0341 tasks. Annex D contains the descriptive statistics for the remainder of the 0341 tasks. The words “metric” and “task” are used interchangeably throughout this Appendix; they both refer to the experimental task.

The two tables below display the results for 11 additional 0341 metrics. Table D B displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent differences between integration levels.

Table D C displays ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare the two groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the HD group is different from that in the C group. We present basic inferential statistics for four tasks.

Table D B – 0341 Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
7-km Hike; first km (minutes)*	C	20	10.66	0.59	7.40%
	HD	21	11.44	0.66	
7-km Hike; second km (minutes)*	C	20	11.05	0.62	10.94%
	HD	21	12.26	0.95	
7-km Hike; third km (minutes)*	C	20	11.20	1.61	32.43%
	HD	21	14.83	3.13	
7-km Hike; fourth km (minutes)*	C	20	12.52	2.33	18.54%
	HD	22	14.84	2.86	
7-km Hike; fifth km (minutes)*	C	20	13.88	2.75	32.56%
	HD	21	18.40	2.86	
7-km Hike; sixth km (minutes)*	C	20	14.14	1.29	21.44%
	HD	22	17.17	2.58	
7-km Hike; seventh km (minutes)*	C	20	15.28	2.11	19.51%
	HD	22	18.27	3.56	
60-mm Mortar Engagement; Movement to MFP #1 (seconds)*	C	65	49.66	10.76	35.92%
	HD	32	67.50	19.22	
60-mm Mortar Engagement; First Three Rounds (seconds)	C	19	36.91	5.09	-0.12%
	HD	20	36.86	4.83	
60-mm Mortar Engagement; Movement to MFP #2 (minutes)*	C	65	1.24	0.28	27.34%
	HD	32	1.58	0.39	
60-mm Mortar Engagement; Second Three	C	19	32.50	4.65	15.77%
	HD	22	37.62	5.80	

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
Rounds (seconds)*					

Table D C – 0341 ANOVA Results and Welch's T-Test Results

Metric	F statistic (df)	F Test P-value	Comparison	2-sided P-Value	1-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
60-mm Mortar Engagement; Movement to MFP #1 (seconds)*	34.38 (1, 95)	< 0.01*	HD-C	< 0.01†	< 0.01†	0.25†	0.42†	0.22†	0.43†
60-mm Mortar Engagement; First Three Rounds (seconds)	0.00 (1, 37)	0.98	HD-C	0.98	0.49	-2.12	2.03	-2.73	2.64
60-mm Mortar Engagement; Movement to MFP #2 (minutes)*	24.21 (1, 95)	< 0.01*	HD-C	< 0.01†	< 0.01†	0.28†	0.45†	0.25†	0.48†
60-mm Mortar Engagement; Second Three Rounds (seconds)*	9.54 (1, 39)	< 0.01*	HD-C	< 0.01*	< 0.01*	3.00	7.25	2.37	7.88

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality.

Additional Task Results:
7-km Hike by km. The general trend for the 7-km hike was that the difference between the HD and C groups increased over the course of the hike.

60-mm Mortar Engagement; Movement to MFP #1. The data for the HD group are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.03 but not normally distributed for the C group, which had a p-value of <0.01. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a two-sided p-value of <0.01.

The C group had a mean of 49.66 seconds. This time is statistically significantly faster than the HD mean of 67.50 seconds. The HD group was 35.92% slower than the C group.

- **Contextual Comments.** For references regarding the importance of speed, see Annex D. The C group squads were able to get in position to bring their weapon system into action 18 seconds faster than the HD squads, on average. This difference could be a significant advantage in a combat engagement.

60-mm Mortar Engagement; First Three Rounds. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.21 for the C group and 0.90 for the HD group.

The C group had a mean of 36.91 seconds. This time is slower (but not statistically significant) than the HD group mean of 36.86 seconds. The HD group was 0.12% faster than the C group.

- **Contextual Comments**. This event did not elicit a particular difference in the squad types. However, this is not a particularly physically demanding event by itself.

60-mm Mortar Engagement; Movement to MFP #2. The data for the HD group are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.55 but not normally distributed for the C group, which had a p-value of <0.01. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a two-sided p-value of <0.01.

The C group had a mean of 1.24 minutes. This time is statistically significantly faster than the HD group mean of 1.58 minutes. The HD group was 27.34% slower than the C group.

- **Contextual Comments**. The importance of speed has been iterated several times before. The C group squads were able to get in position to bring their weapon system into action 20 seconds faster than the HD squads, on average. When considered in combination with the 18-second advantage already gained in the movement to MFP #1 event, the advantage for a combat engagement is even more significant.

60-mm Mortar Engagement; Second Three Rounds. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.65 for the C group and 0.88 for the HD group.

The C group had a mean of 32.50 seconds. This time is statistically significantly faster than the HD group mean of 37.62 seconds. The HD group was 15.77% slower than the C group.

- **Contextual Comments.** This is essentially the same event as the “First Three Rounds” portion of the test already described. With that in mind, it is not clear why there is a significant difference between the squad types for this event when such a difference is absent in the first one. The HD groups were, on average, slightly slower than the first event (just under a second), while the C groups were over four seconds faster on average. The observer’s comments do not show any particular reason this difference occurred. Without amplifying information as to a cause for this difference, it can still be reiterated that faster is better.

THIS PAGE INTENTIONALLY LEFT BLANK

Annex E.

Infantry Assaultman (MOS 0351) and Infantry Anti-tank Missileman (MOS 0352)

This annex details the Infantry Assaultman (MOS 0351) and Infantry Anti-tank Missileman (MOS 0352) combined portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 2 March to 26 April 2015 at Range 107 and Range 110, aboard the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, CA. The combined 0351 and 0352 assessment is referred to in this annex as the 035X assessment. The sections outline the Infantry 035X scheme of maneuver (SOM), limitations, deviations, data set description, descriptive and basic inferential statistics, and modeling results.

E.1 Scheme of Maneuver

E.1.1 Experimental Cycle Overview

The Infantry Assaultman (0351) and Anti-tank Missileman (0352), or MOS 035X, assessment of the GCEITF took place in a field environment aboard MCAGCC, Twentynine Palms, CA. The assessment consisted of 21 trial cycles, each of which was a 2-day test cycle, conducted over the course of 55 days. After every 4 days of trials, the Marines spent 1 recovery day at Camp Wilson. The 0351 and 0352 MOSs were combined for the assessment to maximize the participation of every Marine and generate a larger sample. The experimental unit was composed of a pair of two-person Assault Teams making an assault squad or an anti-armor squad. Each squad consisted of four volunteers and one direct-assignment (non-volunteer) squad leader. Each member was trained to fill each billet within the team/squad: gunner and assistant gunner. The assessment was executed under the supervision of MCOTEA functional test managers and a range Officer in Charge (OIC)/Range Safety Officer (RSO) from the GCEITF.

E.1.2 Experimental Details

The 2-day 035X assessment replicated offensive and defensive tasks. The 035X MOS began each cycle on Day 1/Offensive tasks and concluded with Day 2/Defensive tasks. Two 035X squads executed each trial cycle: a control (C) nonintegrated squad and a high-density (HD) integrated squad with two, three, or four females.

Day 1 of the trial cycle was executed on Range 107 and consisted of supporting a squad attack as an assault squad employing the shoulder-launched, multipurpose assault weapon (SMAW). Each assault squad moved approximately 1 km to an assault position (AP) wearing an assault load and carrying a personal weapon (M-4), two SMAW launchers, and four rockets. The squad moved all personnel and gear over an 8-

ft obstacle/wall. After negotiating the obstacle, the squad staged its assault packs and followed-in-trace of a rifle squad conducting a live-fire assault. During the counterattack, the assault squad moved from a cold position to a hot position and engaged two enemy armored vehicles with the SMAW. Then, the assault squad conducted a 100-m casualty evacuation (CASEVAC) of a 220-lb dummy.

Day 2 of the trial cycle was executed on Range 110 and consisted of defensive actions. The squad started the day with a 7-km forced march from Range 107 to Range 110 wearing an approach load and carrying a personal weapon, two SMAW launchers, and four rockets. After arriving at Range 110, the squad then had to mount and dismount the M-41 Saber system to a High-Mobility Multipurpose Wheeled Vehicle (HMMWV). Each squad then carried four tube-launched, optically tracked, wire-guided (TOW) missiles to a firing line where two Saber systems were already assembled and calibrated. Finally, the two gunners engaged two designated targets as varying distances. At the conclusion of Day 2, the Marines were reorganized into new squads for the next experimental cycle.

E.1.3 Additional Context

Throughout the assessment, Marines bivouacked at Range 107, sleeping in two-man tents. During trial execution, Marines wore/carried prescribed loads for each task. Weighing packs each day prior to the 7-km forced march ensured consistency. After each trial day, the Marines operated under company leadership, performing minimal physically demanding tasks. The Marines who were not part of an assessed squad conducted the same experimental subtasks after the assessed squads to ensure equity between individuals participating in a trial cycle and those not chosen for that particular cycle. These tasks will be discussed in detail in the loading section that follows.

Fatigue surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. It allows outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) to be accounted for in the analysis of the performance data. Workload surveys collected the volunteers' perceived average and maximal levels of exertion during the performance specified tasks. Cohesion surveys provided a method of collecting subjective data relating to each squad's ability to work as a team and its overall perspective on the cohesiveness of the squad.

E.1.4 Experimental Tasks

E.1.4.1 1-km Movement

Assaulting an enemy position never starts from a static position; first, a movement must be conducted to the AP. The distance from the line of departure to the AP depends on myriad factors. Based on time and space constraints, this distance was set at just under

1 km for the experimental event. Each 035X squad moved this distance as quickly as possible while carrying an assault load in addition to two SMAW launchers and four rockets divided among the four members.

E.1.4.2 Negotiate an Obstacle

The conduct of an attack often involves reducing or negotiating an obstacle. It is common in an urban environment to make entry through a window or over a wall (obstacle). One of the more difficult tasks is climbing over a wall with a fighting load. Each squad negotiated an 8-ft wall by getting all four Marines and their equipment over as quickly as possible.

E.1.4.3 SMAW Engagement

An assault squad often accompanies a rifle squad into an attack. The assault squad provides anti-armor fires to destroy fortifications or other designated targets. When employing the SMAW, the squad generally prepares its rockets from a position of cover, commonly referred to as a cold position. Then it moves to an exposed position with a good line of sight to the enemy, commonly referred to as a hot position. During the experiment, the SMAW squad followed-in-trace of the rifle squad to a cold position where the gunners prepared a rocket for employment. The Marines then moved approximately 50 m to a hot position where they engaged an armored vehicle, conducted a rapid reload, engaged a second armored vehicle, and displaced back to the cold position. This task required Marines to move rapidly under a load and accurately engage targets while fatigued. This task determined the time and accuracy of employment.

E.1.4.4 CASEVAC

Casualties are an inevitable part of conducting combat operations. Units train to become proficient in triaging, handling, and transporting casualties. When a casualty is sustained, it is essential to move with a sense of urgency to get the injured Marine to the appropriate level of care. Each 035X squad moved a 220-lb dummy 100 m from a position of cover to a casualty collection point (CCP) while also transporting their assault packs and two SMAW launchers. The 035X squads could use a variety of techniques for transport but had to carry the dummy off the ground. This task determined the squad’s proficiency in moving a simulated casualty to a CCP. After the CASEVAC, each Marine took a fatigue and workload survey to assess overall fatigue and workload of the entire offensive task (see GCEITF Experimental Assessment Plan [EAP], Annex D).

E.1.4.5 7-km Hike

Infantry units must move through all sorts of terrain on foot. Units train by conducting a forced march with an approach load at a sustained rate of march. For the assessment, each 035X squad moved a distance of 7.20 km as quickly as possible while carrying an approach load, two SMAW launchers, and four rockets spread-loaded across all four members. This route was flat (minimal elevation change) and featured an unimproved surface with varying degrees of conditions (compact dirt and loose sand). The squads moved as fast as the slowest person. This task determined the squad's rate of movement over a 7.20-km route carrying the approach load and the crew-served weapon. Each Marine took a fatigue and workload survey after completion of the 7.20-km hike (see GCEITF EAP, Annex D).

E.1.4.6 Mount/Dismount Drill

The TOW missile is generally employed from a mounted platform, such as a HMMWV. Members of an anti-armor squad must mount and dismount the Saber system to a HMMWV. During the assessment, the anti-armor squad mounted a Saber system onto a HMMWV. A squad leader verified that Marines had properly mounted the Saber system, and then the squad dismounted the system. Marines did not conduct the calibration process due to its technical and timely procedures. This task was selected for the physical strength required to lift, curl, and press the Saber components up and lower them down by members of the squad. This task determined the time it took to mount and dismount the Saber system.

E.1.4.7 TOW Engagement

The TOW missile is a heavy anti-armor weapon that may be employed from a tripod to provide defensive fires. During the assessment, a precalibrated Saber system was mounted on a tripod at a designated firing position. Each anti-armor squad moved four TOW missiles to the firing line and engaged two designated targets, firing each missile in sequence. This task was chosen for the strength required to move each missile and the ability to employ the TOW while fatigued. This task determined the time to move the TOW missile to the firing line and gunner accuracy. At the conclusion of the TOW engagement, each Marine took a fatigue and workload survey to assess his or her fatigue and workload during the execution of the mount/dismount drill and engagement task (see GCEITF EAP, Annex D).

At the completion of the 2-day cycle, Marines took a cohesion survey to record their cohesion during the execution of the 2-day trial cycle (see GCEITF EAP, Annex M).

E.1.5 Loading Plan

The loading plan ensured, to the greatest extent possible, equity of physical activity among all volunteers throughout the duration of the experimental assessment. Every trial and task was conducted in the same manner and sequence to ensure consistency. Due to the number of volunteers, a handful of Marines were not part of an assessed squad for each 2-day cycle. Collaboration with company leadership determined that the best method of loading nonassessed Marines was to form them into a separate squad and have them perform the same tasks as an assessed squad to experience the same conditions and physical strain. Minor modifications were permitted due to the reduced size of the squad.

E.1.6 Scheme of Maneuver Summary

The 035X experiment consisted of a 2-day trial cycle composed of an offensive and defensive day. The offensive day involved four subtasks based on supporting a squad attack: 1-km movement, negotiating an obstacle, SMAW engagement, and CASEVAC. The defensive day involved three subtasks: a 7.20-km forced march, mount/dismount drill, and TOW engagement. During the course of the experiment, the 0351 and 0352 squads executed 2 pilot trial cycles and 21 record trial cycles. During trial execution, Marines rotated through every billet within the assault/anti-armor squad, carrying components of the crew-served load.

E.2 Limitations

E.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a manner similar to those in an operational environment. Under certain situations, artificial limitations or interruptions were introduced that changed or altered the way a task would normally be performed. While these limitations represent a degree of artificiality, they do not detract significantly from our abilities to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for 035X assessment.

E.2.2 Relative Difficulty of Record Test

The 035X GCEITF assessment was designed to gather data associated with some of the most physically demanding tasks within the 035X MOSs. These tasks in isolation do not fully replicate life experienced by a Marine during a typical field exercise (FEX) or a combat environment and did not obtain the maximum cumulative load that could be placed on an Infantry Marine. With limited time available, only selected 035X tasks were assessed. Other tasks/duties outside the assessment were minimized due to specific experimental constraints and human factors. During a typical FEX, it is common for Marines to conduct 24-hour operations that include daytime and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. The offensive day SOM took squads approximately 1 hour to complete, and the defensive day SOM took approximately 3 hours to complete. Outside the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

Another primary concern in designing the 035X assessment was to ensure that it was achievable and sustainable for a 60-day period. The 7-km forced march distance was selected based on the amount of training time available prior to the assessment. However, many of the loads carried were decreased. For instance, the crew-served load was altered from the Javelin Missile Simulator Rounds and Command Launch Unit for the Day 2 (defensive SOM) to the SMAW and dummy rockets. Additionally, the Marines were authorized 1 day off after every 4 days of training. This artificial recovery period is not achievable when conducting training or combat operations.

A final factor affecting the relative difficulty of the record test pertains to the intangible physiological impact of the volunteers' ability to drop on request (DOR) at any point during a trial. Any time a volunteer dropped during a trial cycle, that squad/team performed the following subtasks with fewer personnel. This factor could have affected the cohesion of each squad and could have influenced its performance.

E.2.3 Artificial Emplacement

Artificialities were introduced for a 60-day period. On Day 2 of the 035X experiment, the assault/anti-armor Marines engaged targets with the TOW system. Once the 7-km forced march was completed with the SMAW crew-served weapon system, the Marines arrived at a range with a TOW system already emplaced in a static firing position. The Marines engaged two targets per weapon system for each weapon system. Once Marines fired the four missiles, the trial was complete, and the TOW system remained on the firing line.

E.2.4 Number of Volunteer Participants

For the 035X experiment, 6 male and 6 female volunteers began the experiment, but only 5 males and 4 females completed the assessment. The results presented in this annex are based on the performance of 9 to 12 Marines.

E.2.5 Limitations Summary

The 035X assessment was designed to replicate realistic training in a field environment. The end state was to create an experiment in which the volunteers felt they were conducting realistic and operationally relevant tasks. Certain unavoidable limitations to the assessed tasks and non-assessed operating environment introduced a level of artificiality that would not normally have been present in a field training or combat environment.

E.3 Deviations

E.3.1 Integration Level of the Squad

The EAP stated that the 035X MOS would have three levels of integration: C, low-density (LD) (one female), and HD (two females). Prior to the beginning of the first record trial cycle, the population could no longer support an LD group. Initially, there were two females in the HD squad. Due to the number of DORs, however, the number of females in the squad changed based on the availability of male and female Marines.

For the majority of trial cycles, there was an all-male squad and an all-female squad. However, due to injuries, there were some occasions when an HD squad consisted of three female Marines.

E.4 Data Set Description

E.4.1 Data Set Overview

The 035X portion of the experiment consisted of 2 pilot trial cycles and 21 record trial cycles. The pilot trial cycles were conducted from 2 to 6 March 2015. Pilot trial cycle data are not used in analysis due to variations in the conduct of the experiment. We based all analysis on the 21 record trial cycles executed from 7 March to 26 April 2015.

E.4.2 Record Test Volunteer Participants

At the beginning of the first record trial cycle, there were six male 035X volunteers and six female volunteers. Several Marines voluntarily withdrew, or were involuntarily withdrawn, during the execution of the experiment. The final number of volunteers was five males and four females.

E.4.3 Planned, Executed, and Analyzed Trial Cycles

Table E-1 displays number of trial cycles planned, executed, and analyzed by task. The planned number of trial cycles for the 035X MOS per Section 7.5.3 of GCEITF EAP is 120 trial cycles, or 40 trial cycles per planned integration level (C, LD, and HD). The original plan called for six squads per day (two per integration level) over the 20 trial cycles. However, due to the number of Marines who voluntarily withdrew or were involuntarily withdrawn from the experiment prior to the execution of the first record trial cycle, only one squad of the C and HD integration levels remained. We chose to keep the HD integration level to employ more Marines in the experiment.

While the number of male and female volunteers did not support the ability to execute another C group during a trial cycle, there were times when another HD group could execute additional whole or partial trials. A partial trial was a trial in which a two-Marine unit could execute the tasks where data were collected at the two-Marine unit level. The planned number of trials in Table E-1 reflects 21 planned trial cycles for each integration level and accounts for the extra trials the HD integration level completed.

There are several occurrences of missing data for the 7-km hike by individual kilometer. The individual kilometer times were derived from GPS data. Early in the experiment, the Garmin GPSs were set to record a volunteer’s position every second. Due to the storage space on the GPS and length of the trial, when volunteers executed the 7-km hike and then their follow-on tasks, the GPS could not hold all the data. Therefore, it overwrote the hike data. Once the problem was found, the GPSs were corrected to record location every 2 seconds.

Table E-1. 035X Planned, Executed, and Analyzed Trial Cycles[1]

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
7-km Hike	C	21	21	21	
	HD	21	22	21	Run 1 on Mar 8 disregarded due to TIR
SMAW Engagement by Time	C	21	21	20	Remove Apr 6 due to TIR
	HD	21	24	24	
SMAW Engagement Accuracy	C	21	21	20	Remove Apr 6 due to TIR
	HD	21	24	24	
Negotiate Obstacle	C	21	21	21	
	HD	21	22	22	
CASEVAC	C	21	21	21	
	HD	21	22	20	Remove Mar 7 and Mar 17 due to TIR
Mount/Dismount Saber	C	21	21	19	Remove Apr 2 and Apr 12 due to TIR
	HD	21	22	20	Remove Apr 2 and Apr 12 due to TIR
TOW Engagement	C	21	21	19	Remove Mar 18 and Apr 4 due to hangfire
	HD	21	24	23	Remove Mar 20 due to hangfire
1-km Hike	C	21	21	21	
	HD	21	22	22	
7-km Hike; 1-km Time	C	21	21	19	No data: Mar 8, Mar 18.
	HD	21	22	16	No data: run1 Mar 8, run 2 Mar 8, Mar 10, Mar 13, Mar 15, Mar 18.
7-km Hike; 2-km Time	C	21	21	19	No data: Mar 8, Mar 18.
	HD	21	22	16	No data: run1 Mar 8, run 2 Mar 8, Mar 10, Mar 13, Mar 15, Mar 18.
7-km Hike; 3-km Time	C	21	21	16	No data: Mar 8, Mar 18. High outlier (GPS error): Mar 30, Apr 7, Apr 9
	HD	21	22	18	No data: run1 Mar 8, run 2 Mar 8, Mar 15, Mar 18.
7-km Hike; 4-km Time	C	21	21	16	No data: Mar 8, Mar 18. Negative result: Mar 30, Apr 7. Low outlier Apr 9.
	HD	21	22	19	No data: run1 Mar 8, run 2 Mar 8, Mar 18.
7-km Hike; 5-km Time	C	21	21	19	No data: Mar 18. High outlier (GPS error): Mar 30.
	HD	21	22	19	No data: run 2 Mar 8, Mar 18. Remove Mar 8 run 1 due to TIR.
7-km Hike; 6-km Time	C	21	21	17	No data: Mar 18. Low outlier (GPS error) Mar 30, Apr 24 and 26.
	HD	21	22	20	No data: Mar 18. Remove Mar 8 run 1 due to TIR.
7-km Hike; 7-km Time	C	21	21	19	No data: Mar 18. Low outlier (GPS error) Mar 30.
	HD	21	22	20	No data: Mar 18. Remove Mar 8 run 1 due to TIR.
Mount Saber	C	21	21	19	Remove Apr 2 and Apr 12 due to TIR

[1] A TIR in this table refers to a Test Incident Report, which is a report the test team or direct assignment leaders completed when an incident occurred that affected the natural execution of a trial. If a data point is removed due to a TIR, it is because the TIR affected the data in such a way that it is not comparable to the rest of the data set.

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
onto Vehicle	HD	21	22	20	Remove Apr 2 and Apr 12 due to TIR
Dismount Saber from Vehicle	C	21	21	20	Remove Apr 2 due to TIR
	HD	21	22	21	Remove Apr 2 due to TIR
TOW Engagement Accuracy	C	21	21	21	
	HD	21	24	24	

E.5 Descriptive and Basic Inferential Statistics

E.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common infantry tasks and are indicative of unit-level proficiency during either field exercises or combat operations. This section presents results for 7 out of 18 tasks. The Appendix to this Annex contains the descriptive statistics for the remainder of the 035X tasks. The words “metric” and “task” are used interchangeably throughout this annex. They both refer to the experimental task.

Each squad consisted of four volunteer Marines and a direct assignment squad leader. There were two integration levels for all tasks: a C group was not gender integrated, and an HD group was gender integrated with two, three, or four female Marines.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA) (or nonparametric tests as necessary), and scatter plots. The subsequent sections will cover each task in detail. Lastly, contextual comments, additional insights, and subjective comments (as applicable) tying back to each experimental task are incorporated.

Use caution when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading because of differing factors between MOS tasks, such as distances, techniques, leadership, load carried, group size, and group composition.

E.5.2 035X Selected Task Descriptive Statistic Results

The two tables that follow display the results for the seven selected 035X metrics. Table E-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations (SDs). Table E-3 displays ANOVA results, including metrics and integration levels, p-values suggesting statistical significance, integration level elapsed-time differences, and percentage differences between integration levels. For each task, an ANOVA and t-test were conducted to compare the two groups. If nonparametric tests were needed, Table E-3 displays these results instead of ANOVA and t-test results. If p-values are less than the a priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the HD group is different from that in the C group.

Table E-2. 035X Selected Task Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD
7-km Hike (minutes)*	C	21	97.18	5.99
	HD	21	123.57	12.68
SMAW Engagement by Time (minutes)*	C	20	1.30	0.16
	HD	24	2.06	0.52
SMAW Engagement Accuracy (% hits)*	C	20	0.79	0.19
	HD	24	0.49	0.25
Negotiate Obstacle (minutes)*	C	21	1.53	0.22
	HD	22	3.28	1.24
Negotiate Obstacle **[excluding potential influential point]** (minutes)*	C	21	1.53	0.22
	HD	21	3.04	0.53
CASEVAC (minutes)*	C	21	0.91	0.21
	HD	20	2.35	0.90
Mount/Dismount Saber (minutes)*	C	19	4.26	0.40
	HD	20	6.80	0.97
Engage Target (TOW Engagement (minutes)*	C	19	3.76	1.78
	HD	23	5.58	2.01

*Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a nonparametric equivalent test.

Table E-3. 035X Selected Task ANOVA and Welch's T-test Results

Metric	F-Statistic (df)	F-Test P-Value	Com-parison	Differ-ence	% Differ-ence	2-sided P-Value	1-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
7-km Hike (minutes)††	72.09	<0.01*	HD-C	26.39	27.15%	< 0.01*	< 0.01*	22.37	30.40	21.18	31.59
SMAW Engagement by Time (minutes)††	57.76	< 0.01*	HD-C	0.76	57.99%	< 0.01*	< 0.01*	0.61	0.91	0.56	0.95
SMAW Engagement Accuracy (% hits)*	19.42 (1, 42)	< 0.01*	HD-C	0.30	-37.83%	< 0.01†	< 0.01†	-0.50†	-0.25†	-0.50†	-0.25†
Negotiate Obstacle (minutes)*	134.86‡	< 0.01‡	HD-C	1.75	114.75%	< 0.01†	< 0.01†	1.37†	1.68†	1.32†	1.73†
Negotiate Obstacle **[excluding potential influential point]** (minutes)††	145.18	< 0.01*	HD-C	1.51	99.12%	< 0.01*	< 0.01*	1.35	1.68	1.30	1.73
CASEVAC (minutes)††	52.69	< 0.01*	HD-C	1.44	159.25%	< 0.01*	< 0.01*	1.17	1.72	1.09	1.80
Mount/Dismount Saber (minutes) ††	128.25	< 0.01*	HD-C	2.54	59.82%	< 0.01*	< 0.01*	2.23	2.86	2.14	2.95
TOW Engagement (minutes)*	9.40 (1, 40)	< 0.01*	HD-C	1.82	48.36%	< 0.01†	< 0.01†	1.23†	2.10†	1.18†	2.15†

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a nonparametric equivalent test.
†Results presented are from a Mann-Whitney nonparametric test due to non-normality.
††Indicates that results presented are from Robust ANOVA and Welch's t-tests with p-values compared to 0.05 for Bonferroni adjustment due to unequal variances. The reported F-statistic is a Chi-square statistic from Robust ANOVA, and the F-test p-value is the Robust ANOVA p-value. The p-values in columns labeled "2-sided P-value" and "1-sided P-value" are p-values from Welch's t-tests, and the confidence intervals are from Welch's t-tests.
‡Indicates that results presented are from Robust ANOVA due to unequal variances. The reported F-statistic is a Chi-square statistic from Robust ANOVA, and the F-test p-value is the Robust ANOVA p-value.

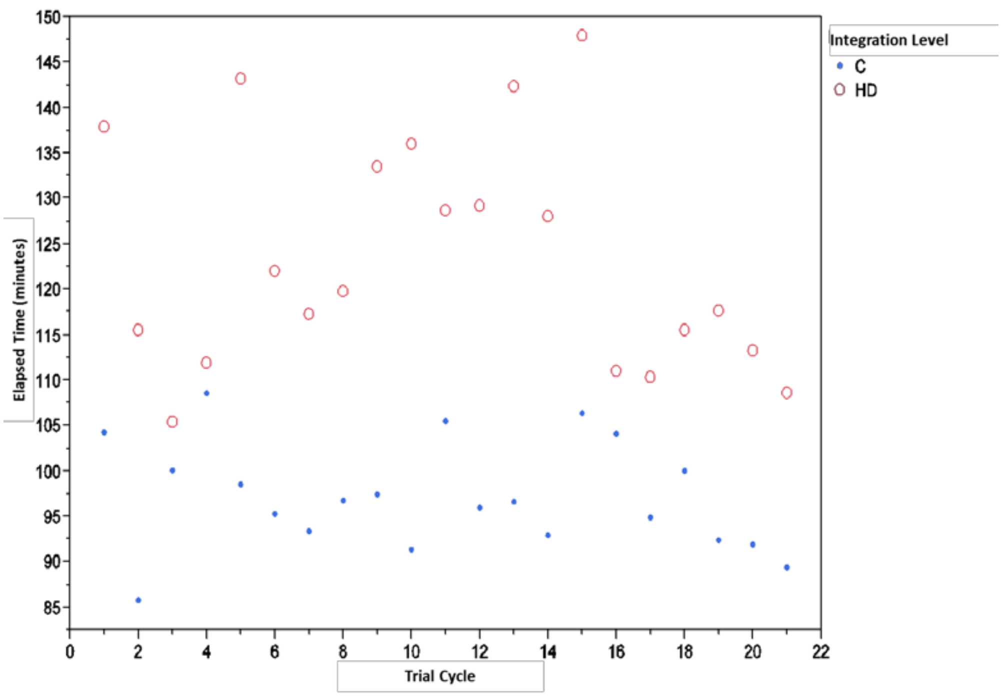
E.5.2.1 7-km Hike

E.5.2.1.1 7-km Hike Overview

This experimental task assessed a squad of four Marines moving 7.20 km while each Marine carried an approach load, an individual weapon (M-4), and a portion of the crew-served weapon load. The crew-served load consisted of two SMAW launchers and four SMAW rockets, resulting in a cumulative load of 116-132 lb per Marine. The recorded time for this task started when the squad departed the Range 107 start point and stopped when the squad arrived at the Range 110 stop point. The squads moved as fast as the slowest person and could take as many breaks as necessary.

Figure E-1 displays all 035X 7-km hike time data. All data on the scatter plot are valid for analysis.

Figure E-1. 7-km Hike



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.78 for the C group and 0.19 for the HD group.

The C group had a mean time of 97.18 minutes. This time is statistically significantly faster than the HD mean time of 123.57 minutes. This difference results in a 27.15%, or 26.39-minute, degradation in hike time between the groups. The HD group had greater variability, as shown by the 6.69-minute increase in SD (5.99 minutes for the C group and 12.68 minutes for the HD group). See Table E-2 and Table E-3 for detailed analytical results.

E.5.2.1.2 7-km Hike Contextual Comments

E.5.2.1.2.1 USMC Hike Standards

The 7-km hike is a task that the following MOSs completed during the experiment: 0311, 0331, 0341, 0351, and 0352, Provisional Infantry, Provisional Machine Gunners, and Combat Engineers. There are varying standards to which we can compare this result. The following sections define those standards, as well as the one we chose as a comparison.

The Infantry T&R Manual (30 August 2013) identifies minimum standards that Marines must be able to perform in combat, including standards for tactical marches. In Chapter 8 of the Infantry T&R Manual in the MOS 0300 Individual Events section, task “0300-COND-1001: March under an approach load” is applicable to all 03XX MOSs, ranks PVT – LtCol. The condition and standard established by this task is: “Given an assignment as a member of a squad, individual weapon, and an approach load, complete a 20 kilometer march in under 5 hours.” The march pace required by this standard is 4 kilometers per hour (km/h). In Chapter 9 of the Infantry T&R Manual in the MOS 0302/0369 Individual Events section, tasks “0302-OPS-2001: Lead an approach march” and “0369-OPS-2501: Lead an approach march” are applicable to MOS 0302 and 0369, ranks SSgt – MGySgt and 2ndLt – LtCol. The condition and standard established by these tasks is: “Given a mission, time constraints, an approach march load, organic weapons, and a route, move 24.8 miles (40 km) in a time limit of 8 hours.” The march pace required by this standard is 5 km/h. Appendix E of the Infantry T&R Manual (Load Terms and Definitions) states: “The approach march load will be such that the average infantry Marine will be able to conduct a 20 mile hike in 8-hours with the reasonable expectation of maintaining 90% combat effectiveness.” The march pace required by this definition is 4.02 km/h.

Chapter 3 of Fleet Marine Force Reference Publication (FMFRP) 3-02A, Marine Physical Readiness Training for Combat (16 June 2004) states: “The normal pace is 30 inches. A pace of 30 inches and a cadence of 106 steps per minute result in a speed of 4.8 km/h or 3 mi/h and a rate of 4 km/h or 2.5 mi/h if a 10-minute rest halt per hour is taken.” Common Infantry practice today is to hike for 50 minutes and take a 10-minute break, while maintaining an overall pace of 4 km/h (resulting in a hiking pace of 4.8 km/h).

Driven by the need to pick an evaluative reference standard, this report follows the T&R Manual's intent to establish minimum standards and uses the 4-km/h march pace for a 20-km march established by task 0300-COND-1001 and supported by the definition of an approach load. Further, while an established reference standard is required to anchor observed performance to the T&R program, more important information is provided by performance differences observed between units that are gender integrated and those that are not.

E.5.2.1.2.2 035X Hike Pace

This result is relevant to both the training and combat environment, as it will take integrated squads more time to conduct foot movements. Per the tactical march standards noted earlier, the Marine Corps standard of hiking is 4.0 km/h. The HD group failed to meet this standard. The C group average pace was 4.45 km/h, while the HD group average pace was 3.50 km/h, finishing 26.4 minutes behind the C group. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), it would take the C group 4 hours and 30 minutes and the HD group 5 hours and 43 minutes to complete the hike, finishing 73 minutes behind the C group. With the exception of one instance, the slowest C group was faster than the fastest HD group throughout the duration of the experiment. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the HD group 100% of the time (21 of 21 trial cycles). Based on the SDs, the variation in performance of the HD group is greater than twice as much as the variation in performance of the C group. This inconsistency in the performance of the HD group leads to greater uncertainty and less confidence in their future performance from the mean.

E.5.2.1.3 7-km Hike Additional Insights

A variety of factors potentially explains the high degree of variability within the HD group: the weather, Day 2 versus Day 4 hike execution, or most notably the crew-served load that varied from 116 to 132 lb.

Based on the USMC standard of a 4-km/h pace over a 7.20-km route (which would result in a 108-minute hike completion time over the 7.20 km), the HD group was 15.57 minutes slower than that standard. In a battlefield situation, in which speed is essential, this delay is advantageous for the enemy. An enemy maneuvering at 4 km/h would have the time to move 1.04 km, shift indirect fires from preplanned targets, commit the employment of their least engaged unit, or conduct a spoiling attack.

Marine Corps Doctrinal Publications (MCDPs) consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to "Being Faster" and states, "Physical speed, moving more miles per hour, is a powerful weapon in itself." MCDP-6 Command and Control also speaks to speed relative to the enemy and states, "The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results." Further insights may be gleaned from the Appendix, which shows the difference in speed by kilometer. In general, the difference in performance increased as the movement got longer.

E.5.2.1.4 7-km Hike Subjective Comments

For subjective comments relating to this task, see the Appendix.

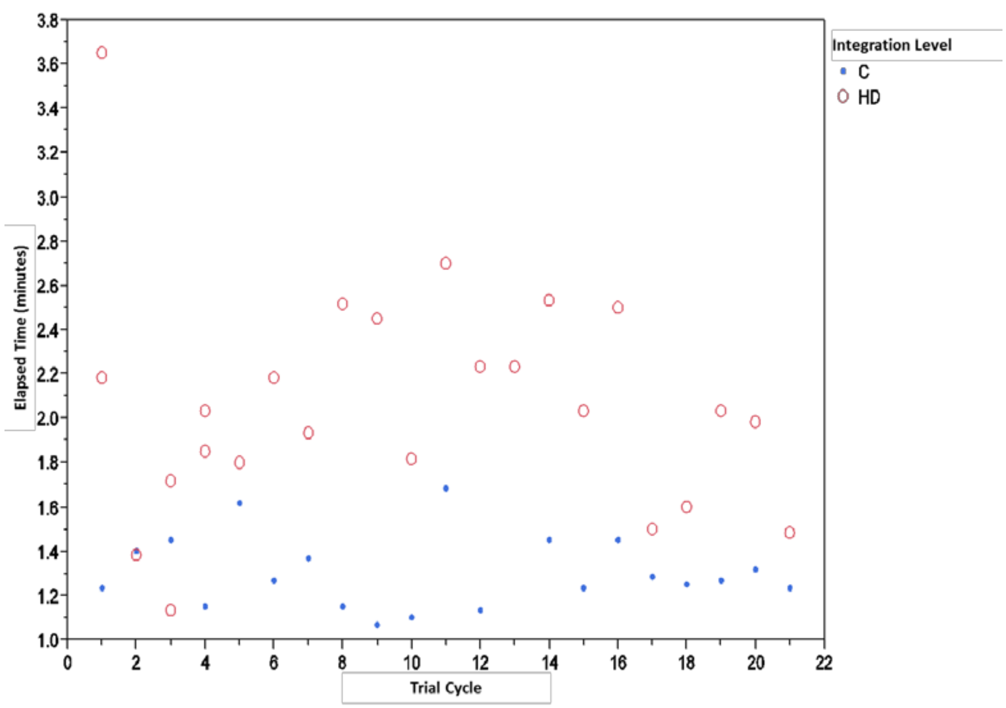
E.5.2.2 SMAW Engagement by Time

E.5.2.2.1 SMAW Engagement by Time Overview

This experimental task assessed the time for an assault squad (composed of a pair of two-Marine SMAW teams) to move approximately 50 meters from a cold position to a hot position and engage two targets with four rockets. One rocket reload per team was conducted between the first and second shots. The recorded time for this task started when the squad departed the cold position and stopped when the last (fourth) rocket was fired. Two SMAW teams conducted this engagement in sequence.

Figure E-2 displays all 035X SMAW engagement time data. All data on the scatter plot are valid for analysis.

Figure E-2. SMAW Engagement by Time



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.24 for the C group and 0.17 for the HD group.

The C group had a mean time of 1.30 minutes. This time is statistically significantly faster than the HD mean time of 2.06 minutes. This difference results in a 57.99%, or 0.76-minute, degradation in engagement time between the groups. The HD group had greater variability, as shown by the 0.36-minute increase in SD (0.16 minute for the C group and 0.52 minute for the HD group). See Table E-2 and Table E-3 for detailed analytical results.

E.5.2.2.2 SMAW Engagement by Time Contextual Comments

The ability to maneuver the SMAW rockets into an advantageous firing position is critically important to the forward progress and tempo of an assaulting unit. The faster this weapon system is brought to bear against an enemy target, the less time for friendly elements to be exposed to a threat. MCDP 1 states, "Speed is rapidity of action. It applies to both time and space. Both forms are genuine sources of combat power. In other words, speed is a weapon." The experimental results show that an HD group takes on average 46 seconds longer, a 58% degradation in performance, to accomplish the same task. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the HD group 95.2% of the time (20 of 21 trial cycles). Based on the SDs, the variation in performance of the HD group is greater than three times the variation in performance of the C group. This inconsistency in the performance of the HD group leads to greater uncertainty and less confidence in its future performance from the mean.

E.5.2.2.3 SMAW Engagement by Time Additional Insights

A purely objective evaluation of 46 seconds is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated squad. Considering the 650-rounds-per-minute sustained rate of fire from an RPK-74 machine-gun bunker, a single enemy machine-gun team would have the time/opportunity fire 498 rounds against an integrated squad still exposed to enemy fire. The resultant trade in casualty exchange could be significant. The difference in this movement and engagement time is even more noteworthy when one considers the accuracy for each SMAW engagement. See the results for the SMAW accuracy, Section E.5.2.3, to compare accuracy versus engagement time. On average, the HD groups were 38% less accurate than the C group.

E.5.2.2.4 SMAW Engagement by Time Subjective Comments

For subjective comments relating to this task, see the Appendix.

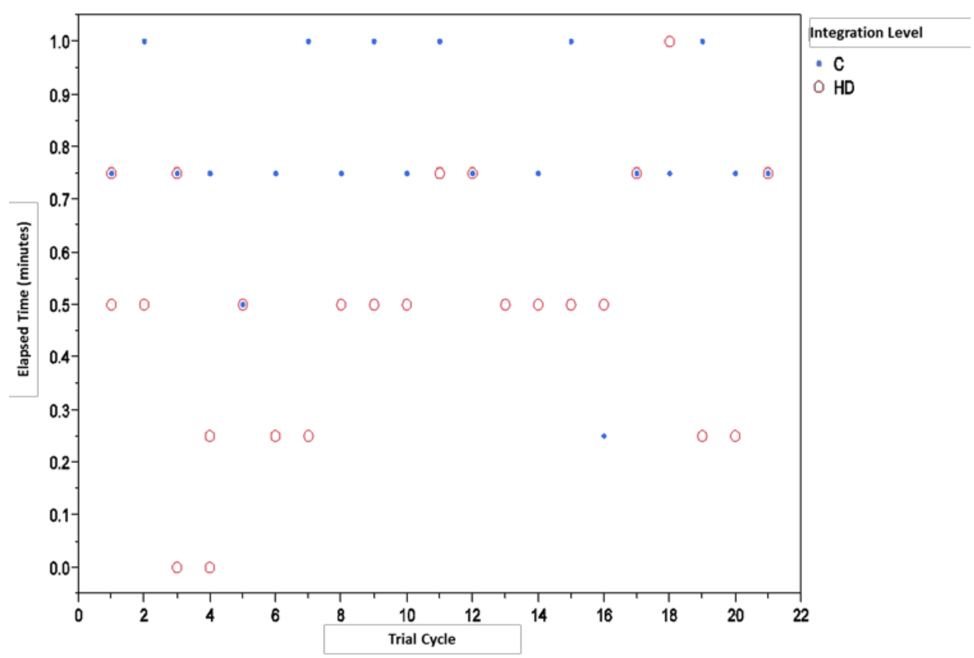
E.5.2.3 SMAW Engagement Accuracy

E.5.2.3.1 SMAW Engagement Accuracy Overview

This experimental task assessed the accuracy of an assault squad engaging two targets with four rockets. Two two-Marine teams, each firing two SMAW rockets, conducted the engagement. The first target was at a distance of 150 m and the second target was at a distance of 250 m. This task was conducted after walking 1 km, bounding 500 m, and sprinting 50 m to arrive at an advantageous firing position. Accuracy was assessed as a hit or miss on a vehicle-sized target.

Figure E-3 displays all 035X SMAW engagement accuracy data. All data on the scatter plot are valid for analysis.

Figure E-3. SMAW Engagement Accuracy



The HD group data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.04, but data are not normally distributed for the C group, which had a p-value of <0.01. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a p-value of <0.01.

The C group had a mean proportion of hit of 0.79. This proportion is statistically significantly higher than the HD proportion of 0.49. This difference results in a 37.88%, or 0.30-percentage-point, degradation in accuracy between the groups. The HD group had greater variability, as shown by the 0.06-percentage point increase in SD (0.19 minute for the C group and 0.25 minute for the HD group). See Table E-2 and Table E-3 for detailed analytical results.

E.5.2.3.2 SMAW Engagement Accuracy Contextual Comments

Accuracy is the measure of military effectiveness with a given weapon system under operational conditions. In combat operations, accuracy is highly desirable in destroying an enemy target. Ammunition, especially rockets, must be employed with precision since they are limited on the battlefield. To extrapolate the probability of hit to a battlefield situation in which there are 10 known targets, a C group would need to carry/fire 13 rockets, while an HD group would need to carry/fire 21 rockets to have the same military effect. Furthermore, on any given day (under the same environmental conditions), the C group was more accurate than the HD group 90% of the time (18 of 20 trial cycles).

E.5.2.3.3 SMAW Engagement Accuracy Additional Insights

This accuracy data are even more noteworthy when one considers the amount of time it took for each SMAW engagement. See the results for the SMAW engagement times, Section E.5.2.2, to compare accuracy versus engagement time. On average, the HD groups took 46 seconds longer to conduct each trial.

E.5.2.3.4 SMAW Engagement Subjective Comments

For subjective comments relating to this task, see the Appendix.

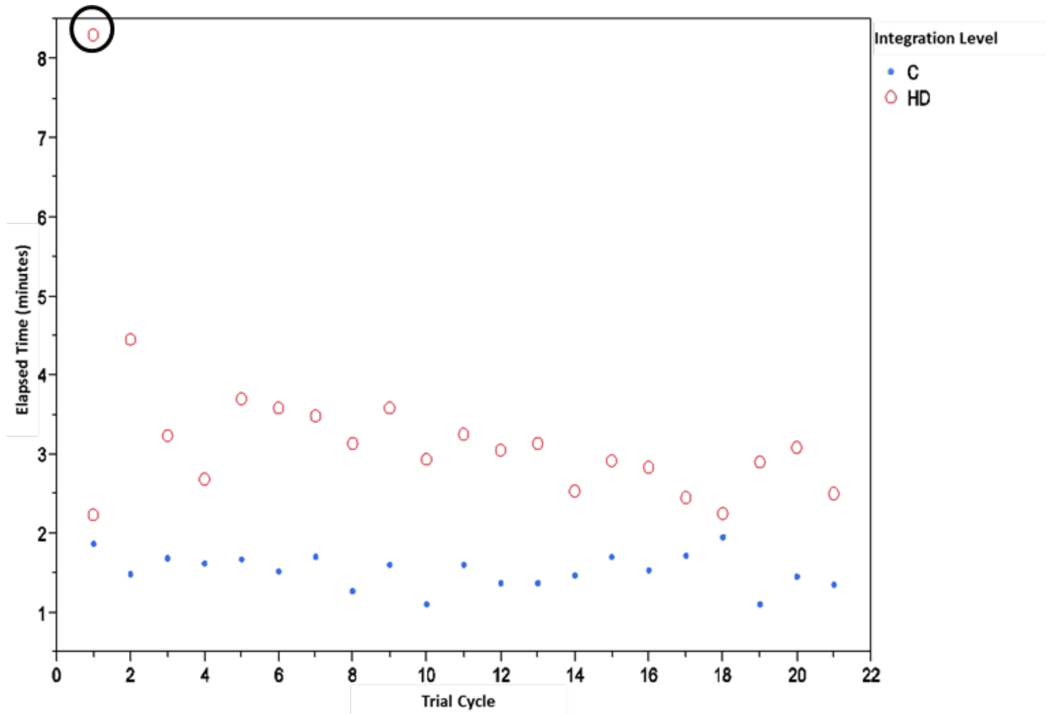
E.5.2.4 Negotiate Obstacle

E.5.2.4.1 Negotiate Obstacle Overview

This experimental task involved a squad of four Marines getting all Marines and equipment over an 8-foot ISO Container (wall obstacle) as quickly as possible. The squad began this task as soon as it finished the 1-km movement. The time started when the first Marine touched the obstacle and stopped when the last Marine touched both feet down on the other side.

Figure E-4 displays all negotiate obstacle data. There was one potential influential point: the HD on trial cycle 1. Because the impact of this point is unknown, we perform all analysis with and without this point. All data on the scatter plot are valid for analysis.

Figure E-4. Negotiate Obstacle with Potential Influential Point Circled



The inclusion of the potential influential point does not change the statistical significance between groups. It does, however, change the SD and percentage differences between the integration levels. Once we remove the potential influential point, the percentage difference between the C group and HD group decreases. The SD for the HD group decreases without the potential influential point. Additionally, the removal of the potential influential point results in the data for HD group being normally distributed. The following sections discuss results with and without the potential influential point.

E.5.2.4.1.1 Negotiate Obstacle Descriptive Statistics with Potential Influential Point

The data for C group are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.77 but not normally distributed for the HD group, which had a p-value of <0.01. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a p-value of <0.01.

The C group had a mean time of 1.53 minutes. This time is statistically significantly faster than the HD mean time of 3.28 minutes. This difference results in a 114.75%, or 1.75-minute, degradation in time between the groups. The HD group had greater variability, as shown by the 1.02-minute increase in SD (0.22 minutes for the C group and 1.24 minutes for the HD group). See Table E-2 and Table E-3 for detailed analytical results.

E.5.2.4.1.2 Negotiate Obstacle Descriptive Statistics without Potential Influential Point

The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.77 for the C group and 0.45 for the HD group.

The C group had a mean time of 1.53 minutes. This time is statistically significantly faster than the HD mean time of 3.04 minutes. This difference results in a 99.12%, or 1.51-minute, degradation in time between the groups. The HD group had greater variability, as shown by the 0.31-minute increase in SD (0.22 minute for the C group and 0.53 minute for the HD group). See Table E-2 and Table E-3 for detailed analytical results.

E.5.2.4.2 Negotiate Obstacle Contextual Comments

Sun Tzu stated, "Speed is the essence of war. Take advantage of the enemy's unpreparedness; travel by unexpected routes and strike him where he has taken no precautions" (MCDP-1, p. 69). Speed and surprise are crucial to success. Furthermore, the enemy is reliably expected to cover obstacles with fires. When possible, obstacles are to be avoided. When not possible, they must be negotiated quickly. While no purely objective standard can be set for the negotiation of the obstacle presented in this task, any decrement in speed translates into increased exposure to enemy fires and greater risk for friendly casualties.

Throughout the assessment, within a given MOS and task, the C and HD groups used the same techniques to accomplish each task. However, this was not the case for this task. The influential point on Trial Cycle 1 represents the only occurrence in which the HD group performed this task using the same technique as the C group as a fully integrated squad (four females). During every other trial, the HD group utilized a mechanical advantage by attaching three web belts and using it as a ladder to aid in getting the last Marine over the obstacle. The HD group constructed this improvised ladder in advance of each trial for the duration of the experiment.

On average, the HD groups were 1.51 minutes slower than the C groups in performing the same task, a 99.12% degradation in performance. Through the course of the experiment, the slowest C group was faster than the fastest HD group. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the HD group 100% of the time (21 of 21 trial cycles). Based on the SDs, the variation in performance of the HD group is greater than twice as much as the variation in performance of the C group. This inconsistency in the performance of the HD group leads to greater uncertainty and less confidence in its future performance from the mean.

E.5.2.4.3 Negotiate Obstacle Additional Insights

During nine HD trials, a partially integrated squad of two or three females was assessed. The mechanical device was not utilized whenever a male was part of the squad. Four of these nine trials represent the four fastest HD times: 7 March 2015, 16 April 2015, 18 April 2015, and 25 April 2015 (Trial Cycles 1, 17, 18, and 21, respectively).

E.5.2.4.4 Negotiate Obstacle Subjective Comments

The subjective leadership log adds greater fidelity by stating that, for the HD group's four fastest trails (Trial Cycles 1, 17, 18, and 21), a male Marine "pulled all other Marines in team over the obstacle; performing majority of the work."

For additional subjective comments relating to this task, see the Appendix.

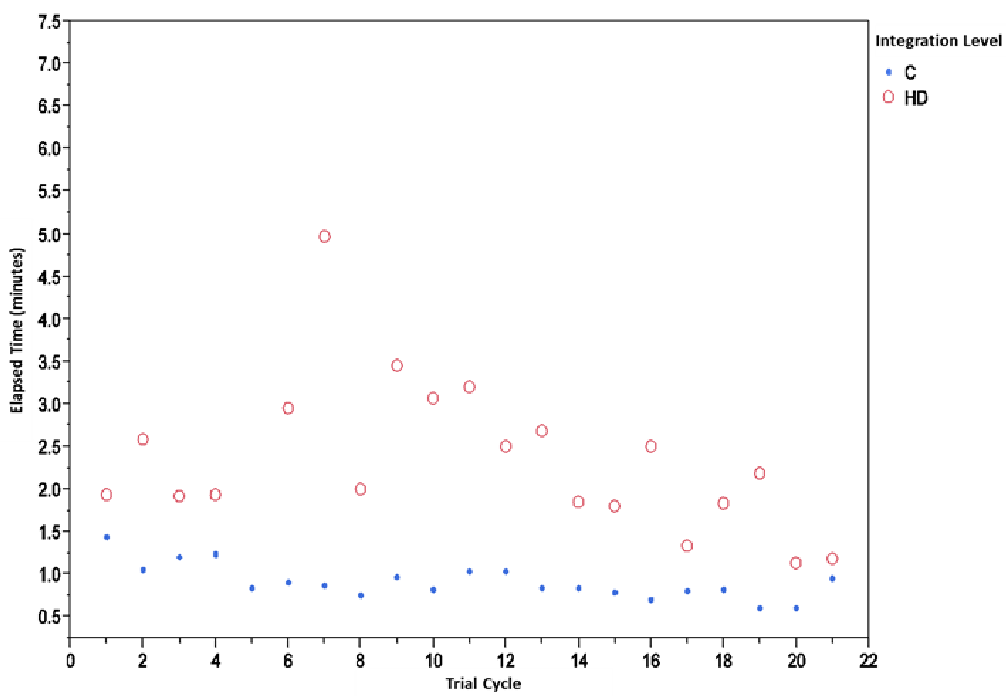
E.5.2.5 CASEVAC

E.5.2.5.1 CASEVAC Overview

This experimental task assessed the squad's ability to move a 220-lb dummy a distance of 100 meters to a CCP while wearing a fighting load, individual weapon, and two SMAW launchers. Marines conducted this task at the conclusion of the SMAW engagement. Squads could use a variety of techniques, but they had to move all personnel and gear the entire distance, as well as carry the dummy off the ground. The recorded time started when Marines touched the dummy and it stopped when the dummy and all members of the squad arrived at the CCP.

Figure E-5 displays all CASEVAC data. All data on the scatter plot are valid for analysis.

Figure E-5. CASEVAC



The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.18 for the C group and 0.06 for the HD group.

The C group had a mean time of 0.91 minute. This time is statistically significantly faster than the HD mean time of 2.35 minutes. This difference results in a 159.25%, or 1.44-minute, degradation in time between the groups. The HD group had greater variability, as shown by the 0.69-minute increase in SD (0.21 minute for the C group and 0.90 minute for the HD group). See Table E-2 and Table E-3 for detailed analytical results.

E.5.2.5.2 CASEVAC Contextual Comments

The implications of this task contain relevance to both the training and combat environment because a casualty must be moved expediently to a higher echelon of medical care. The data demonstrate that integrated squads took 1.44 minutes longer, a 159.25% degradation in performance, to accomplish the same 100-meter casualty movement. With the exception of 3 of 21 trials, the slowest C group was faster than the fastest HD group. During these three exceptions (Trial Cycles 17, 20, and 21), the HD group was partially integrated with three females, vice fully integrated. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the HD group 100% of the time (21 of 21 trial cycles). Based on the SDs, the variation in performance of the HD group is greater than four times the variation in performance of the C group. This inconsistency in the performance of the HD group leads to greater uncertainty and less confidence in their future performance from the mean.

E.5.2.5.3 CASEVAC Additional Insights

While the “Golden Hour” is a common medical planning construct for C2 and logistical support, medical literature supports a “Platinum Ten” philosophy of first response. The U.S. National Library of Medicine references a French article that states that “on the battlefield, the majority of casualties die within ten minutes of the trauma” (*Wounded in Action: The Platinum Ten Minutes and the Golden Hour*, Daban). The fundamental principle is that patient need to be correctly triaged and moved to medical care as fast as possible. Any time degradation reduces the probability of survival.

E.5.2.5.4 CASEVAC Subjective Comments

Seven instances in the leadership subjective logs indicated that an HD group required extra breaks in order to stop and put the dummy down, compared with zero comments for the C group. The implication is that a C group could have maintained its CASEVAC pace for a longer distance, while the HD group could not have maintained its CASEVAC pace over a longer distance. Therefore, the longer the CASEVAC distance, the greater the difference in time will be due to integration.

For additional subjective comments relating to this task, see the Appendix.

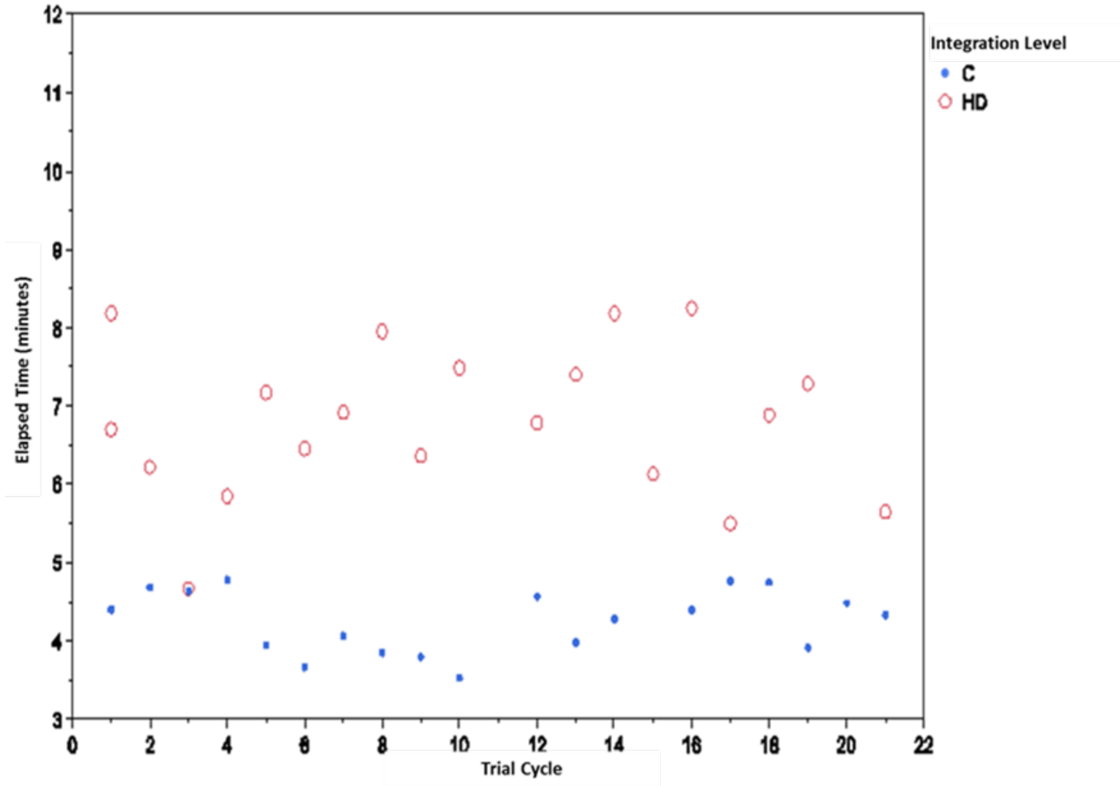
E.5.2.6 Mount/Dismount Saber

E.5.2.6.1 Mount/Dismount Saber Overview

This experimental task assessed the squad's ability to mount a Saber system onto a tactical vehicle (HMMWV) and then dismount it. A brief, non-assessed administrative pause was conducted between the mounting and dismounting subtasks. This task was conducted approximately 10 minutes after the conclusion of the 7-km hike. The recorded time started when Marines touched a component of the Saber, and it stopped when the last component of the Saber was back on the ground.

Figure E-6 displays all 035X mount/dismount Saber data. All data on the scatter plot are valid for analysis.

Figure E-6. Mount/Dismount Saber



The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.22 for the C group and 0.77 for the HD group.

The C group had a mean time of 4.26 minutes. This time is statistically significantly faster than the HD mean time of 6.80 minutes. This difference results in a 59.82%, or 2.54-minute, degradation in time between the groups. The HD group had greater variability, as shown by the 0.57-minute increase in SD (0.40 minute for the C group and 0.97 minute for the HD group). See Table E-2 and Table E-3 for detailed analytical results.

E.5.2.6.2 Mount/Dismount Saber Contextual Comments

The challenging aspect of mounting and dismounting the Saber system is the weight of the components. On average, it took the HD groups 2.54 minutes longer, a 59.82% degradation in performance. With the exception of one trial cycle, the fastest HD group was slower than the slowest C group throughout the duration of the experiment. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the HD group 100% of the time (19 of 19 trial cycles). Based on the SDs, the variation in performance of the HD group is greater than twice the variation in performance of the C group. This inconsistency in the performance of the HD group leads to greater uncertainty because there is less confidence in its future performance from the mean.

E.5.2.6.3 Mount/Dismount Saber Additional Insights

This task began with all components of the Saber system prestaged on the deck within 2 meters of the HMMWV. The Marines were not required to move the Saber system to/from the actual firing position. For an indication of the performance when moving heavy objects, see results for the TOW engagement (Section E.5.2.7), which involved moving a 50-lb TOW missile a distance of 100 meters.

E.5.2.6.4 Mount/Dismount Saber Subjective Comments

For subjective comments relating to this task, see the Appendix.

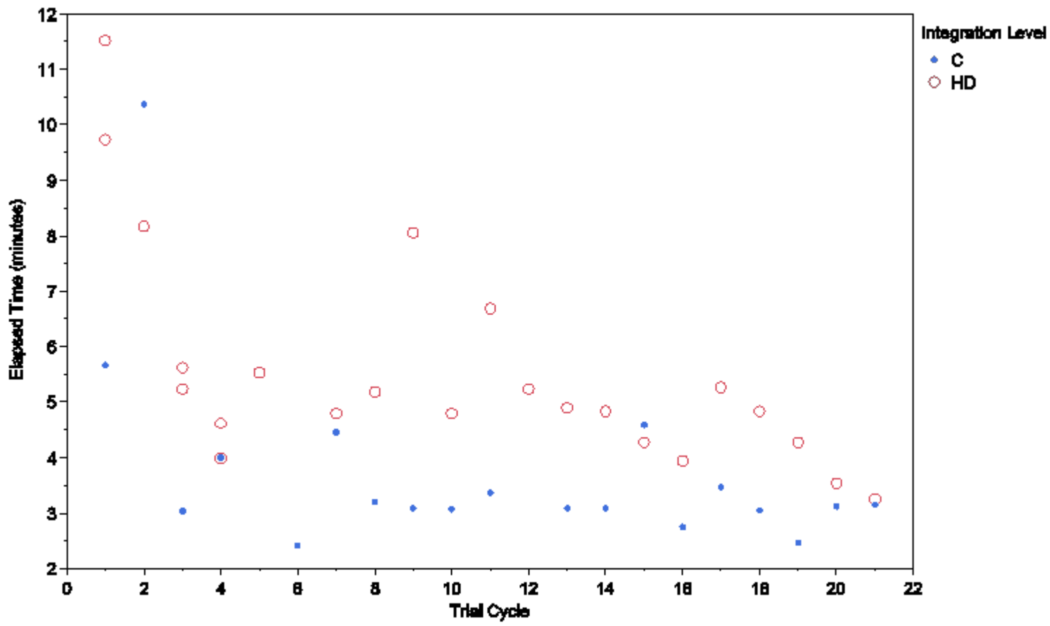
E.5.2.7 TOW Engagement

E.5.2.7.1 TOW Engagement Overview

This experimental task assessed the squad’s time to move four TOW missiles (each weighing 50 lb) approximately 100 m to a firing line and engage four targets. Due to experimental efficiencies, the Saber system was pre-established on a firing line. One missile reload per team was conducted between the first and second shots. The recorded time started when the squad departed the ORP, and it stopped when the last TOW missile was fired.

Figure E-7 displays all 035X TOW engagement data. All data on the scatter plot are valid for analysis.

Figure E-7. TOW Engagement



The data are not normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of <0.01 for the C and HD groups. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a p-value of <0.01.

The C group had a mean time of 3.76 minutes. This time is statistically significantly faster than the HD mean time of 5.58 minutes. This difference results in a 48.36%, or 1.82-minute, degradation in time between the groups. The HD group had greater variability, as shown by the 0.23-minute increase in SD (1.78 minutes for the C group and 2.01 minutes for the HD group). See Table E-2 and Table E-3 for detailed analytical results.

E.5.2.7.2 TOW Engagement Contextual Comments

Each TOW missile weighs over 50 lb. The ability to quickly move a TOW missile to a Saber system and effectively engage targets is important to eliminating enemy targets on the battlefield while minimizing exposure. On average, it took the HD group 1.82 minutes longer than the C group, a 48.36% degradation in performance, to accomplish the same movement and engagement. This provides the enemy that much more time to engage friendly positions. Furthermore, on any given day (under the same environmental conditions), the C group was faster than the HD group 89.5% of the time (17 of 19 trial cycles).

E.5.2.7.3 TOW Engagement Additional Insights

None.

E.5.2.7.4 TOW Engagement Subjective Comments

For subjective comments relating to this task, see the Appendix.

E.6 Statistical Modeling Results

E.6.1 Statistical Modeling Overview

The previous section discussed results only as they pertain to differences due to integration level alone. The goal of statistical modeling, as applied here, is to estimate, simultaneously, the effect of gender integration levels and other relevant variables on squad performance. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For the same seven selected tasks described in the previous section, this section presents an overview of the analysis and results and then presents the modeling results for each of the tasks.

For each task, we describe the significant variables in the model and determine whether they are positively or negatively correlated with the result. A negative correlation indicates that an increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time. The results indicate where certain patch numbers are significant for a given variable. The experiment tracked Marines within the squad by a patch number that associated their random position within the squad to a specific billet. Table E-4 displays the 035X patch numbers and billet titles.

Table E-4. 035X Patch Numbers and Billet Titles

Patch Number	Billet Title
1	Gunner
2	Assistant Gunner
3	Gunner
4	Assistant Gunner

E.6.2 035X Selected Task Method of Analysis

Because there are two different types of responses for the 035X selected tasks, we used two different methodologies. For all of the elapsed time response variables, we used a mixed effects model. For the SMAW hits, we used a multinomial logit model. The following sections describe each of these methodologies in more detail.

E.6.2.1 Mixed Effects Models

Due to the small number of trials, a mixed effects model with all 035X squad members and all types of personnel data does not work for the 035X data set. Thus, we model each personnel variable with integration level separately with a random effect for who filled each position within the 035X squad. For example, age for each member of the 035X squad (four variables), a random effect for who filled each billet, and integration level are modeled with the result (response time) as the response variable. Where maximum likelihood estimation converged, AIC was used for variable selection. Otherwise, we comment on the significance of individual variables in the full model. Variables reported as significant are concluded to be significant based on at least a one-sided test.

E.6.2.2 Multinomial Logit Model

We modeled the SMAW hits using a multinomial logit model where the number of impacts is the response variable and is modeled as a categorical variable. The levels for this variable are 0, 1, 2, 3, and 4, which correspond to the potential number of hits. Each personnel variable was modeled separately due to the small number of trials. For each model, we used the AIC to choose which variables to include in the model.

E.6.3 035X Selected Task Overall Modeling Results

Integration level is significant for all of the selected tasks modeled. For each of these tasks, modeling the random effects for the individuals participating in the task results in changes from the initial results in the descriptive statistics. Each respective task paragraph describes these changes.

The negotiate obstacle task had a potential influential point, and we model this task with and without this point. The HD integration level is significant for all variables in each of these analyses. Not all of the same variables remain significant when modeled with and without influential points. However, analysis with and without the potential influential point for the negotiate obstacle task does not change the overall result of the effect of the integration levels.

E.6.3.1 7-km Hike

We model elapsed time for the 7-km hike as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Age
- Armed Forces Qualification Test (AFQT) score
- General Classification Test (GT) score
- Physical Fitness Test (PFT) crunches
- PFT 3-mile run
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- Height.

The following personnel variables are significant in their respective models and are positively correlated with the 7-km hike time:

- AFQT score of patch 4
- CFT Movement to Contact of patches 1, 2, and 4
- CFT Maneuver Under Fire time of patches 2, 3, and 4.

The following personnel variables are significant in their respective models and are negatively correlated with the 7-km hike time:

- Height of patches 1, 2, 3, and 4
- Weight of patches 1, 2, and 4.

The following personnel variables have no significant variables in their respective models:

- None.

Height of all Marines was significant and negatively correlated for all Marines. This model also included integration level, which was negatively correlated with the 7-km hike time where the coefficient for integration level is -15.98. This is the only model for which integration level is significant and negatively correlated with the 7-km hike time. In the final model with only integration level, the coefficient for integration level is 23.00. Due to this large change in both magnitude and direction, it is highly likely there is a problematic multicollinearity between height and integration level in this model. This result is only reported for completeness as it is impractical and unsubstantial due to the negative correlation.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, the final model includes integration level only where HD has a difference of 23.00 minutes when compared with a C group and a p-value of <0.01. This difference is a decrease from the 26.39-minute difference identified in the descriptive statistics, which is a 12.85% change.

E.6.3.2 SMAW Engagement by Time

We model elapsed time for the SMAW engage targets as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Age
- Height
- Weight
- AFQT score
- GT score
- CFT Movement to Contact
- PFT crunches
- PFT 3-mile run
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the SMAW engage targets time:

- Age of patches 3 and 4
- Height of patch 4
- AFQT score of patch 3

- GT score of patch 3
- CFT Movement to Contact of patches 1 and 2
- CFT Maneuver Under Fire of patches 1, 2 and 3
- PFT 3-mile run of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the SMAW engage targets time:

- Height of patches 1 and 2
- Weight of patch 2
- CFT Movement to Contact of patch 4
- CFT Maneuver Under Fire of patch 4
- PFT crunches of patch 2
- PFT 3-mile run of patch 3.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, the final model includes integration level only where HD has a difference of 0.61 minute when compared with a C group and a p-value of <0.01. This difference is a decrease from the 0.76-minute difference identified in the descriptive statistics, which is a 19.74% change.

E.6.3.3 SMAW Engagement Accuracy

We model SMAW engagement accuracy as a function of each personnel variable and integration level in a separate multinomial logit model. The covariates in each model are the values of each personnel variable for each patch number and integration level. For each model, we report the variables chosen using AIC to remain in the model and if those variables have a positive, negative, or nonlinear correlation with the response.

Using AIC for variable selection, the HD integration level remained in the model in the following variable's respective models:

- Squad leader
- Age
- Weight
- AFQT score

- GT score
- PFT crunches
- PFT 3-mile run
- CFT Movement to Contact
- CFT Maneuver Under Fire
- Rifle score.

Model selection using AIC resulted in the following personnel variables in the final model, and they had a positive correlation with the SMAW engagement accuracy:

- Height of patch 1.

Model selection using AIC resulted in the following personnel variables in the final model, and they had a negative correlation with the SMAW engagement accuracy:

- Height of patches 2 and 4
- CFT Movement to Contact for patch 1.

Model selection using AIC resulted in the following personnel variables in the final model, and they had a nonlinear correlation with the SMAW engagement accuracy:

- PFT 3-mile run for patch 1
- Rifle score for patch 1.

The final model includes integration level only. Integration level is negatively correlated with the SMAW accuracy, meaning changing integration level from C to HD results in a lower probability of hit. Table E-5 shows the probabilities of obtaining a specific number of hits on target by integration level based on the model results.

Table E-5. Probability of SMAW Hit by Integration Level

Number of Hits	C	HD
0	0.00%	8.33%
1	5.00%	20.83%
2	5.00%	41.67%
3	60.00%	25.00%
4	30.00%	4.17%

E.6.3.4 Negotiate Obstacle Overview

We model elapsed time for the negotiate obstacle as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

There was one potential influential point (HD group on trial cycle 1), and we model the negotiate obstacle time with and without this point.

E.6.3.4.1 Negotiate Obstacle with Potential Influential Point

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- All.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the negotiate obstacle time:

- Age of patch 2
- CFT Maneuver Under Fire of patch 1
- PFT crunches of patch 2
- Rifle score of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the negotiate obstacle time:

- Weight of patches 2 and 3
- CFT Movement to Contact of patch 2
- PFT crunches of patch 4.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, the final model includes integration level only where HD has a difference of 1.91 minutes when compared with a C group and a p-value of <0.01. This difference is an increase from the 1.75-minute difference identified in the descriptive statistics, which is an 8.38% change.

E.6.3.4.2 Negotiate Obstacle without Potential Influential Point

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- All.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the negotiate obstacle time:

- Age of patch 3
- CFT Movement to Contact of patch 1
- CFT Maneuver Under Fire of patch 1
- Rifle score of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the negotiate obstacle time:

- Squad leader
- Height of patch 1
- Weight of patch 1.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, the final model includes integration level only where HD has a difference of 1.55 minutes when compared with a C group and a p-value of <0.01. This difference is an increase from the 1.51-minute difference identified in the descriptive statistics, which is a 2.65% change.

E.6.3.5 CASEVAC

We model elapsed time for the CASEVAC as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- All.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the CASEVAC time:

- Age of patch 3
- AFQT score of patch 3
- GT score of patch 3
- CFT Movement to Contact of patch 1
- CFT Maneuver Under Fire of patch 1, 2, and 3
- PFT 3-mile run of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the CASEVAC time:

- Age of patch 2
- Height of patches 1 and 2
- Weight of patch 1
- AFQT score of patch 2
- PFT crunches of patches 2
- Rifle score of patch 1.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, the final model includes integration level only where HD has a difference of 1.45 minutes when compared to a C group and a p-value of <0.01. This difference is an increase from the 1.44-minute difference identified in the descriptive statistics, which is a 0.69% change.

E.6.3.6 Mount/Dismount Saber

We model elapsed time for the mount/dismount Saber as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- All.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the mount/dismount Saber time:

- AFQT score of patch 4
- GT score of patch 4
- CFT Maneuver Under Fire of patch 3
- PFT 3-mile run of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the mount/dismount Saber time:

- Height of patch 3
- CFT Maneuver Under Fire time of patch 2
- PFT crunches of patch 1
- PFT 3-mile run of patch 4.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, the final model includes integration level only where HD has a difference of 2.25 minutes when compared with a C group and a p-value of <0.01. This difference is a decrease from the 2.54-minute difference identified in the descriptive statistics, which is an 11.42% change.

E.6.3.7 TOW Engagement

We model elapsed time for the TOW engagement as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position in the squad. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- All.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the TOW engagement time:

- AFQT score of patch 1
- PFT 3-mile run of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the TOW engagement time:

- Age of patch 1
- AFQT score of patch 2
- PFT crunches of patch 2.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, the final model includes integration level only where HD has a difference of 2.22 minutes when compared with a C group and a p-value of 0.02. This difference is an increase from the 1.82-minute difference identified in the descriptive statistics, which is a 21.98% change.

Appendix to Annex E
035X Supplemental Information

This appendix provides supplemental information for the 035X portion of the GCEITF experiment. It provides information regarding the GCEITF leadership subjective comments and additional descriptive and basic inferential statistics not described in Annex E.

Section 1. GCEITF Leadership Subjective Comments

The GCEITF leadership provided comments on their observations of the experiment throughout it execution. Table E A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table E A – Summary of GCEITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement			Requesting extra breaks			Requires extra assistance			Needs no assistance			Compensating for another Marine			Gear pass off			Other			No category			Total
		C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	
7-km Hike	M	13	0	13	1	0	1	0	0	0	0	3	3	0	0	0	0	0	0	2	0	2	0	0	0	19
	F	0	20	20	0	1	1	0	1	1	0	4	4	0	0	0	0	6	6	0	0	0	0	0	0	32
	Unit	0	3	3	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	9
SMAW Engagement by Time	M	1	3	4	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
	F	0	16	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16
	Unit	0	2	2	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3
SMAW Engagement Accuracy	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1
	F	0	0	0	0	0	0	0	4	4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	5
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Negotiate Obstacle	M	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5	0	0	0	1	0	1	0	0	0	6
	F	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	Unit	0	0	0	0	0	0	0	1	1	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3
CASEVAC to CCP	M	0	0	0	0	0	0	0	0	0	2	2	4	0	0	0	0	0	0	0	0	0	0	0	0	4
	F	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	Unit	0	1	1	0	2	2	0	3	3	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	9
Mount/Dismount TOW; by Squad	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	2
Engage Target (TOW)	M	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	2
	F	0	0	0	0	0	0	0	5	5	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	6
	Unit	0	1	1	0	0	0	1	2	3	0	0	0	0	0	0	0	0	0	2	2	4	0	0	0	8
1-km Hike	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
	Unit	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1

Section 2. Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for 11 additional 035X tasks. Annex E contains the descriptive statistics for the remainder of the 035X tasks. The words “metric” and “task” are used interchangeably throughout this Appendix. They both refer to the experimental task.

The two tables below display the results for 11 additional 035X metrics. Table E B displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent differences between integration levels.

Table E C displays ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare the two groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the HD group is different from that in the C group. We present basic inferential statistics for four tasks.

Table E B – 035X Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
7--km Hike; first km	C	19	10.85	0.49	11.25%
	HD	16	12.07	0.61	
7-km Hike; second km	C	19	11.21	0.55	21.54%
	HD	16	13.63	1.76	
7-km Hike; third km	C	16	11.52	1.62	21.58%
	HD	18	14.01	2.82	
7-km Hike; fourth km	C	16	12.94	3.02	32.40%
	HD	19	17.13	3.10	
7-km Hike; fifth km	C	19	17.60	3.39	13.99%
	HD	19	20.06	4.58	
7-km Hike; sixth km	C	17	13.88	2.37	34.05%
	HD	20	18.61	2.78	
7-km Hike; seventh km	C	19	15.38	2.26	33.21%
	HD	20	20.49	3.68	
1-km Hike	C	21	8.82	0.62	17.33%
	HD	22	10.34	0.45	
Mount Saber	C	19	2.36	0.32	73.77%
	HD	20	4.10	0.76	
Dismount Saber	C	20	1.92	0.23	41.18%

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
	HD	21	2.71	0.62	
TOW accuracy (% hit)	C	21	0.94	0.16	-4.78%
	HD	24	0.90	0.19	

Table E C – 035X ANOVA Results and Welch's T-Test Results

Metric	F statistic (df)	F Test P-value	Comparison	2-sided P-Value	1-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
1-km Hike*	86.58 (1, 41)	< 0.01*	HD-C	< 0.01*	< 0.01*	1.31	1.74	1.25	1.81
Mount Saber*	85.60‡	< 0.01‡	HD-C	< 0.01*	< 0.01*	1.50	1.98	1.42	2.05
Dismount Saber*	30.23‡	< 0.01‡	HD-C	< 0.01*	< 0.01*	0.60	0.98	0.55	1.04
TOW Accuracy (% hit)	0.76 (1, 43)	0.39	HD-C	0.35†	0.18†	-3.25e-5†	4.83e-5†	-1.69e-5†	3.65e-5†

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality.
‡Indicates results presented are from Robust ANOVA due to unequal variances. The reported F-statistic is a Chi-square statistic from Robust ANOVA, and the F-test p-value is the Robust ANOVA p-value.

Additional Task Results:

7-km Hike by km. The general trend for the 7-km hike was that the difference between the HD and C groups increased over the course of the hike.

1-km Hike. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.97 for the C group and 0.01 for the HD group.

The C group had a mean time of 8.82 minutes. This time is statistically significantly slower than the HD group mean time of 10.34 minutes. The HD group was 17.33% slower than the C group.

Mount Saber. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.26 for the C group and 0.72 for the HD group.

The C group had a mean of 2.36 minutes. This time is statistically significantly slower than the HD group mean of 4.10 minutes. The HD group was 73.77% slower than the C group.

- **Contextual Comments.** One challenging aspect of mounting the Saber system is the weight of the sub-components and missile. This task affects combat capability during the preparation phase of a combat operation. Based on the standard deviations, the variation in performance of the HD group is twice as great as the variation in performance of the C group. This inconsistency in the performance of the HD group leads to greater uncertainty, as there is less confidence in their future performance from the mean.

Dismount Saber. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.23 for the C group and 0.04 for the HD group.

The C group had a mean of 1.92 minutes. This time is statistically significantly slower than the HD group mean of 2.71 minutes. The HD group was 41.18% slower than the C group.

- **Contextual Comments**. One challenging aspect of dismounting the Saber system is the weight of the sub-components. This task affects combat capability during the preparation and conduct phases of an operation. Increased time to complete this task during mission conduct reduces time available for the target engagement process and increases exposure time. Based on the standard deviations, the variation in performance of the HD group is twice as great as the variation in performance of the C group. This inconsistency in the performance of the HD group leads to greater uncertainty, as there is less confidence in their future performance from the mean.
- Lift limitations identified when lifting the sub-components of this weapon system pose additional anti-tank missilemen capability concerns, as it is not uncommon for Marines to be required to conduct a combat reload of a TOW missile from inside or outside a tactical vehicle. The timely execution of these transitional tasks is critical to the lethality of all mounted combat units, from vehicle team to Combined Anti-armor Team. If an inability to quickly dismount subcomponents of this system is related to time required to load and unload the TOW missile, additional combat capability may be affected.

TOW Accuracy (% hit). The data are not normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of <0.01 for the C and HD groups.

The C group had a mean percent hit of 94%. This percentage is higher (but not statistically significantly) than the HD group mean of 90%. The HD group was 4.78% less accurate than the C group.

Annex F.
Provisional Infantry Rifleman (PI)

This annex details the Provisional Infantry (PI) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed 2 March – 26 April 2015 at Range 107 and Range 110, aboard the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, CA. The sections outline the PI Scheme of Maneuver (SOM), Limitations, Deviations, Data Set Description, Descriptive and Basic Inferential Statistics, Subjective Comments, and Modeling Results.

F.1 Scheme of Maneuver

F.1.1 Experimental Cycle Overview

The PI assessment of the GCEITF took place in a field environment aboard MCAGCC, Twentynine Palms, CA. The assessment consisted of 21 trial cycles, each of which was a 2-day test cycle, conducted over the course of 55 days. After every 4 days of trials, the Marines received 1 recovery day spent at Camp Wilson. Each squad consisted of 12 volunteers and a direct assignment (non-volunteer) squad leader. Each member of the squad was trained to fill each billet within the fireteam: fireteam leader, grenadier, automatic rifleman, and rifleman. The rifle squad contained three fireteams. The assessment was executed under the supervision of MCOTEA functional test managers and a range officer in charge (OIC)/range safety officer (RSO) from the GCEITF.

F.1.2 Experimental Details

The 2-day PI assessment was modeled to replicate defensive and offensive tasks. The PI conducted defensive tasks on Day 1 and offensive tasks on Day 2. Three PI squads executed each trial cycle: a control (C) non-integrated squad, a low-density (LD) gender-integrated squad with two females, and a high-density (HD) gender-integrated squad with four or five females.

Day 1 of the PI trial cycle was executed on Range 110 and consisted of defensive actions. The day started with a 7-km forced march from Range 107 to Range 110, for which Marines were wearing an approach load and carrying personal weapons. After a 10-minute operational pause, each fireteam spent 2 hours digging two-man fighting positions wearing a fighting load. Two Marines from each fireteam provided security while the other two dug, rotating every 15 minutes.

Day 2 of the PI trial cycle was executed on Range 107 and consisted of a squad-reinforced attack. Each squad moved approximately 1 km to an assault position wearing an assault load and carrying personal weapons. As a squad, they moved all personnel and gear over an 8-ft obstacle/wall. After negotiating the obstacle, the squad staged its assault packs, deployed on-line, and conducted a 425-m live-fire assault to a limit of advance wearing the fighting load. They immediately repelled an enemy counterattack by fire for 90 seconds. Finally, each fireteam conducted a 100-m casualty evacuation (CASEVAC) of a 220-lb. At the conclusion of Day 2, the composition of the squad was dissolved and a new sample of males and females were randomly assigned for the following cycle.

F.1.3 Additional Context

Throughout the duration of the assessment, Marines bivouacked at Range 107, sleeping in two-man tents. Prior to the Day 2 trial, each Marine zeroed his or her weapon system to maximize accuracy. During trial execution, Marines wore/carried prescribed loads for each task. Packs were weighed each day prior to the 7-km forced march to ensure consistency. After each trial day, the Marines operated under the guidance of their Company leadership, performing minimal physically demanding tasks. Not all Marines were selected for each trial cycle. Those Marines who were not part of an assessed squad conducted the same experimental subtasks to ensure equity in physical loading between individuals participating in a trial cycle and those not chosen via random selection for that particular cycle. A detailed discussion of these tasks is in the Loading Plan section below.

Fatigue surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. Outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) are accounted for in the analysis of the performance data. Workload surveys collected the volunteers' perceived average and maximal level of exertion during the performance of specified tasks. Cohesion surveys provided a method of collecting subjective data relating to each squad's ability to work as a team and their overall perspective on the cohesiveness of the squad.

F.1.4 Experimental Tasks

F.1.4.1 7-km Hike

Moving under a load is one of the most fundamental tasks of an infantry unit; it is both physically and mentally demanding. Units train by conducting tactical marches with an approach load at increasing distances. The Infantry Training and Readiness (T&R) Manual states that "the approach march load will be such that the average infantry Marine will be able to conduct a 20 mile hike in 8-hours with the reasonable expectation of maintaining 90% combat effectiveness." During the GCEITF assessment, each PI squad moved a distance of 7.20 km, from Range 107 to Range 110. This route was flat (minimal elevation change) and conducted on an unimproved surface with varying degrees of conditions (compact dirt and loose sand). The squads moved as fast as the slowest person, carrying an approach load and an individual weapon. The individual weapons included the M-4, the M-16A4 with M203 attachment, and the M27, resulting in a cumulative load of 96 – 101 lb per Marine. The primary purpose of this task was to determine the squad's pace over a 7.20-km route while carrying the approach load. Each Marine took a fatigue and workload survey after completion of the 7.20-km hike (see GCEITF Experimental Assessment Plan [EAP], Annex D).

F.1.4.2 Digging Fighting Holes

Protection is vital to infantry units, especially in a defensive posture. Infantry units commonly construct fighting positions to conceal positions and minimize exposure to enemy fire. The most physically demanding aspect of constructing a doctrinal fighting position is digging out the fighting hole. The terrain at Twentynine Palms consisted of hard, compact sand and rocks. Each fireteam dug 2 two-man fighting holes in a time limit of 2 hours, maintaining 50% security, meaning that two Marines dug while two provided security in the prone position. The Marines swapped positions every 15 minutes. The purpose of this task was to determine the fireteam's rate of work while digging fighting positions. After 2 hours, each Marine took a fatigue and workload survey (see GCEITF EAP, Annex D).

F.1.4.3 1-km Movement

Assaulting an enemy position never starts from a static position; first, a movement must be conducted to the assault position (AP). The distance from the line of departure to the AP is dependent upon myriad factors. Based on time and space constraints during the GCEITF assessment, this distance was a little less than 1 km. Each PI squad moved this distance as quickly as possible while carrying an assault load and individual weapon. The Infantry T&R Manual states that "the assault load is the load that is needed during the actual conduct of the assault." This task caused moderate fatigue for the Marines prior to commencing the attack.

F.1.4.4 Negotiate an Obstacle

The conduct of an attack often involves reducing or negotiating an obstacle. It is common in an urban environment to enter through a window or over a wall (obstacle). One of the more difficult tasks is climbing over a wall with a fighting load. Each squad negotiated an 8-ft wall by getting all Marines and equipment over as quickly as possible. Although the technique was not dictated, the PI squads used three launch-points (one per fireteam) from which to lift Marines up onto the wall.

F.1.4.5 Squad Attack

The mission of the rifle squad is "to locate, close-with, and destroy the enemy by fire and maneuver, and repel the enemy's assault by fire and close combat." Conducting a live fire-and-movement task is fundamental to the infantry community and operationally relevant. Having moved to an AP, each PI squad deployed into an on-line formation to maximize firepower in the direction of the enemy and then began movement-to-contact with weapons at the ready. Upon contact with the enemy (targets being presented), the Marines laid down a heavy volume of fire to gain fire superiority. They then began buddy rushes to close with the enemy objective. Each squad conducted approximately 300 m of buddy rushes, engaging a total of twelve GCEITF Targeting Systems (GTSs) with direct-fire weapons and three machinegun bunkers with the M16A4 with M203 grenade launcher. After all targets were destroyed (no targets remaining), the squad moved another 125 m to a limit of advance (LOA) and prepared for an enemy counterattack. This task is physically demanding and involves a combination of accuracy, tempo, and squad cohesion.

F.1.4.6 Repel Counterattack

At the conclusion of every assault, the attacking unit must be prepared for the enemy to regroup and organize a counterattack. Upon consolidation at the LOA, the PI squad oriented its weapons in the direction of the enemy's retreat. Upon target presentation, the counterattack commenced. It continued for 90 seconds. Six GTSs were used during the course of the counterattack. The primary purpose of this task was to determine the PI squad's accuracy engaging targets approximately 300 m away after conducting an offensive squad attack.

F.1.4.7 CASEVAC

Casualties are an inevitable part of conducting combat operations. Units train to become proficient in triaging, handling, and transporting a casualty. After a Marine is injured, it is essential to move the casualty to the appropriate level of care as quickly as possible. During the GCEITF assessment, each PI fireteam was assigned a casualty at the conclusion of the live-fire attack and counterattack. Each fireteam had to move a 220-lb casualty (dummy and equipment) a distance of 100 meters to a casualty collection point (CCP). The team could use a variety of techniques for transport, but had to carry the dummy off the ground (not drag any part). The primary purpose of this task was to determine the fireteam's proficiency at moving a simulated casualty to a CCP. After the CASEVAC, each Marine took a fatigue and workload survey to assess overall fatigue and workload of the entire offensive task (see GCEITF EAP, Annex D).

At the completion of the 2-day cycle, Marines also took a cohesion survey to record cohesion during the execution of the 2-day trial cycle (See GCEITF EAP, Annex M).

F.1.5 Loading Plan

Due to the number of volunteers, a handful of Marines were not part of an assessed squad for each 2-day cycle. The primary purpose of the loading plan was to ensure, to the greatest extent possible, equity of physical activity among all volunteers throughout the duration of the experimental assessment. Every trial and task was conducted in the same manner and sequence to ensure consistency. Collaboration with the Company leadership determined that the best method of loading non-assessed Marines was to have them perform the same tasks as an assessed PI squad so that they would experience the same conditions and physical strain. Every trial and task was conducted in the same manner and sequence to ensure consistency.

F.1.5.1 Variations

A variety of factors, including safety and the number of loading Marines, introduced variations to the loading plan.

- In some instances, the loading Marines formed a quasi-squad and conducted the trial after all assessed PI squads were done for the day. At other times, the quasi-squad was too small, so the loading Marines were attached to an assessed squad, in which case they operated on the flanks and were not given ammunition. At no point in time did a loading Marine aid or interfere with an assessed PI Marine/squad.
- The loading Marines did not always dig for 2 hours. To gain efficiencies, the loading Marines would rotate between digging, assisting in weighing buckets of sand, and refilling the holes of the assessed squads. Any such variation was carefully calculated to ensure that loading Marines were doing an amount of work equivalent to that of the assessed Marines.

F.1.6 Scheme of Maneuver Summary

The PI experiment consisted of a 2-day trial cycle comprising a defensive and an offensive day. The defensive day involved two subtasks: 7-km forced march and digging a fighting position. The offensive day involved five subtasks based around a squad live-fire attack: 1-km movement, negotiating an obstacle, fire and movement, counterattack, and CASEVAC. During the course of the experiment, the PI squad executed 2 pilot trial cycles and 21 record trial cycles. During trial execution, Marines rotated through every billet within the rifle squad, carrying the respective weapon system.

F.2 Limitations

F.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a similar manner to those in an operational environment. However, under certain situations, artificial limitations or interruptions were introduced that changed or altered the way a task would normally be performed. While these limitations represent a degree of artificiality, they do not detract significantly from our abilities to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for the PI assessment.

F.2.2 Relative Difficulty of Record Test

The PI GCEITF assessment was designed to gather data associated with some of the most physically demanding tasks that an open MOS Marine could perform when assigned to an Infantry unit with collateral responsibilities as a rifleman. However, these tasks in isolation did not fully replicate life experienced by a Marine during a typical field exercise (FEX) or in a combat environment and did not obtain the maximum cumulative load that could be placed on a Marine assigned to a ground combat element. With limited time available, only selected PI tasks were assessed. Other tasks/duties outside of the assessment were minimized due to specific experimental constraints and human factors. During a typical FEX, it is common for Marines to conduct 24-hour operations that include performing daytime and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. It took PI squads approximately 4 hours to complete the defensive day and approximately 1 hour to complete the offensive day. Outside the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

Another concern in designing the PI assessment was to ensure that it was achievable and sustainable for a 60-day period. The 7-km forced march distance was selected based on the training time available prior to the assessment. However, many of the loads carried were decreased. Once they were over the wall/obstacle, the provisional infantrymen staged their assault packs without security rather than wearing them during the fire-and-movement. The provisional infantrymen also did not carry ammunition, radios, batteries, or other equipment often required when operating in a tactical environment. The Marines were authorized 1 day off after every 4 days of training; this artificial recovery period is not achievable when conducting training or combat operations.

A final factor affecting the relative difficulty of the record test pertains to the intangible physiological impact of the volunteers' ability to drop on request (DOR) at any point during a trial. Any time a volunteer dropped during a trial cycle, that PI squad performed the following subtasks with fewer personnel. This factor affected the cohesion of each squad and influenced its performance.

F.2.3 Digging a Fighting Position

Several artificialities were present as the volunteers dug their fighting holes. Preparing a defensive fighting position involves many continuing actions by those not actually digging the hole, such as clearing fields of fire and creating sectors of fire. However, no continuing defensive actions were conducted other than lying in the prone position and providing notional security. The Marines digging did not have to fill up a 5-gallon bucket for measurement; they either filled sandbags or built up a parapet. These artificialities may have influenced their performance and morale.

F.2.4 Number of Volunteer Participants

A total of 33 male and 8 female volunteers began the experiment; of those, 29 males and 6 females completed the assessment. The results presented in this annex are based on the performance of 35 – 41 Marines. For population analysis (by gender and MOS), which examines the volunteer population and the overall Marine Corps population, see Population Analysis Annex Q.

F.2.5 Limitations Summary

The PI assessment was designed to replicate realistic training in a field environment. The end-state was to create an experiment in which the volunteers felt they were conducting realistic and operationally relevant tasks, but unavoidable limitations to the assessed tasks and non-assessed operating environment introduced a level of artificiality not normally present in a field training or combat environment.

F.3 Deviations

F.3.1 Concentration for the Provisional Rifle Squad

Marines participated in this experiment voluntarily and could leave the experiment at any time, for any reason, making the total number of human subjects available highly volatile. Trials were not executed if there were not enough participants to carry out the assigned mission, and attrition affected the integration levels for the PI experiment. The EAP stated that the PI Rifle Squad would have three levels of integration: a C group, an LD group, and an HD group, with zero females, two females, and four females, respectively. Due to the low number of male volunteers, the MCOTEA research team had to add a female to the high-density group for some trials, resulting in the final high-density integration level being either four or five females.

F.4 Data Set Description

F.4.1 Data Set Overview

The PI portion of the experiment consisted of 2 pilot trial cycles and 21 record trial cycles. The pilot trial cycles were conducted 2 March – 6 March 2015. Pilot trial cycle data are not used in analysis, due to variations in the conduct of the test. We based all analysis on the 21 record trial cycles executed 7 March – 26 April 2015.

F.4.2 Record Test Volunteer Participants

At the beginning of the first record trial cycle, there were 33 male PI volunteers and 8 female volunteers. Several of those voluntarily withdrew, or were involuntarily withdrawn, during the execution of the experiment. The final number of volunteers was 29 males and 6 females.

F.4.3 Planned, Executed and Analyzed Trial Cycles

Table F-1 displays the number of trial cycles planned, executed, and analyzed by task. The planned number of trial cycles for the PI MOS per Section 7.5.3 of GCEITF EAP was 78 trials, or 26 trials per planned integration level (C, LD, and HD). Due to the change in trial execution, only 21 trial cycles could be executed. The planned number of trial cycles in Table F-1 reflects the 21 planned trial cycles for each integration level.

Of note, there are several occurrences of missing data for the 7-km hike by individual kilometer. The individual kilometer times were derived from GPS data. Early in the experiment, the Garmin GPSs were set to record a volunteer's position every second. Due to the storage space on the GPS and length of the trial when volunteers executed the 7-km hike and then 2 hours of preparing a fighting position, the GPS could not hold all of the data. Therefore, it overwrote the hike data. Once the problem was found, the GPSs were corrected to record location every 2 seconds.

Table F-1. PI Planned, Executed, and Analyzed Trial Cycles

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
7-km Hike	C	21	21	21	
	LD	21	18	18	
	HD	21	18	18	
Movement to LOA	C	21	21	19	Missing TRACR 12 Apr (data problem); 18 Mar controller
	LD	21	18	16	Missing TRACR 12 Apr (data problem); 18 Mar controller
	HD	21	18	16	Missing TRACR 12 Apr (data problem); 18 Mar controller
Attack & C-Atk Percent Hits	C	21	21	19	Missing TRACR 12 Apr (data problem), 24 Apr
	LD	21	18	17	Missing TRACR 12 Apr (data problem)
	HD	21	18	16	Missing TRACR 13 Mar, TRACR 12 Apr (data problem)
CASEVAC by FT	C	105	96	96	
	LD	42	41	41	Partial trial 20 Mar
	HD	42	36	34	No 15 Mar (2 HD FTs)
Prepare Fighting Positions by FT	C	105	96	95	27 Mar missing data
	LD	42	43	43	
	HD	42	36	34	No 14 Mar (2 HD FTs)
1-km Hike	C	21	21	21	
	LD	21	18	18	
	HD	21	18	18	
Negotiate	C	21	21	20	2 Apr high outlier

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
Obstacle	LD	21	18	18	
	HD	21	18	17	12 Apr high outlier
7-km Hike; 1km Time	C	21	21	17	No data: 7 Mar, 9 Mar, 12 Mar, 14 Mar
	LD	21	18	15	No data: 7 Mar, 9 Mar, 12 Mar
	HD	21	18	16	No data: 7 Mar, 12 Mar
7-km Hike; 2km Time	C	21	21	17	No data: 7 Mar, 9 Mar, 12 Mar, 14 Mar
	LD	21	18	15	No data: 7 Mar, 9 Mar, 12 Mar
	HD	21	18	16	No data: 7 Mar, 12 Mar
7-km Hike; 3km Time	C	21	21	17	No data: 7 Mar, 9 Mar, 12 Mar, 14 Mar
	LD	21	18	15	No data: 7 Mar, 9 Mar, 12 Mar
	HD	21	18	16	No data: 7 Mar, 12 Mar
7-km Hike; 4km Time	C	21	21	17	No data: 7 Mar, 9 Mar, 12 Mar, 14 Mar
	LD	21	18	15	No data: 7 Mar, 9 Mar, 12 Mar
	HD	21	18	16	No data: 7 Mar, 12 Mar
7-km Hike; 5km Time	C	21	21	17	No data: 7 Mar, 9 Mar, 12 Mar, 14 Mar
	LD	21	18	14	No data: 7 Mar, 9 Mar, 12 Mar; 8 Apr high outlier
	HD	21	18	16	No data: 7 Mar, 12 Mar
7-km Hike; 6km Time	C	21	21	17	No data: 7 Mar, 9 Mar, 12 Mar, 14 Mar
	LD	21	18	15	No data: 7 Mar, 9 Mar, 12 Mar
	HD	21	18	15	No data: 7 Mar, 12 Mar; 23 Apr high outlier
7-km Hike; 7km Time	C	21	21	18	No data: 7 Mar, 12 Mar, 14 Mar
	LD	21	18	15	No data: 7 Mar, 9 Mar, 12 Mar
	HD	21	18	16	No data: 7 Mar, 12 Mar

F.5 Descriptive and Basic Inferential Statistics

F.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common infantry tasks and are indicative of unit-level proficiency during either field exercises or combat operations. This section presents results for 7 out of 16 tasks. The PI Appendix contains the descriptive statistics for the remainder of the PI tasks. The words “metric” and “task” are used interchangeably throughout this Annex. They both refer to the experimental task.

Each fireteam consisted of four volunteer Marines: the fireteam leader, automatic rifleman, grenadier, and rifleman. Each squad consisted of 12 volunteer Marines (three fireteams) with a direct assignment (non-volunteer) squad leader. There were three integration levels for all tasks. For squad-level tasks, a C group was non gender integrated, an LD was gender integrated with two female Marines, and an HD group was gender integrated with four or five female Marines. For fireteam-level tasks, a C group was non gender integrated, an LD group was gender integrated with one female Marine, and an HD group contained at least two, but not more than three, females.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA), Tukey Tests (or non-parametric tests as necessary), and scatter plots. The subsequent sections will cover each task in detail. Lastly, contextual comments, additional insights, and subjective comments (as applicable) tying back to each experimental task are incorporated.

Caution must be used when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks, such as distances, techniques, leadership, load carried, group size, and group composition.

F.5.2 PI Selected Task Descriptive Statistics Results

The two tables below display the results for the seven selected PI metrics. Table F-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations. Table F-3 displays ANOVA and Tukey Test results, including metrics and integration levels, p-values suggesting statistical significance, elapsed-time differences between the integration levels, and percentage differences between integration levels. For each task, an ANOVA was conducted to compare the three groups and Tukey Tests were conducted to compare each pair of two groups. In cases where non-parametric tests were needed, Table F-3 displays these results instead of ANOVA and Tukey Test results. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the result was not found to be the same across all three groups.

Table F-2. PI Selected Task Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD
7-km Hike (minutes)*	C	21	82.75	4.75
	LD	18	90.76	6.71
	HD	18	91.05	3.85
Movement to LOA (minutes)†	C	19	3.01	0.44
	LD	16	3.22	0.36
	HD	16	3.37	0.43
Attack & C-Atk Percent Hits (%)	C	19	0.39	0.05
	LD	17	0.37	0.04
	HD	16	0.36	0.07
CASEVAC by FT (minutes)*	C	96	1.31	0.34
	LD	41	1.57	0.34
	HD	34	1.63	0.42
Prepare Fighting Positions by FT (lbs.)	C	95	5146.13	963.27
	LD	43	5257.46	1078.25
	HD	34	5184.84	1097.58
1-km Hike (minutes)*	C	21	8.87	0.58
	LD	18	9.55	0.56
	HD	18	9.37	0.44
Negotiate Obstacle (minutes)*	C	20	1.75	0.19

Metric	Integration Level	Sample Size	Mean	SD
	LD	18	1.78	0.24
	HD	17	2.13	0.34

*Indicates that there is a statistically significant difference in a two-sided hypothesis test between integration levels according to ANOVA and those according to a non-parametric equivalent test.

†Indicates that there are contradicting statistical significance results between ANOVA and a non-parametric equivalent test.

Table F-3. PI Selected Task ANOVA and Tukey Test Results

Metric	F statistic (df)	F Test P-value	Comparison	Difference	% Difference	2-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
7-km Hike (minutes)*	16.23 (2, 54)	< 0.01*	LD-C	8.01	9.68%	< 0.01*	5.10	10.92	4.50	11.52
			HD-C	8.30	10.03%	< 0.01*	5.39	11.22	4.79	11.82
			HD-LD	0.29	0.32%	0.98	-2.73	3.32	-3.35	3.94
Movement to LOA (minutes)*	3.39 (2, 48)	0.76†	LD-C	0.22	7.15%	0.04†	0.15†	0.37†	0.08†	0.40†
			HD-C	0.36	12.00%	0.01†	0.23†	0.55†	0.02†	0.60†
			HD-LD	0.15	4.52%	0.16†	0.00†	0.28†	-0.02†	0.37†
Attack & C-Atk Percent Hits (%)	2.25 (2, 49)	0.12	LD-C	-0.03	-6.62%	0.33	-0.06	0.01	-0.06	0.01
			HD-C	-0.04	-9.63%	0.11	-0.07	-0.01	-0.08	0.00
			HD-LD	-0.01	-3.23%	0.81	-0.05	0.02	-0.05	0.03
CASEVAC by FT (minutes)*	14.56 (2, 168)	< 0.01*	LD-C	0.26	20.12%	< 0.01*	0.15	0.38	0.13	0.40
			HD-C	0.33	24.88%	< 0.01*	0.20	0.45	0.18	0.47
			HD-LD	0.06	3.97%	0.73	-0.08	0.20	-0.11	0.23
Prepare Fighting Positions (lbs.)	0.18 (2, 169)	0.84	LD-C	111.33	2.16%	0.82	-211.46	434.13	-276.03	498.70
			HD-C	38.71	0.75%	0.98	-312.26	389.68	-382.47	459.89
			HD-LD	-72.62	-1.38%	0.95	-475.67	330.42	-556.29	411.04
1-km Hike (minutes)*	8.634 (2, 54)	0.58†	LD-C	0.68	7.65%	< 0.01†	0.38†	0.78†	0.35†	0.85†
			HD-C	0.50	5.66%	<0.01†	0.32†	0.67†	0.27†	0.73†
			HD-LD	-0.18	-1.85%	0.73†	-0.25†	0.12†	-0.30†	0.17†
Negotiate Obstacle (minutes)*	11.51 (2, 52)	< 0.01*	LD-C	0.03	1.88%	0.92	-0.12	0.18	-0.15	0.21
			HD-C	0.38	21.80%	< 0.01*	0.23	0.53	0.20	0.56
			HD-LD	0.35	19.55%	< 0.01*	0.19	0.50	0.16	0.53

*Indicates that there is a statistically significant difference in a two-sided hypothesis test between integration levels according to ANOVA and those according to a non-parametric equivalent test.

**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.

†Results presented are from a Kruskal-Wallis and Mann-Whitney non-parametric tests due to non-normality.

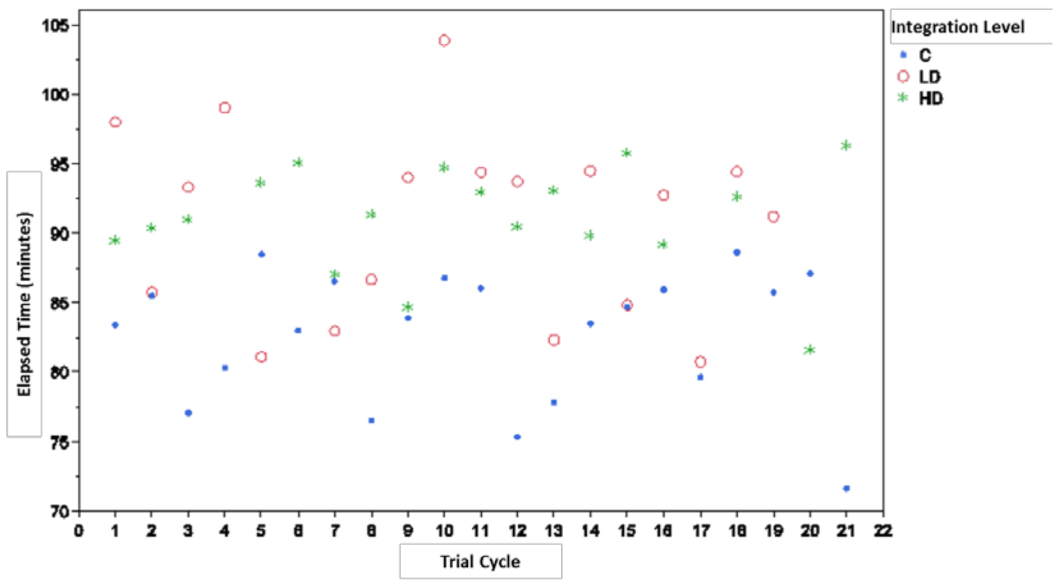
F.5.2.1 7-km Hike

F.5.2.1.1 7-km Hike Overview

This experimental task assessed a squad of 12 Marines moving 7.20 km while each Marine carried an approach load and an individual weapon. The individual weapons included the M-4, the M-16A4 with M203 attachment, and the M27, resulting in a cumulative load of 96 – 101 lb per Marine. The recorded time for this task started when the squad departed the Range 107 start point and stopped when the squad arrived at the Range 110 stop point. Each squad moved as fast as the slowest person and could take as many breaks as necessary.

Figure F-1 displays all PI 7-km hike data. All data on the scatter plot are valid for analysis.

Figure F-1. 7-km Hike



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in p-values of 0.06 for the C group, 0.24 for the LD group, and 0.28 for the HD group.

The C group had a mean time of 82.75 minutes. This is statistically significantly faster than the LD mean time of 90.76 minutes and the HD mean time of 91.05 minutes. These differences result in 9.68% (8.01-minute) and 10.03% (8.30-minute) degradations in time for the LD and HD groups, respectively. Additionally, the LD group had greater variability as shown by the 1.96-minute increase in standard deviation (SD), and the HD had less variability as shown by the 0.90-minute decrease in SD (4.75 minutes for the C group, 6.71 minutes for the LD group, and 3.85 minutes for the HD group). The LD group was faster than the HD group. There was a 0.32% (0.29-minute) degradation in hike time from the LD to HD group, but this difference is not statistically significant. See Table F-2 and Table F-3 for detailed analytical results.

F.5.2.1.2 7-km Hike Contextual Comments

F.5.2.1.2.1 USMC Hike Standards

The 7-km hike is a task that the following MOSs completed during the experiment: 0311, 0331, 0341, 0351 and 0352, PI, Provisional Machine Gunners, and Combat Engineers. There are varying standards to which we can compare this result. The following sections define those standards, as well as the one we chose as a comparison.

The Infantry T&R Manual (30 Aug 2013) identifies minimum standards that Marines must be able to perform in combat, to include standards for tactical marches. In Chapter 8 of the Infantry T&R Manual in the MOS 0300 Individual Events section, task "0300-COND-1001: March under an approach load" is applicable to all 03XX MOSs, ranks Pvt – LtCol. The condition and standard established by this task is: "Given an assignment as a member of a squad, individual weapon, and an approach load, complete a 20 kilometer march in under 5 hours." The march pace required by this standard is 4 kilometers per hour (km/h). In Chapter 9 of the Infantry T&R Manual in the MOS 0302/0369 Individual Events section, tasks "0302-OPS-2001: Lead an approach march" and "0369-OPS-2501: Lead an approach march" are applicable to MOS 0302 and 0369, ranks SSgt – MGySgt and 2ndLt – LtCol. The condition and standard established by these tasks is: "Given a mission, time constraints, an approach march load, organic weapons, and a route, move 24.8 miles (40 km) in a time limit of 8 hours." The march pace required by this standard is 5 km/h. Appendix E of the Infantry T&R Manual (Load Terms and Definitions) states: "The approach march load will be such that the average infantry Marine will be able to conduct a 20 mile hike in 8-hours with the reasonable expectation of maintaining 90% combat effectiveness." The march pace required by this definition is 4.02 km/h.

Chapter 3 of Fleet Marine Force Reference Publication (FMFRP) 3-02A, Marine Physical Readiness Training for Combat (16 Jun 2004), states: "The normal pace is 30 inches. A pace of 30 inches and a cadence of 106 steps per minute result in a speed of 4.8 km/h or 3 mi/h and a rate of 4 km/h or 2.5 mi/h if a 10-minute rest halt per hour is taken." Common Infantry practice today is to hike for 50 minutes and take a 10-minute break, while maintaining an overall pace of 4 km/h (resulting in a hiking pace of 4.8 km/h).

Driven by the need to pick an evaluative reference standard, this report follows the T&R Manual's intent to establish minimum standards and uses the 4-km/h march pace for a 20-km march established by task 0300-COND-1001 and supported by the definition of an approach load. Further, while an established reference standard is required to anchor observed performance to the T&R program, more important information is provided by performance differences observed between gender-integrated and non-gender-integrated units.

F.5.2.1.2.2 PI 7-km Hike Pace

This result is relevant to both the training and combat environments as it will take integrated squads more time to conduct foot marches. Per the tactical march standards referenced above, the Marine Corps standard of hiking is 4.0 km/h. All integration levels surpassed this standard. The C group's average pace was 5.22 km/h, while the LD and HD groups' average paces were 4.76 km/h and 4.74 km/h, respectively. Assuming that these paces could be maintained for a 20-km task (an optimistic assumption that does not account for any further degradation of performance), the HD and LD groups would be approximately 23 minutes behind the C group at the finish. Finally, on any given day (under the same environmental conditions), the C group was faster than the HD group 94.4% of the time (17 of 18 trial cycles). The C group was faster than the LD group 88.9% of the time (16 of 18 trial cycles). The HD group was faster than the LD group 60% of the time (9 of 15 trial cycles).

F.5.2.1.3 7-km Hike Additional Insights

The difference in the 7-km hike time and speed was statistically significant, and the size of the difference has some intuitive significance. Common Infantry practice today is to hike for 50 minutes and take a 10-minute break, while maintaining an overall pace of 4 km/h (resulting in a hiking pace of 4.8 km/h). Given a 4.8 km/h rate of march, the LD and HD failed to maintain that pace.

A purely objective evaluation of 0.45 – 0.46 km/h decrease in pace is elusive, but Marine Corps Doctrinal Publications (MCDPs) consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to “Being Faster” and states, “Physical speed, moving more miles per hour, is a powerful weapon in itself.” MCDP-6 Command and Control also speaks to speed relative to the enemy and states, “The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results.” The difference observed between the C and HD or LD groups was over 8 minutes. For an indication of performance when the load is increased, see the 0331, 0341, and 035X 7-km hike data and contextual comments.

F.5.2.1.4 7-km Hike Subjective Comments

The subjective leadership comments reveal that the majority of the Marines responsible for falling behind and slowing the movement within integrated groups were female Marines. There were 32 comments about males slowing the movement within C groups. For slowing movement within the LD group, 13 comments related to females and 8 comments related to males. For slowing movement within the HD groups, 27 comments related to females, while only 2 comments related to males. It was observed that as the integration level increased, the more frequently female Marines were responsible for a degradation in performance.

For additional subjective comments relating to this task, see the Appendix.

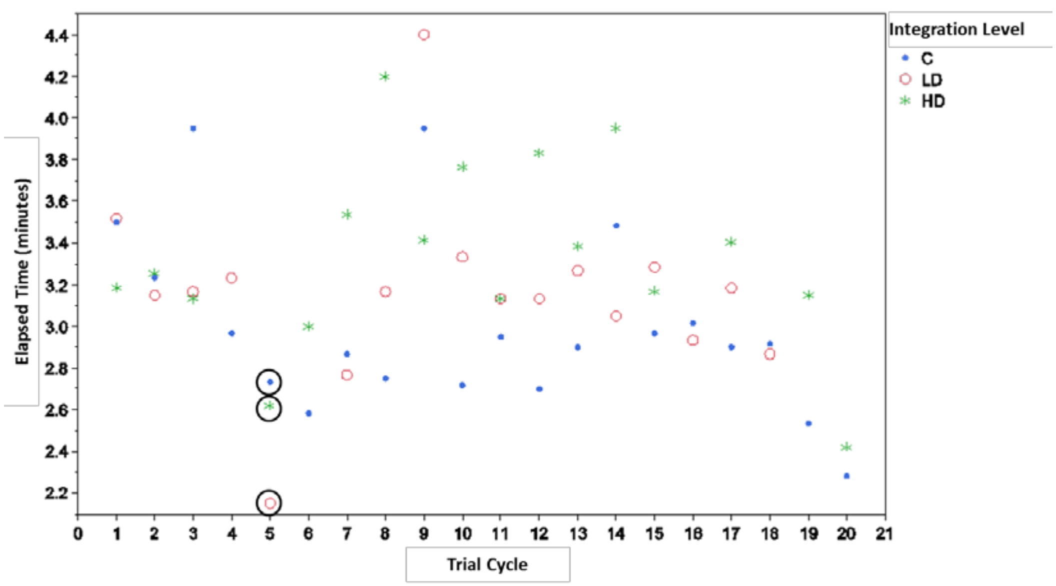
F.5.2.2 Movement to LOA

F.5.2.2.1 Movement to LOA Overview

This experimental task assessed a squad moving approximately 125 meters to an LOA immediately after conducting a fire-and-movement assault. The recorded time started when the last target went down after the assault and stopped when the last member of the squad was prone at the LOA.

Figure F-2 displays all PI Movement to LOA time data. On trial cycle 5, data for the C, LD, and HD groups were removed from the analysis due to data collection errors; data omitted from the analysis are circled in black. With the exception of the trial cycle 5 data points, all data on the scatter plot are valid for analysis.

Figure F-2. Movement to LOA



The data for the C and HD groups are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in p-values of 0.06 and 0.51, respectively, but they are not normally distributed for the LD group, which had a p-value of < 0.01.

The C group had a mean of 3.01 minutes; the LD, 3.22 minutes; and the HD, 3.37 minutes. These differences result in 7.15% (0.22-minute) and 12.00% (0.36-minute) degradations in time for the LD and HD groups, respectively. The LD group was faster than the HD group. There was a 4.52% (0.15-minute) degradation in movement time from the LD to HD group.

The difference in movement to LOA time was not statistically significant between the LD and HD groups in a two-sided Mann-Whitney Test but was statistically significant in a one-sided Mann-Whitney Test. The difference in movement to LOA time was statistically significant between the LD and C groups as well as the HD and C groups in a Mann-Whitney Test. The difference in movement to LOA time was not statistically significant for the LD and C groups in a Tukey Test. Because of a lack of normality, we recommend using the Mann-Whitney test results (reported in Table F-3).

The near significance (between the LD and HD groups) of the non-parametric results suggests that further study of this task is warranted. See Table F-2 and Table F-3 for detailed analytical results.

F.5.2.2.2 Movement to LOA Contextual Comments

The ability to close with the objective after having conducted a live-fire attack is a crucial aspect of maintaining the momentum during offensive operations. In this case, the LD squads were exposed to the enemy 13.20 seconds (7.15%) longer and HD squads were exposed to the enemy 21.60 seconds (12.00%) longer than the C group squads. This longer exposure means that these squads have that much less time to prepare for a counterattack or conduct follow-on actions. On any given day (under the same environmental conditions), the C group was faster than the LD group 58.8% of the time (10 out of 17 trial cycles) and faster than the HD group 76.5% of the time (13 out of 17 trial cycles). On any given day, the LD group was faster than the HD group 69.2% of the time (9 out of 13 trial cycles), while they finished in the same amount of time for one trial.

F.5.2.2.3 Movement to LOA Additional Insights

A purely objective evaluation of 13.20 or 21.60 seconds is elusive, while the previous citations emphasizing the importance of speed still apply (see Section F.5.2.1.3). However, the performance decrement in this task is approximately 7% or 12% (depending on type of integrated squad) and may possess some practical significance on the battlefield that would reduce the survivability of an integrated squad. Considering the 12 – 15 rounds-per-minute sustained rate of fire for both the M4 and AK-47 rifles, a single enemy fighter would have the opportunity of two to three more well-aimed shots on Marines in an integrated squad moving to the limit of advance. Similarly, the delay degrades our own pursuit of the enemy by fire, denying our slowest Marines two to three well-aimed engagements on the enemy from a fixed position at the LOA. The resultant trade in casualty exchange could be significant.

F.5.2.2.4 Movement to LOA Subjective Comments

The subjective leadership comments reveal that the majority of the Marines displaying fatigue within integrated groups were female Marines. There were 54 comments about males slowing the movement within C groups. Of the 62 comments related to Marines showing fatigue within the LD group, 35 related to females and 27 related to males. Of the 61 comments related to Marines showing fatigue within the HD group, 44 related to females, while only 17 related to males. We observe that as the integration level increased, female Marines were increasingly responsible for degradation in performance.

For additional subjective comments relating to this task, see the Appendix.

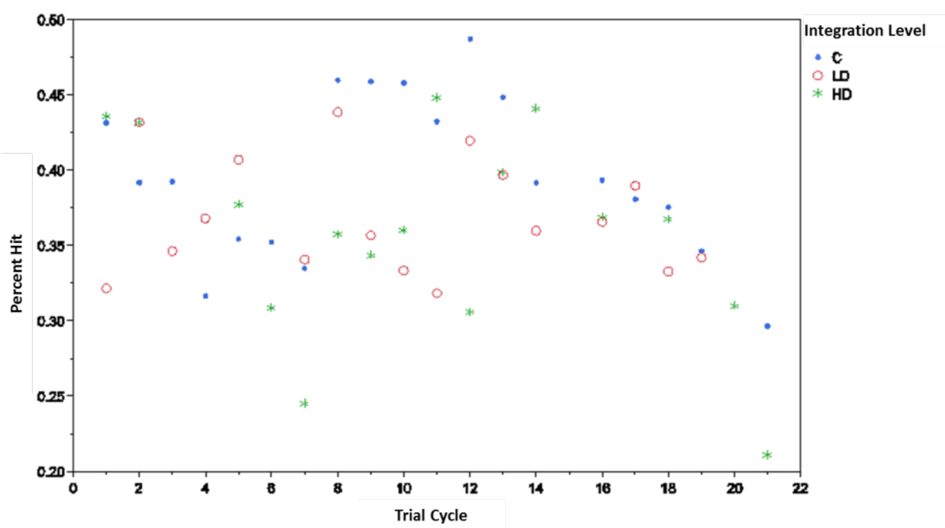
F.5.2.3 Attack & Counterattack (C-Atk) Percent Hit

F.5.2.3.1 Attack & C-Atk Overview

This experimental task assessed the accuracy of a rifle squad engaging 12 GTSs during a 300-meter fire-and-movement assault. Each GTS captured the precise location of a round that impacted a target silhouette using a location of hit and miss (LOMAH) sensor. The GTSs were equally spread out over the downrange area at varying distances and were exposed according to pre-determined parameters that were consistent for every squad attack. The accuracy was determined by dividing the number of hits on each target by the total amount of ammunition expended by each squad.

Figure F-3 displays all Attack & C-Atk Percent Hit data. All data on the scatter plot are valid for analysis.

Figure F-3. Attack & C-Atk Percent Hits



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.65 for the C group, 0.20 for the LD group, and 0.36 for the HD group.

The C group had an average percent hit of 39%. This difference is higher (but not statistically significantly) for both the LD 37% and the HD 36%. These differences result in degradations of 6.62% (3 percentage points) and 9.63% (4 percentage points) in accuracy for the LD and HD groups, respectively. The HD group was 3.23% (1 percentage point) less accurate (but not statistically significantly) than the LD group. See Table F-2 and Table F-3 for detailed analytical results.

F.5.2.3.2 Attack & C-Atk Contextual Comments

F.5.2.3.2.1 Attack & C-Atk Contextual Comments Overview

In combat operations, accuracy is highly desirable in destroying or effectively suppressing an enemy position. In the execution of this task, the LD group sustained an average of 6.62% decrement in the percentage of hits for the number of shots taken, and the HD group sustained an average of 9.63% decrement in the percentage of hits for the number of shots taken. Operationally, this equates to a degradation of accuracy and/or an increase in ammunition expenditure. On any given day (under the same environmental conditions), the C group was more accurate than the LD group 70.6% of the time (12 out of 17 trial cycles) and more accurate than the HD group 66.7% of the time (10 out of 15 trial cycles). On any given day, the HD group was more accurate than the LD group 58.3% of the time (7 out of 12 trial cycles).

F.5.2.3.2.2 Analysis by Weapon System and Gender

The use of a Weapons Player Pack (WPP) on each weapon system allowed data to be captured on each shot taken by a Marine during the conduct of the attack and c-atk. When synchronized with the data obtained from the GTS, a shooter-to-shot correlation was possible. For the analysis of shot accuracy, accuracy percentages by gender and weapon were analyzed that measured percent hits and percent hits and near misses where a hit indicates that a round hit the target silhouette and a near miss indicates that the LOMAH sensor detected a round within a 1.5-meter detection arc.

Since the accuracy results by gender were collected, a t-test was used for this analysis. Table F-4 displays the shot accuracy results by gender and weapon system.

Table F-4. Shot Accuracy by Gender and Weapon System[1]

	Probability of Hit				Probability of Hit & Near Miss			
Weapon	**F**	**M**	**% Difference**	**2-sided p-value**	**F**	**M**	**% Difference**	**2-sided p-value**
M4 0311*	.28	.42	33%	<0.01	.73	.75	3%	0.32
M27 0311**	.25	.43	42%	<0.01	.58	.69	16%	.014
M16A4/M203 0311**	.15	.28	46%	<0.01	.50	.67	25%	<0.01
M4 PI**	.37	.44	16%	.02	.73	.79	8%	0.03
M27 PI	.37	.38	3%	.7571	.66	.69	4%	.37
M16A4/M203 PI**	.15	.26	42%	<0.01	.59	.70	16%	0.05

*Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the percent hits values for the gender.

[1] The M16A4/M203 shot accuracy data are only for 5.56mm ammunition shot from the weapon and do not include the 40mm practice round accuracy.

**Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the percent hits and the percent hits and near miss values for the gender.

The shot accuracy results by gender and weapon reveal that there is a statistically significant difference between genders of the Marines carrying M4 and M16A4/M203 within the PI squads. The overall accuracy declined and the percent difference increased as the weight of the weapon system increased. The M4 was the lightest weapon and yielded the best results, while the M16A4/M203 was the heaviest weapon and yielded the worst results.

One might think that experience level of the Marines (male vs. female) influenced the results. However, when compared to the 0311 results, the PI results disprove this conjecture, as both the male and female PI Marines came to the unit without being trained in Infantry tactics at the fleet learning center and their performance mirrors the results of the 0311 Marines. With the exception of the difference in M27 accuracy, which was not statistically significant, the PI accuracy and percent difference increased as the weapon got heavier. Furthermore, a close look at the unit training plan, methodology, leadership, and ammunition expended will show that the 0311 and PI Marines in the GCEITF were well prepared for this assessment, minimizing the impact of past experience. Therefore, one would conclude that given the same type and amount of training, female Marines would have less lethality on the battlefield than male Marines.

F.5.2.3.3 Attack & C-Atk Additional Insights

To explore the operational effect of a 2-percentage point difference in percentage of hits by integration level (see Table F-2), we conducted a Monte Carlo Simulation using the Lanchester Square Law as a model of a tactical-level engagement. For more information on the Lanchester Square Law, see the Methodology Annex. We chose the Lanchester Square Law for rudimentary analysis using a well-known combat model to explore the potential effects of the differences discovered between integration levels.

We make several important assumptions for this Lanchester model in terms of parameter selection. We chose parameters that model a worst-case scenario for a squad engagement in which the squad faces an equivalent adversary in terms of capability. The first assumption is that the force sizes are equal, which is a 12-on-12 fight. The second assumption is the relative rate of fire between forces, and we assume the ratio of friendly to enemy rate of fire to be equal to one. Finally, with respect to accuracy, the enemy was assumed to be as good as the best experimental group, which was the C group. Thus, the enemy probability of hit was set at 0.39.

Two hundred thousand simulated engagements were run for both the C and LD groups. A force won an engagement when its opposition was eliminated. By construct, the C group, with a probability of hit equal to the enemy’s at 0.39, won 50% of the engagements with an average of three Marines remaining. This, the even-fight case, is useful only to assess the effect of the LD group's decrement in probability of hit. In another two hundred thousand simulated engagements, the LD group, with a probability of hit equal to 0.37, won 49.2% of the engagements with three Marines remaining. Based on these results, we would expect the LD group to win 0.8% fewer tactical-level engagements. One may expect a similar comparison between the C and HD groups to fall between the above result and the result obtained for the 0311 simulation.

F.5.2.3.4 Attack & C-Atk Subjective Comments

For subjective comments relating to this task, see the Appendix.

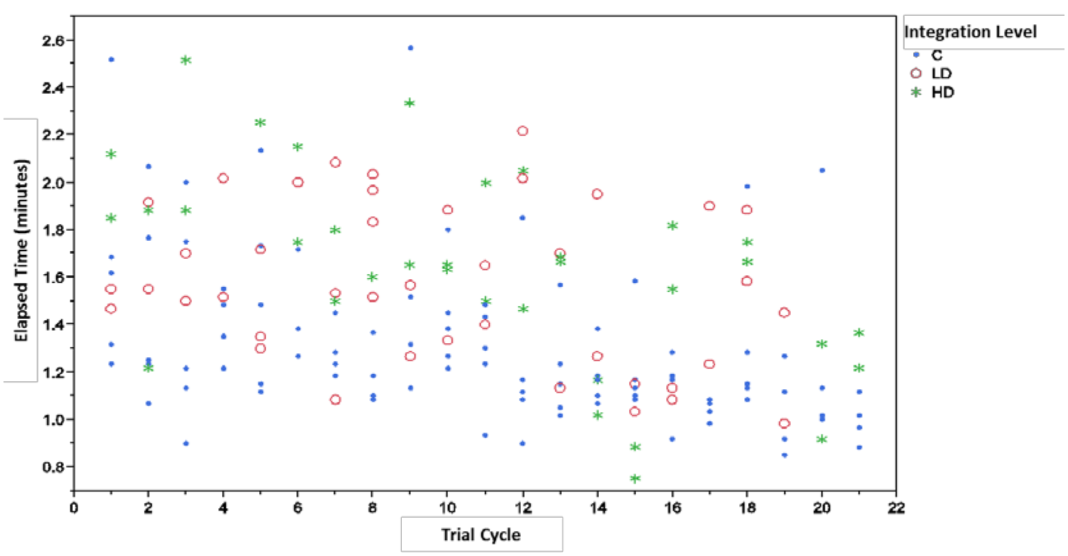
F.5.2.4 CASEVAC by FT

F.5.2.4.1 CASEVAC by FT Overview

This experimental task assessed a fireteam's ability to move a 220-lb dummy a distance of 100 meters to a CCP while wearing a fighting load and individual weapon. Marines conducted this task at the conclusion of the squad attack/counterattack. At the discretion of each fireteam, Marines used a two-Marine, a three-Marine, or a four-Marine carry in order to move the casualty. The recorded time started when a member of the fireteam touched the dummy, and it stopped when the dummy and all members of the fireteam arrived at the CCP.

Figure F-4 displays CASEVAC by FT data. All data on the scatter plot are valid for analysis.

Figure F-4. CASEVAC by FT



For this task, sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

The C group had a mean time of 1.31 minutes. This is statistically significantly faster than the LD mean time of 1.57 minutes and the HD mean time of 1.63 minutes. These differences result in a 20.12% (0.26-minute) and 24.88% (0.33-minute) degradation in times for the LD and HD groups, respectively. The LD group was faster than the HD group. There was a 3.97% (0.06-minute) degradation in hike time from the LD to the HD group, but this difference is not statistically significant. See Table F-2 and Table F-3 for detailed analytical results.

F.5.2.4.2 CASEVAC Contextual Comments

The implications of this task contain relevance to both the training and combat environment as survival is dependent on expeditious movement of the casualty to higher levels of care. The data demonstrate that LD groups took 15.6 seconds longer (a 20.12% degradation in performance) and the HD groups took 19.2 seconds longer (a 24.88% degradation in performance) to accomplish the same 100-meter casualty movement.

F.5.2.4.3 CASEVAC Additional Insights

While the “Golden Hour” is a common medical planning construct for C2 and logistical support, medical literature supports a “Platinum Ten” philosophy of first response. The U.S. National Library of Medicine references a French article that states, “On the battlefield, the majority of casualties die within ten minutes of the trauma” (*Wounded in Action: The Platinum Ten Minutes and the Golden Hour*, Daban). The fundamental principle is that a patient needs to be correctly triaged and moved to medical care as quickly as possible. Any time degradation will reduce the probability of survival.

F.5.2.4.4 CASEVAC Subjective Comments

For subjective comments relating to this task, see the Appendix.

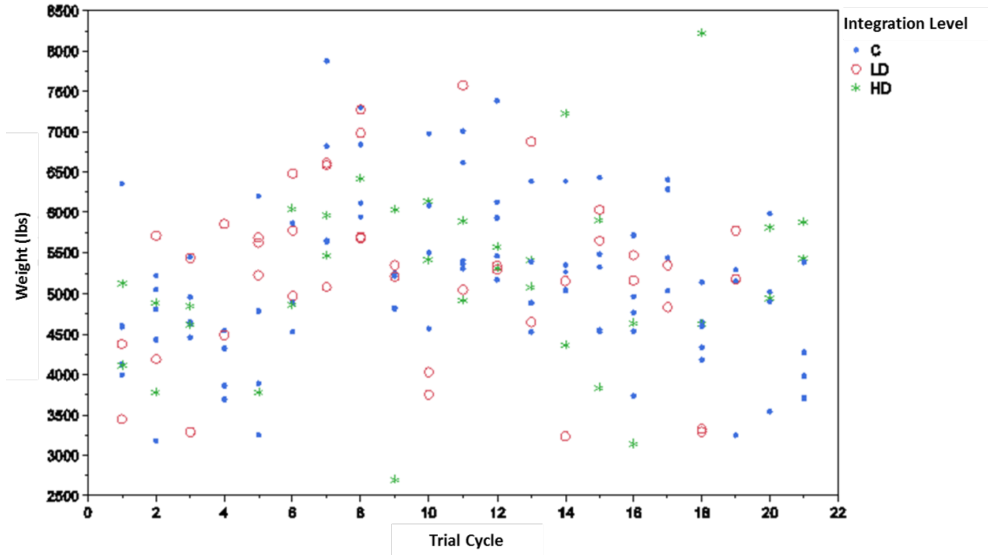
F.5.2.5 Prepare Fighting Positions

F.5.2.5.1 Prepare Fighting Position Overview

This experimental task assessed the amount of earth moved by a fireteam in a two-hour period. Each fireteam was required to dig 2 two-man fighting holes while wearing a fighting load and maintaining 50% security (only two Marines were digging at any given time). All the earth was scooped into buckets and weighed by a non-assessed Marine. The recorded time began 10 minutes after finishing the 7-km hike, and time stopped 2 hours after beginning to dig.

Figure F-5 displays all PI Prepare Fighting Positions data. Five invalid data points were removed from both the plot and the analysis. All data on the scatter plot are valid for analysis.

Figure F-5. Prepare Fighting Positions



For this task, sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

The C group moved a mean weight of 5,146.13 lb. This difference is higher (but not statistically significantly) for both the LD mean weight of 5,257.46 lb and the HD mean weight of 5184.84 lb. These differences result in 2.16% (111.33-lb) and 0.75% (38.71-lb) degradations in weight moved for the LD and HD groups, respectively. The HD group moved 1.38% (72.62 lb) less (but not statistically significantly) in weight than the LD group. See Table F-2 and Table F-3 for detailed analytical results.

F.5.2.5.2 Prepare Fighting Positions Contextual Comments

Marine Corps Warfighting Publication (MCWP) 3-11.2 Marine Rifle Squad specifies the dimensions of a two-person fighting hole: 6 ft long X 2 ft wide X 5 ft deep (assuming occupants who are 5'6" – 6' tall). The volume of this fighting hole is 60 ft^3. Two such holes are required for a fireteam, and together they have a combined volume of 120 ft^3. In Engineering Manual 1110-1-1905 (Engineering and Design: Bearing Capacity of Soils), the Army Corps of Engineers estimates that collapsible soil (of the type found at MCAGCC Twentynine Palms) has an average density of 85 lb/ft^3. Thus, a total of 10,200 lb of dirt would have to be moved in order to construct two fighting positions to the MCWP 3-11.2 standard.

The C group moved earth at a rate of 2,573.07 lb/hr. At this rate, the C group would have completed digging its fighting holes in 3 hours and 58 minutes. The LD group moved earth at a rate of 2,628.73 lb/hr. At this rate, the LD group would have completed digging its fighting holes in 3 hours and 53 minutes. The HD group moved earth at a rate of 2,592.42 lb/hr. At this rate, the HD group would have completed digging its fighting holes in 3 hours and 56 minutes. Thus, the three groups each would have completed digging their fighting holes within 5 minutes of each other.

F.5.2.5.3 Prepare Fighting Positions Additional Insights

Motivation and perspective provided by the unit leadership was a contributing factor to the overall morale and work ethic of the fireteams and each individual Marine. Leadership and teamwork were not explicitly measured in this tactical task but are thought to have a larger impact than anticipated. The PI squads performed nearly 20% better than the 0311 control and integrated squads at this same task. Of note, the PI Platoon Leadership was observed constantly engaging with their Marines, walking around and inspecting the holes, helping weigh buckets of sand, and providing historical relevance after the assessment. This engaged style of leadership is thought to have resulted in greater workload output.

F.5.2.5.4 Prepare Fighting Position Subjective Comments

For subjective comments relating to this task, see the Appendix.

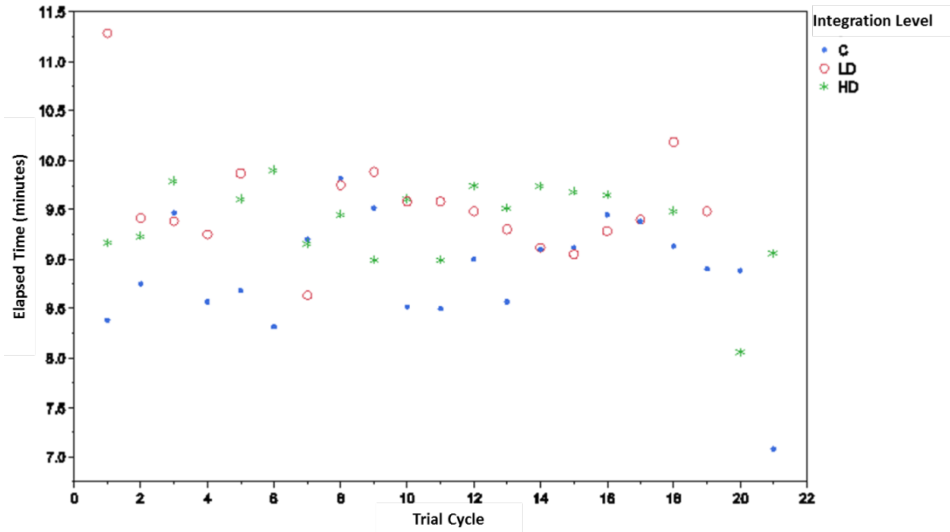
F.5.2.6 1-km Hike

F.5.2.6.1 1-km Hike Overview

This experimental task assessed a squad of 12 Marines moving approximately 1 km while each Marine carried an assault load and individual weapon. Each squad moved as quickly as possible to reinforce a notional friendly squad pinned down by enemy fire. The recorded time started when the squad departed the assembly area on Range 107 and ended upon reaching a designated attack position just prior to the wall/obstacle.

Figure F-6 displays all PI 1-km hike data. All data on the scatter plot are valid for analysis.

Figure F-6. 1-km Hike



The data are not normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of < 0.01 for the C, LD and HD groups. Because of a lack of normality, we recommend using the Mann-Whitney test results (reported in Table F-3). We proceed with presenting ANOVA results because they were confirmed by Mann-Whitney Tests.

The C group had a mean time of 8.87 minutes. This time is statistically significantly faster than the LD mean time of 9.55 minutes and the HD mean time of 9.37 minutes. These differences result in 7.65% (0.68-minute) and 5.66% (0.50-minute) degradations in time for the LD and HD groups, respectively. The HD group was faster than the LD group. There was a 1.85% (0.18-minute) degradation in hike time from the HD to LD group, but this difference is not statistically significant. See Table F-2 and Table F-3 for detailed analytical results.

F.5.2.6.2 1-km Hike Contextual Comments

The Infantry T&R Manual states that “the maximum assault load weight will be such that an average infantry Marine will be able to conduct combat operations indefinitely with minimal degradation in combat effectiveness.” While moving 1 km with the assault load, the integrated squad type each moved at least 5.66% slower on average than the C group, making them 30 seconds less responsive reinforcing an adjacent unit. On any given day (under the same environmental conditions), the C group was faster than the LD group 72.2% of the time (13 out of 18 trial cycles), and faster than the HD group 77.8% of the time (14 out of 18 trial cycles). On any given day, the LD group was faster than the HD group 53.3% of the time (8 out of 15 trial cycles).

F.5.2.6.3 1-km Hike Additional Insights

A purely objective evaluation of 30 seconds is elusive but may possess some practical significance on the battlefield that would reduce the survivability of a unit waiting on an integrated squad. Considering the 12-15 rounds-per-minute sustained rate of fire for both the M4 and AK-47 rifles, a single enemy fighter would have the opportunity to take six to seven more well-aimed shots on Marines while waiting for reinforcements from an integrated squad. A fireteam of enemy fighters would have time to call in indirect 82mm mortar fire or maneuver during this time delay. The resultant trade in casualty exchange could be significant.

F.5.2.6.4 1-km Hike Subjective Comments

For subjective comments relating to this task, see the Appendix.

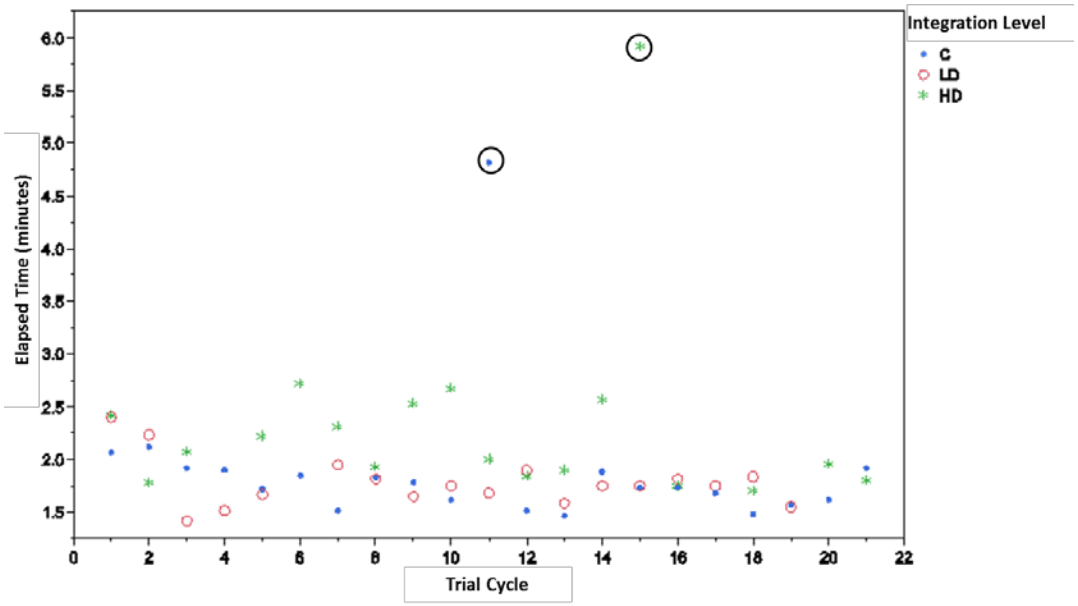
F.5.2.7 Negotiate Obstacle

F.5.2.7.1 Negotiate Obstacle Overview

This experimental task assessed the time it took to get a squad of 12 Marines and all equipment over an 8-foot-tall ISO Container (wall/obstacle). The squad began this task as soon as it finished the 1-km movement. The recorded time started when the first Marine touched the wall and stopped when the last Marine touched both feet down on the other side.

Figure F-7 displays all PI Negotiate Obstacle data. On trial cycle 11, the C group data point was removed for analysis as it represents a data outlier, and on trial cycle 15, the HD data point was removed for analysis as it represents a data outlier. With the exception of these points, all data on the scatter plot are valid for analysis.

Figure F-7. Negotiate Obstacle



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.53 for the C group, 0.07 for the LD group, and 0.07 for the HD group.

The C group had a mean time of 1.75 minutes. This is faster (but not statistically significantly) than the LD mean time of 1.78 minutes. The C group mean time is statistically significantly faster than the HD mean time of 2.13 minutes. These differences result in 1.88% (0.03-minute) and 21.80% (0.38-minute) degradations in time for the LD and HD groups, respectively. The LD mean time was statistically significantly faster than the HD group. There was a 19.55% (0.35-minute) degradation in hike time from the LD to the HD group. See Table F-2 and Table F-3 for detailed analytical results.

F.5.2.7.2 Negotiate Obstacle Contextual Comments

Sun Tzu stated, “Speed is the essence of war. Take advantage of the enemy’s unpreparedness; travel by unexpected routes and strike him where he has taken no precautions” (MCDP-1, p. 69). Speed and surprise are crucial to success. Furthermore, the enemy is reliably expected to cover obstacles with fires. When possible, obstacles are to be avoided. When not possible, they must be negotiated quickly. While no purely objective standard can be set for the negotiation of the obstacle presented in this task, any decrement in speed translates into increased exposure to enemy fires and greater risk for friendly casualties. On average, the LD groups were 2 seconds slower than the C groups and the HD groups were 23 seconds slower in performing the same task. On any given day (under the same environmental conditions), the C group was faster than the LD group 58.8% of the time (10 out of 17 trial cycles) and faster than the HD group 86.7% of the time (13 out of 15 trial cycles).

On any given day, the LD group was faster than the HD group 76.9% of the time (10 out of 13 trial cycles).

F.5.2.7.3 Negotiate Obstacle Additional Insights

Of note, a large discrepancy may be observed when comparing the PI obstacle time with the 0311 obstacle time. This difference can be explained by the different techniques used by each group. The 0311 squads provided their own security and only used one launch-point up the wall, while the PI squads did not establish security and used three launch-points to get over the wall. Each group (0311 and PI) was consistent within itself, but the two groups should not be compared with each other.

F.5.2.7.4 Negotiate Obstacle Subjective Comments

According to the subjective leadership observations, 25 recorded comments about Marines requiring extra assistance to complete this task applied to female Marines, while only 2 applied to male Marines who were members of LD or HD squads. There were also four comments citing a male Marine who was a member of an LD or HD squad as compensating for another Marine, and no such comments for the female squad members.

For additional subjective comments for this task, see the Appendix.

F.6 Statistical Modeling Results

F.6.1 Statistical Modeling Results Overview

The previous section discussed results only as they pertain to differences due to integration level alone. The goal of statistical modeling, as applied here, is to estimate, simultaneously, the effect of gender integration levels and other relevant variables on squad performance. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For the same seven selected tasks described in the previous section, this section first presents an overview of the analysis and results, and then presents the modeling results for each of the tasks.

For each task, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates that an increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time but not a desired outcome for the percent hits or for pounds of earth moved. The results indicate where certain patch numbers are significant for a given variable. The experiment tracked Marines within the rifle squad by a patch number that associated their random position within the squad to a specific billet. Table F-5 displays the patch numbers and associated billet titles for the rifle squad.

Table F-5. Patch Numbers and Billet Titles for the Rifle Squad

Patch Number	Billet Title
1	FT 1 Fireteam Leader
2	FT 1 Automatic Rifleman
3	FT 1 Grenadier
4	FT 1 Rifleman
5	FT 2 Fireteam Leader
6	FT 2 Automatic Rifleman
7	FT 2 Grenadier
8	FT 2 Rifleman
9	FT 3 Fireteam Leader
10	FT 3 Automatic Rifleman
11	FT 3 Grenadier
12	FT 3 Rifleman

F.6.2 PI Selected Tasks Method of Analysis

A mixed effects model does not work for the PI data set because of the low number of volunteers and trials. Therefore, we model the PI selected tasks using ordinary least squares regression.

For the majority of each of the primary metrics for the rifle squad, there are only 60 observations for each result. Because there are 12 Marines in a rifle squad and so few results, the regression model does not have a sufficient number of degrees of freedom to create a model using all types of personnel data for each squad member for each result. Thus, each variable combined with integration level is modeled separately. For example, age for each member of the rifle squad (12 variables) and integration level are modeled with the result (response time, percentage hits, or pounds of earth moved) as the response variable. Where possible, a backward stepwise regression, using AIC, determined which variables are optimal in the model. If there were missing values, backward stepwise could not run and significant variables are reported based on p-values from the overall model.

F.6.3 PI Selected Tasks Overall Modeling Results

There are no personnel data variables that are both statistically significant and have a practical impact on the model. Each time personnel data variables are statistically significant in a model, their effects are practically negligible, conflicting, and/or incomplete for the squad — i.e., there are no tasks for which a variable is significant for all, or even most, members of the rifle squad.

Integration level, however, consistently appears as statistically significant in each of the tasks (except prepare defensive position), and its effect is clear, causal, and practical. Therefore, integration level is the best variable for describing performance for each of these tasks. Refer to Section F.5 to see the ANOVA summary for each of the PI tasks mentioned below.

F.6.3.1 7-km Hike

We model elapsed time for the 7-km hike as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

Both the HD and LD integration levels are significant and positively correlated with the response for the models that include the following personnel variables:

- All.

Both the HD and LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following variables are significant in their respective models and are positively correlated with the 7-km hike time:

- Height of patch 5,
- Weight of patch 5,
- AFQT score of patch 1,
- GT score of patch 1, 2, and 11,
- CFT MTC of patches 5 and 8,
- CFT MANUF of patches 5 and 8,
- PFT 3-mile run of patch 6,
- Rifle score of patches 3, 4, and 10.

The following variables are significant in their respective models and are negatively correlated with the 7-km hike time:

- Age of patches 4 and 8,
- Height of patch 8,
- Weight of patch 6,
- AFQT score of patch 10,
- CFT MTC of patches 3, 4, and 11,
- CFT MANUF time of patches 2, 3, 4, 6, 11, and 12,
- PFT crunches of patches 2, 3, 6, 7, and 9,
- Rifle score of patches 1, 5, and 8.

The following personnel variables have no significant variables in their respective models:

- None.

There are no patterns for any personnel variables for the 7-km hike. See Section F.5.2.1 for the ANOVA summary of this task.

F.6.3.2 Movement to LOA

We model elapsed time for the movement to LOA as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following variables:

- Age,
- Weight,
- AFQT score,
- GT score,
- PFT crunches,
- PFT 3-mile run.

The LD integration level is significant and positively correlated with the response for the models that include the following variables:

- Age,
- AFQT score,
- GT score,
- PFT 3-mile run.

The HD and LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following variables are significant in their respective models and are positively correlated with the movement to the LOA time:

- Age of patches 5 and 7;
- Height of patch 4,
- Weight of patch 4,
- GT score of patches 7 and 11,
- CFT MTC of patches 5, 6, 9, 10, and 12,
- PFT 3-mile run of patch 3,
- Rifle score of patches 2 and 7.

The following variables are significant in their respective models and are negatively correlated with the movement to the LOA time:

- Age of patches 3, 4, and 8,
- Height of patches 2, 3, 6, 9, and 12
- Weight of patches 2, 3, and 12,
- CFT MTC of patches 3, 4, and 11,
- CFT MANUF time of patch 4,
- PFT crunches of patches 5 and 6
- Rifle score of patches 4 and 8.

The following personnel variables have no significant variables in their respective models:

- None.

There are no patterns for any personnel variables for the movement to LOA time. See Section F.5.2.2 for the ANOVA summary of this task.

F.6.3.3 Attack & C-Atk Percent Hits

We model the attack & c-atk percent hits as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD and LD integration levels are significant and positively correlated with the response for the models that include the following variables:

- None.

The HD integration level is significant and negatively correlated with the response for the models that include the following variables:

- Age,
- AFQT score,
- GT score,
- CFT MTC,
- CFT MANUF.

The LD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following variables are significant in their respective models and are positively correlated with the percent of hits for the attack and counterattack:

- Age of patches 3, 6, 7, 8, and 12,
- Height of patch 12,
- Weight of patches 7 and 12,
- AFQT score of patch 2, 8, and 9,
- GT score of patch 9,
- CFT MTC of patch 9,
- CFT MANUF of patch 9,
- PFT crunches of patches 1 and 8,
- PFT 3-mile run of patch 1 and 8,

- Rifle score of patches 2 and 7.

The following variables are significant in their respective models and are negatively correlated with the percent of hits for the attack and counterattack:

- Age of patch 10,
- Height of patches 3 and 10,
- Weight of patch 3,
- AFQT score of patches 5 and 12,
- CFT MTC of patch 5.

The following personnel variables have no significant variables in their respective models:

- Rifle score.

There are no patterns for any personnel variables for the attack and c-atk percent hits. See Section F.5.2.3 for the ANOVA summary of this task.

F.6.3.4 CASEVAC by FT

We model elapsed time for the CASEVAC as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

Both the HD and LD integration levels are significant and positively correlated with the response for the models that include the following variables:

- All.

The HD and LD integration level are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following variables are significant in their respective models and are positively correlated with the CASEVAC by FT time:

- Height of patch 3,
- Weight of patch 3,
- CFT MTC of patches 1 and 2,

- CFT MANUF of patches 1 and 2,
- PFT 3-mile run of patch 1.

The following variables are significant in their respective models and are negatively correlated with the CASEVAC by FT time:

- Height of patches 1 and 4,
- Weight of patch 1,
- AFQT score of patch 1,
- GT score of patches 1 and 2,
- PFT crunches of patch 1.

The following personnel variables have no significant variables in their respective models:

- None.

There are no patterns for any personnel variables for the CASEVAC time. See section F.5.2.4 for the ANOVA summary of this task.

F.6.3.5 Prepare Fighting Position by FT

We model the amount of earth moved to prepare a fighting position as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report any statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

Both the HD and LD integration levels are significant and positively correlated with the response for the model that includes:

- CFT MANUF.

The HD and LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following variables are significant in their respective models and are positively correlated with the amount of earth moved for the defensive position:

- Age of patches 1 and 2,
- Weight of patch 2,

- AFQT score of patch 2,
- Rifle score of patch 3.

The following variables are significant in their respective models and are negatively correlated with the amount of earth moved for the defensive position:

- CFT MTC of patch 2,
- CFT MANUF of patches 2, 3, and 4,
- PFT 3-mile run of patch 2.

The models for the following variables have no significant variables in the model:

- Height,
- GT score,
- PFT crunches.

The following personnel variables have no significant variables in their respective models:

- None.

There are no patterns for any personnel variables for the amount of earth moved to prepare a fighting position. See Section F.5.2.5 for the ANOVA summary of this task.

F.6.3.6 1-km Hike

We model elapsed time for 1-km hike as a function of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report any statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

Both HD and LD Integration levels are significant and positively correlated with the response for the models that include the following variables:

- Age,
- Height,
- Weight,
- AFQT score,
- GT score,

- PFT crunches,
- PFT 3-mile run,
- Rifle score.

The LD integration level is significant and positively correlated with the response in the models that include the following variables:

- CFT MTC,
- CFT MANUF.

The HD and LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following variables are significant in their respective models and are positively correlated with the 1-km hike time:

- Age of patches 3 and 5,
- Height of patch 1,
- Weight of patch 1,
- CFT MTC of patch 7,
- CFT MANUF time of patch 10,
- PFT 3-mile run time of patch 4.

The following variables are significant in their respective models and are negatively correlated with the 1-km hike time:

- AFQT score of patches 6, 9, 10, and 12,
- CFT MTC of patch 12,
- PFT crunches of patch 6,
- Rifle score of patch 10.

The following personnel variables have no significant variables in their respective models:

- None.

There are no patterns for any personnel variables for the 1-km hike time. See section F.5.2.6 for the ANOVA summary of this task.

F.6.3.7 Negotiate Obstacle

We model elapsed time for the negotiate obstacle of each personnel variable and integration level in a separate model. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following variables:

- Age,
- Height,
- Weight,
- AFQT score,
- CFT MTC,
- CFT MANUF,
- PFT crunches,
- PFT 3-mile run,
- Rifle score.

The LD integration level is significant and positively correlated with the response for the models that include the following variables:

- Weight,
- PFT crunches.

The HD integration level is significant and negatively correlated with the response for the models that include the following variables:

- None.

The LD integration level is significant and negatively correlated with the response for the model that includes:

- GT score.

The following variables are significant in their respective models and are positively correlated with the negotiate obstacle time:

- Age of patch 5,

- Height of patches 4 and 8,
- Weight of patches 3 and 4,
- CFT MTC of patches 3, 5, and 9,
- PFT crunches of patches 6, 8, and 12,
- PFT 3-mile run of 1, 5, 6, 7, and 9.

The following variables are significant in their respective models and are negatively correlated with the negotiate obstacle time:

- Age of patch 12,
- AFQT score of patches 5, 10, and 11,
- GT score of patches 3, 5, 10, and 12,
- CFT MTC of patches 8 and 11,
- PFT crunches of patches 1 and 3,
- PFT 3-mile run of patch 8,
- Rifle score of patches 1, 8, and 9.

The following personnel variables have no significant variables in their respective models:

- None.

There are no patterns for any personnel variables for the negotiate obstacle time. See Section F.5.2.7 for the ANOVA summary of this task.

Appendix to Annex F
Provisional Infantry Supplemental Information

This appendix provides supplemental information for the Provisional Infantry (PI) portion of the GCEITF experiment. It provides information regarding the GCEITF leadership subjective comments and additional descriptive and basic inferential statistics not described in Annex F.

Section 1. GCEITF Leadership Subjective Comments

The GCEITF leadership provided comments on their observations of the experiment throughout its execution. Table F A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table F A – Summary of GCEITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement				Requesting extra breaks				Requires extra assistance				Needs no assistance				Compensating for another Marine				Gear pass off				Other				No category				Total
		C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	
7-km Hike	M	35	7	1	43	1	0	1	2	0	1	0	1	1	2	0	3	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	52
	F	0	10	21	31	0	2	4	6	0	1	2	3	0	2	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44
	Unit	0	0	0	0	0	1	0	1	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Movement to LOA	M	25	17	5	47	0	0	0	0	24	10	12	46	11	11	4	26	0	2	1	3	0	0	0	0	0	0	0	0	1	0	0	1	123
	F	1	22	17	40	0	0	0	0	3	13	27	43	0	3	4	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90
	Unit	0	0	1	1	0	0	0	0	0	1	1	2	4	4	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	12
CASEVAC to CCP; FT1 Elapsed time	M	0	1	0	1	0	0	0	0	0	0	0	0	22	17	7	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47
	F	0	1	0	1	0	0	0	0	0	2	0	2	0	7	4	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
CASEVAC to CCP; FT2 Elapsed Time	M	1	0	0	1	0	0	0	0	0	0	0	0	15	15	8	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39
	F	0	0	0	0	0	0	0	0	0	2	3	5	0	6	13	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASEVAC to CCP; FT3 Elapsed Time	M	0	0	0	0	0	0	0	0	0	0	0	0	31	15	21	67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	67
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Attack & C-Atk (% hits)	M	0	0	0	0	0	0	0	0	2	2	0	4	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	6
	F	0	1	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Prepare Defensive Position; Total Weight FT1	M	6	3	0	9	0	0	0	0	6	0	0	6	6	3	4	13	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	30
	F	0	1	0	1	0	0	0	0	0	0	0	0	0	1	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Prepare Defensive Position; Total Weight FT2	M	3	0	2	5	0	0	0	0	5	1	0	6	6	4	3	13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	25
	F	0	0	2	2	0	0	0	0	0	1	0	1	0	0	7	7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	11
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Prepare Defensive Position; Total Weight FT3	M	2	2	0	4	0	0	0	0	2	2	4	8	9	2	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1-km Hike	M	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	2
	F	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Negotiate Obstacle	M	0	0	0	0	0	0	0	0	3	1	1	5	1	0	1	2	1	1	3	5	0	0	0	0	0	0	0	0	0	0	0	0	12
	F	0	0	1	1	0	0	0	0	0	8	17	25	0	1	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	3

Section 2. Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for nine additional PI tasks. Annex F contains the descriptive statistics for the remainder of the PI tasks. The words “metric” and “task” are used interchangeably throughout this Appendix; they both refer to the experimental task.

The two tables below display the results for nine additional PI metrics. Table F B displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent differences between integration levels.

Table F C displays ANOVA and Tukey Test results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA was conducted to compare the three groups and Tukey Tests were conducted to compare each pair of two groups. If non-parametric tests were needed, Table F C displays these results instead of ANOVA and Tukey Test results. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the result was not found to be the same across all three groups. We present basic inferential statistics for two tasks (one with a potential influential point).

Table F B – PI Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
7-km Hike; first km	C	17	9.74	0.61	7.67%	9.00%	1.23%
	LD	15	10.48	0.52			
	HD	16	10.61	0.37			
7-km Hike; second km	C	17	10.07	0.46	9.37%	9.70%	0.30%
	LD	15	11.02	0.52			
	HD	16	11.05	0.47			
7-km Hike; third km	C	17	9.81	0.48	11.94%	8.95%	-2.67%
	LD	15	10.98	1.65			
	HD	16	10.69	0.54			
7-km Hike; fourth km	C	17	10.37	0.53	6.87%	7.60%	0.68%
	LD	15	11.08	0.76			
	HD	16	11.15	0.59			
7-km Hike; fifth km	C	17	12.39	1.47	18.97%	18.94%	0.02%
	LD	14	14.74	3.21			
	HD	16	14.74	2.93			
7-km Hike; sixth km	C	17	13.56	2.07	4.72%	9.02%	4.11%
	LD	15	14.20	2.56			
	HD	15	14.78	2.53			
7-km Hike; seventh km	C	18	12.87	1.34	0.94%	2.95%	1.99%
	LD	15	13.00	1.23			
	HD	16	13.25	0.68			
Attack(Hits on Target)*	C	19	563.74	76.75	-6.21%	-17.25%	-11.77%
	LD	17	528.71	60.84			
	HD	17	466.47	154.24			
Attack **[excluding potential influential point]**	C	19	563.74	76.75	-6.21%	-12.22%	-6.40%
	LD	17	528.71	60.84			
	HD	16	494.88	103.67			

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
(Hits on Target)*							
C-Atk (Hits on Target)	C	20	59.00	23.04	-2.59%	10.81%	13.75%
	LD	17	57.47	20.17			
	HD	16	65.38	32.52			

Table F C – PI ANOVA and Tukey Test Results

Metric	F statistic (df)	F Test P-value	Comparison	2-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
Attack(Hits on Target)††	6.17	0.05*	LD-C	0.14	-65.06	-5.00	-73.89	3.83
			HD-C	0.03*	-151.83	-42.70	-168.14	-26.39
			HD-LD	0.14	-115.46	-9.01	-131.45	6.98
Attack **[excluding potential influential point]** (Hits on Target)*	3.10 (2, 49)	0.05*	LD-C	0.41	-82.51	12.44	-92.29	22.23
			HD-C	0.04*	-117.11	-20.61	-127.06	-10.66
			HD-LD	0.46	-83.36	15.70	-93.57	25.91
C-Atk (Hits on Target)	0.45 (2, 50)	0.64	LD-C	0.98	-16.17	13.11	-19.19	16.13
			HD-C	0.74	-8.51	21.26	-11.58	24.33
			HD-LD	0.65	-7.55	23.36	-10.74	26.55

*Indicates there is a statistically significant difference in a two-sided hypothesis test between integration levels according to ANOVA or a non-parametric equivalent test.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.
††Indicates results presented are from Robust ANOVA and Welch's t-tests with p-values compared to 0.033 for Bonferroni adjustment due to unequal variances. The reported F-statistic is a Chi-square statistic from Robust ANOVA, and the F-test p-value is the Robust ANOVA p-value. The p-values in columns labeled "2-sided P-value" and "1-sided P-value" are p-values from Welch's t-tests, and the confidence intervals are from Welch's t-tests.

Additional Task Results:

7-km Hike by km. The general trend for the 7-km hike was that the difference between the LD and C groups, the HD and C groups, and the HD and LD groups increased over the course of the hike.

Attack (hits on target). The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.51 for the C group, 0.12 for the LD group, and 0.02 for the HD group.

The C group had a mean of 563.74 hits on target. This number is higher (but not statistically significant) than the LD group mean of 528.71 and statistically significantly higher than the HD group mean of 466.47. The LD group put 6.21% less shots on target than the C group. The HD group put 17.25% less shots on target than the C group. The HD group put 11.77% less shots on target than the LD group.

Attack with Potential Influential Point Removed (Hits on Target). The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.51 for the C group, 0.12 for the LD group, and 0.95 for the HD group.

The C group had a mean of 563.74 hits on target. This number is higher (but not statistically significant) than the LD group mean of 528.71 and statistically significantly higher than the HD group mean of 494.88. The LD group put 6.21% less shots on target than the C group. The HD group put 12.22% less shots on target than the C group. The HD group put 6.40% less shots on target than the LD group.

Counterattack (hits on Target). The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.97 for the C group, 0.22 for the LD group, and 0.72 for the HD group.

The C group had a mean of 59.00 hits on target. This number is higher (but not statistically significantly) than the LD group mean of 57.47 and lower (but not statistically significantly) than the HD group mean of 65.38. The LD group put 2.59% less shots on target than the C group, and the HD group put 10.81% more shots on target than the C group. The HD group put 13.75% more shots on target than the LD group.

Annex G.

Provisional Infantry Machine Gunner (PIMG)

This annex details the Provisional Infantry Machine Gunner (PIMG) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 2 March – 26 April 2015 at Range 107 and Range 110, aboard the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, CA. The sections outline the PIMG Scheme of Maneuver (SOM), Limitations, Deviations, Data Set Description, Descriptive and Basic Inferential Statistics, and Modeling Results.

G.1 Scheme of Maneuver

G.1.1 Experimental Cycle Overview

The Provisional Infantry Machine-gunner (PIMG) assessment of the GCEITF took place in a field environment aboard MCAGCC, Twentynine Palms, CA. The assessment consisted of 21 trial cycles, each of which was a 2-day test cycle, conducted over the course of 55 days. Marines spent 1 recovery day at Camp Wilson after every 4 days of trials. Each machine-gun squad consisted of three volunteers and a direct-assignment (non-volunteer) squad leader. Each member of the squad was trained to fill each billet within the squad: gunner, assistant gunner, and ammo man. For consistency throughout the report, the term “squad” will be used for the M240B medium machine-gun squad. The assessment was executed under the supervision of MCOTEA functional test managers and a range Officer in Charge (OIC)/Range Safety Officer (RSO) from the GCEITF.

G.1.2 Experimental Details

The 2-day PIMG assessment replicated offensive and defensive tasks. The PIMG squads began each cycle on the Defensive task. Two PIMG squads executed each trial cycle: a control (C) all-male squad and a high-density (HD) all-female squad.

Day 1 of the trial cycle was executed on Range 110 and consisted of defensive actions. The day started with a 7-km forced march from Range 107 to Range 110 wearing an approach load and carrying personal weapons, crew-served weapons, and ammunition. Each heavy machine-gun squad prepared a M2 heavy machine gun for employment. The squads engaged three targets based on a prescribed course of fire with 400 rounds of .50-cal ammunition. Immediately upon the gun going out of action, the machine-gun squad displaced the machine gun to a designated location. The trials concluded with a mount and dismount of the M2 machine gun to a HMMWV.

Day 2 of the trial cycle was executed on Range 107 and consisted of supporting a squad attack from the support-by-fire (SBF) position then displaced to the limit of advance. The PIMG squad repelled an enemy counterattack by fire for 90 seconds. Finally, each squad conducted a 100-m CASEVAC of a 220-lb dummy. At the conclusion of Day 2, the Marines reorganized into a new squad for the next experimental cycle.

G.1.3 Additional Context

Throughout the assessment, Marines bivouacked at Range 107, sleeping in two-man tents. Prior to the experiment, each machine gun was zeroed to maximize accuracy. During trial execution, Marines wore/carried prescribed loads for each task. Weighing packs each day prior to the 7-km forced march ensured consistency. After each trial day, the Marines operated under the guidance of their Company leadership, performing minimal physically demanding tasks. The Marines who were not part of an assessed PIMG squad conducted the same experimental subtasks after the assessed squads to ensure equity between individuals participating in a trial cycle and those not chosen for that particular cycle. These tasks will be discussed in detail in the loading section below.

Fatigue surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. It allows for outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) to be accounted for in the analysis of the performance data. Workload surveys collected the volunteers' perceived average and maximal level of exertion during the performance specified tasks. Cohesion surveys provided a method of collecting subjective data relating to each machine-gun squad's ability to work as a team and their overall perspective on the cohesiveness of the squad.

G.1.4 Experimental Tasks

G.1.4.1 7-km Hike

Infantry units must move through all sorts of terrain on foot. Units train by conducting a forced march with an approach load at a sustained rate of march. For the assessment, each PIMG squad moved a distance of 7.2 km as quickly as possible while carrying an approach load, and the M240B machine gun, tripod, spare barrel, and ammunition spread-loaded across all three members. This task determined the squad's rate of movement over a 7.2-km route while carrying the approach load and their crew-served weapon. Each Marine took a fatigue and workload survey after completion of the 7.2-km hike.

G.1.4.2 M2 Emplacement, Engagement, and Displacement

Providing defensive fires with the M2 heavy machine gun entails moving the system to a position of advantage, engaging the enemy, and conducting a rapid displacement.

Oftentimes, the M2 is employed from the tripod. During the assessment, each PIMG squad emplaced the M2 on a tripod at a specified firing location. The assessment began with targets exposed and engaged by the squad. Three targets and 400 rounds were allocated for this course of fire. Immediately upon going out of action, the squad displaced from the firing line to a designated location, moving a heavy load a short distance, manipulating a weapon while fatigued, and accurately engaging targets. This task determined accuracy and displacement times.

G.1.4.3 M2 Mount and Dismount

The M2 heavy machine gun is often employed from a vehicular platform, such as a HMMWV. To mount this system, the squad must lift all components from the ground to the turret and assemble the system. Similarly, to dismount the system, the squad must manually lower each component to the ground. During the assessment, each PIMG squad worked together to mount and dismount an M2 from a HMMWV. This task required the strength to lift, manipulate, and lower heavy components. This task determined the time for a squad of three Marines to fully mount and dismount the M2 from a tactical vehicle.

At completion of the mount and dismount tasks, Marines took a fatigue and workload survey to assess their fatigue and workload during execution (see GCEITF EAP, Annex D).

G.1.4.4 1-km Movement

Assaulting an enemy position never starts from a static position; first, a movement must be conducted to the assault position (AP). The distance from the line of departure to the AP is dependent upon myriad factors. Based on time and space constraints, this distance was set at just under 1 km for the experimental event. Each PIMG squad moved this distance as quickly as possible while carrying an assault load in addition to the M240B machine gun, tripod, spare barrel, and ammunition (600 rounds of 7.62 mm) divided among the members.

G.1.4.5 Occupy and Engagement from SBF Position

Prior to commencing an assault, it is common for the machine-gun section to occupy a position of overwatch and provide SBF for the attacking unit. Movement to a SBF position varies in distance based on the terrain, which is often challenging. From the SBF position, the PIMG squad must rapidly get their gun into action, acquire targets, and accurately engage the enemy. During the assessment, each PIMG squad moved approximately 100 meters from an AP to the SBF position and emplaced the M240B. Once the squad leader confirmed “gun up,” they engaged targets presented in a predetermined course of fire. Three targets and 400 rounds were allocated during this course of fire. This task determined how quickly the squad could move to a SBF position and get the gun into action, as well as determined the squad's accuracy while engaging targets.

G.1.4.6 Displace to the LOA

After providing initial suppressive fires from a SBF position to support an assault, machine-gun squads generally move to another position of advantage. Given the command, they must break down their weapon system and displace to a follow-on firing position, quickly. During the assessment, each PIMG squad displaced from their initial SBF position approximately 300 meters to a limit of advance/secondary SBF position. From this new position, they prepared to repel the enemy counterattack. This task determined how much time it took for a machine-gun squad to displace and prepare for an enemy counterattack.

G.1.4.7 Repel Counterattack

At the conclusion of every assault, the attacking unit must be prepared for the enemy to regroup and organize a counterattack. Upon consolidation at the LOA, the squad oriented its weapons in the direction of the enemy's retreat. Upon targets being presented, the counterattack commenced for 90 seconds. Two targets and 200 rounds were allocated during the machine-gun counterattack course of fire.

G.1.4.8 CASEVAC

Casualties are an inevitable part of conducting combat operations. Units train to become proficient in triaging, handling, and transporting a casualty. When a casualty is sustained, it is essential to move with a sense of urgency to get the injured Marine to the appropriate level of care. At the conclusion of the live-fire and counterattack, each PIMG squad moved a 220-lb dummy 100 meters from a position of cover to a casualty collection point (CCP) while also transporting their assault packs and crew-served weapon system (M240B, tripod, and spare barrel). The machine-gun squad could use a variety of techniques for transport but had to carry the dummy off the ground. This task determined the PIMG squad's proficiency in moving a simulated casualty to a CCP. After the CASEVAC, each Marine took a fatigue and workload survey to assess overall fatigue and workload of the entire offensive task (see GCEITF EAP, Annex D).

At the completion of the 2-day cycle, Marines took a cohesion survey to record their cohesion during execution of the 2-day trial cycle (see GCEITF EAP, Annex M.)

G.1.5 Loading Plan

The loading plan ensured, to the greatest extent possible, equity of physical activity amongst all volunteers throughout the duration of the experimental assessment. Every trial and task was conducted in the same manner and sequence to ensure consistency. Due to the number of volunteers, a handful of Marines were not part of an assessed squad each 2-day cycle. Collaboration with the Company leadership determined the best method of loading non-assessed Marines was to have them perform the same tasks as an assessed squad to experience the same conditions and physical strain. Minor modifications were permitted due to the reduced squad of the squad, such as conducting a trial as the fourth Marine in a 3-Marine element when not enough individuals were available to form another squad.

G.1.6 Scheme of Maneuver Summary

The Provisional Infantry Machine-gun experiment consisted of a 2-day trial cycle comprising of a defensive day and an offensive day. The defensive day involved four subtasks: a 7.2-km forced march, M2 engagement, displacement, and mount/dismount drill. The offensive day involved five subtasks based around supporting a squad attack: 1-km movement, movement and occupy a SBF position, displacement, repel a counterattack, and CASEVAC. During the course of the experiment, the PIMG squad executed 2 pilot trial cycles and 21 record trial cycles. During trial execution, Marines rotated through every billet within the machine-gun squad, carrying components of the crew-served load.

G.2 Limitations

G.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a similar manner to those in an operational environment. Under certain situations, artificial limitations or interruptions were introduced that changed or altered the way a task would normally be performed. While these limitations represent a degree of artificiality, they do not detract significantly from our abilities to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for PIMG assessment.

G.2.2 Relative Difficulty of Record Test

The PIMG GCEITF assessment was designed to gather data associated with some of the most physically demanding tasks that an open MOS Marine could perform when assigned to an Infantry unit with the collateral responsibilities as a machine-gunner. These tasks in isolation do not fully replicate life experienced by a Marine during a typical FEX or a combat environment and did not obtain the maximum cumulative load that could be placed on a Marine assigned to a ground combat element. With limited time available, only selective tasks were assessed. Other tasks/duties outside of the assessment were minimized due to specific experimental constraints and human factors. During a typical FEX, it is common for Marines to conduct 24-hour operations that include day and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. The offensive day SOM took squads approximately 1 hour to complete, and the defensive day SOM took approximately 3 hours to complete. Outside the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

Another concern in designing the PIMG assessment was making it was achievable and sustainable for a 60-day period. The 7-km forced march distance was selected based on the training time available prior to the assessment. However, many of the loads carried were decreased; the crew-served load was altered from the M2 heavy machine gun to the M240B medium machine gun. Marines were authorized 1 day off after every 4 days of training; this artificial recovery period is not achievable when conducting training or combat operations.

A final factor affecting the relative difficulty of the record test pertains to the intangible physiological impact of the volunteers' ability to DOR at any point during a trial. Any time a volunteer dropped during a trial cycle, that squad performed the following subtasks with fewer personnel. This factor could have affected the cohesion of each squad and influenced its performance.

G.2.3 Geometries of Fire and Conditions Set

Several artificialities were present during the M240B live-fire portion of the assessment. Although the SBF position was realistic, the geometries of fire for the initial and secondary position were offset to prevent fires from interfering with the rifle-squad assessment. The PIMG course of fire began when they reached the SBF position, rather than waiting for the tactical conditions to be established. The loss of tactical realism in basing fires off a maneuver element resulted in this task being less challenging than in training or combat, in which the squad leader and gunner must make intuitive decisions regarding rates of fire, target precedence, and shifting and ceasing fire criteria.

G.2.4 Number of Volunteer Participants

For the PIMG experiment, four male and four female volunteers began the experiment, but by the end three males and four females completed the assessment. The results presented in this annex are based on the performance of 7 to 8 Marines.

G.2.5 PIMG Limitations Summary

The PIMG assessment was designed to replicate realistic training in a field environment. The end-state was to create an experiment in which the volunteers felt they were conducting realistic and operationally relevant tasks, but unavoidable limitations to the assessed tasks and non-assessed operating environment introduced a level of artificiality not normally present in a field training or combat environment.

G.3 Deviations

G.3.1 M2 Heavy Machine-gun Employment

The EAP stated that just prior to the M2 .50-cal engagement, the Marines would be assessed moving the M2 and ammunition approximately 100 meters to the firing line. Discussion with Company leadership about the relevancy of this subtask informed the decision not to make the emplacement a timed event, but to administratively move the M2 to the firing line. This subtask was not assessed as a standalone measurement.

G.4 Data Set Description

G.4.1 Data Set Overview

The PIMG portion of the experiment consisted of 2 pilot trial cycles and 21 record trial cycles. The pilot trial cycles were conducted from 2 to 6 March 2015. Pilot trial cycle data are not used in analysis due to variations in the conduct of the test. We based all analysis on the 21 record trial cycles executed from 7 March to 26 April 2015.

G.4.2 Record Test Volunteer Participants

At the beginning of the first record trial, there were four male PIMG volunteers and four female volunteers. There was one male DOR.

G.4.3 Planned, Executed and Analyzed Trial Cycles

Table G-1 displays the number of trial cycles planned, executed and analyzed by task. The planned number of trial cycles for the PIMG MOS per Section 7.5.3 of the GCEITF EAP is 78 trial cycles or 26 per planned integration level (C, LD, and HD). The plan called for 4 squads per day over the 20 trial cycles. Due to the number of Marines who voluntarily withdrew or were involuntarily withdrawn from the experiment prior to the execution of the first record trial cycle, only one squad of the C and HD integration levels remained. The planned number of trial cycles in Table G-1 reflects 21 planned trial cycles for each integration level.

Of note, there are several occurrences of missing data for the 7-km hike by individual kilometer. The individual kilometer times were derived from GPS data. Early in the experiment, the Garmin GPS's were set to record a volunteer's position every second. Due to the storage space on the GPS and length of the trial when volunteers executed the 7-km hike and then follow-on tasks, the GPS could not hold all of the data and overwrote the hike data. Once the problem was found, the GPS's were corrected to record location every 2 seconds.

Table G-1. PIMG Planned, Executed, and Analyzed Trial Cycles[1]

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of trials used in analysis	Notes
7-km Hike	C	21	19	19	
	HD	21	21	21	
M240B Movement & Emplacement	C	21	19	19	
	HD	21	21	21	
M240B Engagement	C	21	19	17	TIR: Apr 7, Apr 12
	HD	21	21	20	TRACR malfunction Apr 12
CASEVAC	C	21	19	18	Task failure-Mar 8
	HD	21	21	19	Task failure-Mar 10, Mar 18
M2 Mount/Dismount	C	21	19	19	
	HD	21	21	21	
M240B Displace to LOA	C	21	19	19	
	HD	21	21	20	High outlier: Mar 8
M2 Displacement	C	21	19	18	High outlier: Apr 11
	HD	21	21	21	
1-km Hike	C	21	19	19	
	HD	21	21	21	
7-km Hike; 1km Time	C	21	19	19	
	HD	21	21	17	No data: Mar 7, Mar 9, Mar 12, Mar 14
7-km Hike; 2km Time	C	21	19	19	
	HD	21	21	17	No data: Mar 7, Mar 9, Mar 12, Mar 14
7-km Hike; 3km Time	C	21	19	17	High outlier: Mar 7, Apr 11
	HD	21	21	16	High outlier: Apr 3; No data: Mar 7, Mar 9, Mar 12, Mar 14
7-km Hike; 4km Time	C	21	19	18	High outlier: Mar 7
	HD	21	21	18	Low outlier: Apr 3; No data: Mar 7, Mar 12
7-km Hike; 5km Time	C	21	19	19	
	HD	21	21	19	No data: Mar 7, Mar 12
7-km Hike; 6km Time	C	21	19	19	
	HD	21	21	19	High outlier: Apr 18; No data: Mar 7
7-km Hike; 7km Time	C	21	19	19	
	HD	21	21	20	No data: Mar 7
M240B Movement	C	21	19	19	
	HD	21	21	21	
M240B Emplacement	C	21	19	19	
	HD	21	21	20	Low outlier: Mar 18
M240B Attack Hits on Target	C	21	19	17	TRACR malfunction: Apr 12; TIR: Apr 7
	HD	21	21	20	TRACR malfunction: Apr 12
M240B Counterattack hits on Target	C	21	19	16	TRACR malfunction: Apr 12; Missing Apr 2; TIR: Apr 7
	HD	21	21	20	TRACR malfunction: Apr 12
Mount M2;	C	21	19	19	

[1] A TIR in this table refers to a Test Incident Report, which is a report the test team or direct assignment leaders completed when an incident occurred that affected the natural execution of a trial. If a data point is removed due to a TIR, it is because the TIR affected the data in such a way that it is not comparable to the rest of the data set.

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of trials used in analysis	Notes
Elapsed Time	HD	21	21	21	
Dismount M2; Elapsed Time	C	21	19	19	
	HD	21	21	21	
M2 Engagement	C	21	19	19	
	HD	21	21	21	

G.5 Descriptive and Basic Inferential Statistics

G.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common infantry tasks and are indicative of unit-level proficiency during either field exercises or combat operations. This section presents results for 7 out of 16 tasks. The Appendix to this Annex contains the descriptive statistics for the remainder of the PIMG tasks. The words “metric” and “task” are used interchangeably throughout this Annex; both refer to the experimental task.

Each machine-gun squad consisted of three volunteer Marines: the gunner, assist-gunner, and ammo man. A direct assignment (non-volunteer) squad leader led each squad. There were two integration levels for all tasks. A C group was all male and a HD group had three females.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA) (or non-parametric tests as necessary), and scatter plots. The subsequent sections will cover each task in detail. Lastly, contextual comments, additional insights, and subjective comments (as applicable) tying back to each experimental task are incorporated.

Use caution when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks, such as distances, techniques, leadership, load carried, group size, and group composition.

G.5.2 PIMG Selected Tasks Overview

The two tables below display the results for the seven selected PIMG metrics. Table G-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations. Table G-3 displays ANOVA results, including metrics and integration levels, p-values suggesting statistical significance, integration level elapsed-time differences, and percentage differences between integration levels. For each task, an ANOVA and t-test were conducted to compare the two groups. If non-parametric tests were needed, Table G-3 displays these results instead of ANOVA and t-test results. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the HD group is different from that in the C group.

Table G-2. PIMG Selected Task Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD
7-km Hike (minutes)*	C	19	103.34	8.68
	HD	21	128.96	14.19
M240B Movement & Emplacement (minutes)*	C	19	3.81	0.43
	HD	21	4.38	0.48
M240B Engagement (% hits)	C	17	32.31%	15.61%
	HD	20	31.71%	9.60%
CASEVAC (minutes)*	C	18	4.04	1.61
	HD	19	6.55	2.21
M2 Mount/Dismount (minutes)*	C	19	6.00	0.82
	HD	21	5.47	0.80
M240B Displace to LOA (minutes)*	C	19	4.98	0.44
	HD	20	5.87	0.85
M2 Displacement (minutes)*	C	18	1.71	0.26
	HD	21	1.98	0.59
M2 Displacement **[excluding potential influential point]** (minutes)*	C	18	1.71	0.26
	HD	20	1.88	0.41

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table G-3. PIMG Selected Task ANOVA and Welch's T-test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-sided P-Value	1-sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
7-km Hike (minutes)*	46.22 (1, 38)	< 0.01*	HD-C	25.62	24.79%	< 0.01*	< 0.01*	20.81	30.43	19.39	31.85
M240B Movement & Emplacement (minutes)*	15.77 (1, 38)	< 0.01*	HD-C	0.57	15.10%	< 0.01*	< 0.01*	0.39	0.76	0.33	0.82
M240B Engagement (% hits)	0.02 (1, 35)	0.89	HD-C	-0.01	-1.87%	0.89	0.45	-0.06	0.05	-0.08	0.07
CASEVAC (minutes)*	15.42 (1, 35)	< 0.01*	HD-C	2.51	62.10%	< 0.01*	< 0.01*	1.68	3.34	1.44	3.58
M2 Mount/Dismount (minutes)*	4.26 (1, 38)	0.05*	HD-C	-0.53	-8.84%	0.05*	0.02*	-0.87	-0.19	-0.96	0.98
M240B Displace to LOA (minutes)*	16.67 (1, 37)	< 0.01*	HD-C	0.89	17.91%	< 0.01*	< 0.01*	0.61	1.17	0.53	1.26
M2 Displacement (minutes)*	2.68††	0.10††	HD-C	0.27	15.79%	0.10†	0.05†	0.05†	0.28†	0.00†	0.33†
M2 Displacement **[excluding**	2.41 (1, 36)	0.13	HD-C	0.18	10.25%	0.13	0.06*	0.03	0.32	-0.01	0.36

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-sided P-Value	1-sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
potential influential point] (minutes)											

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality.
††Indicates results presented are from Robust ANOVA due to unequal variances. The reported F-statistic is a Chi-square statistic from Robust ANOVA, and the F-test p-value is the Robust ANOVA p-value.

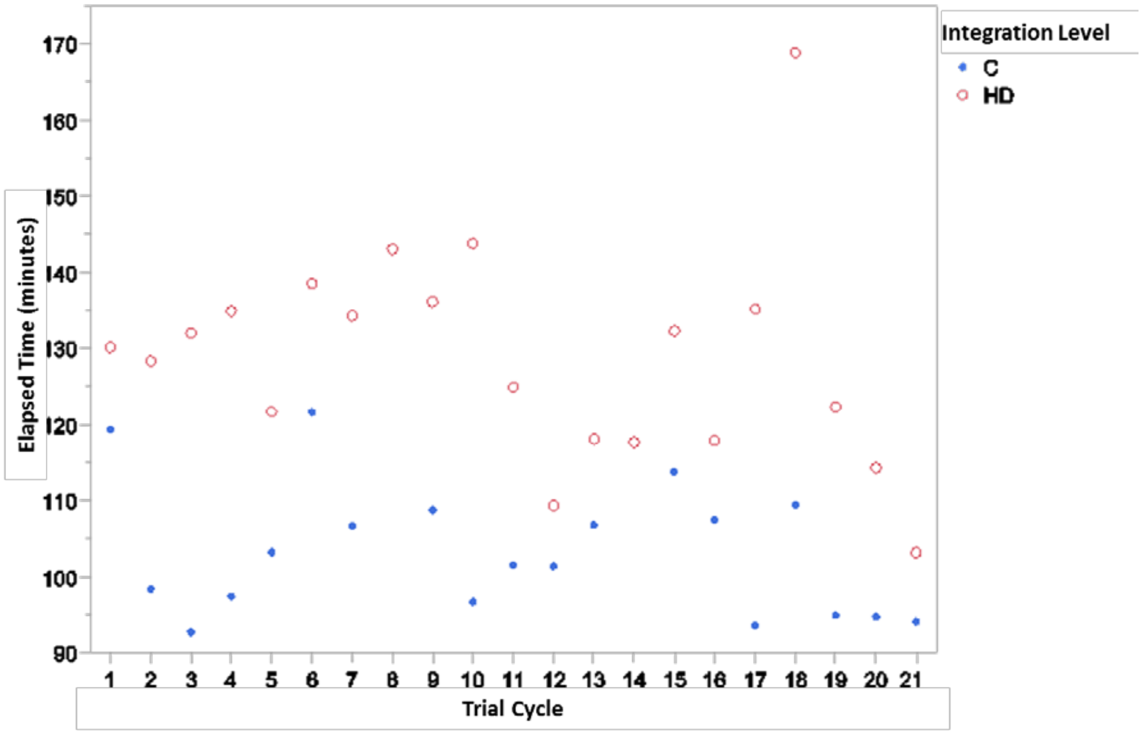
G.5.2.1 7-km Hike

G.5.2.1.1 7-km Hike Overview

This experimental task assessed a squad of three Marines moving 7.20 km while each Marine carried an approach load, and individual weapon (M-4), and a portion of the crew-served weapon load. The crew-served load consisted of an M240B medium machine gun, tripod, A-bag with spare barrel, and four cans of ammo resulting in a cumulative load of 118-130 lb per Marine. The recorded time for this task started when the squad departed the Range 107 start point and stopped when the squad arrived at the Range 110 stop point.

Figure G-1 displays all PIMG 7-km hike data. All data on the scatter plot are valid for analysis.

Figure G-1. 7-km Hike



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.13 for the C group and 0.32 for the HD group.

The C group had a mean time of 103.34 minutes. This time is statistically significantly faster than the HD mean time of 128.98 minutes. This difference results in a 24.79%, or 25.62-minute, degradation in hike time between the groups. The HD group had greater variability in times, as shown by the 5.51-minute increase in standard deviation (SD) (8.68 minutes for the C group and 14.19 minutes for the HD group). See Table G-2 and Table G-3 for detailed analytical results.

G.5.2.1.2 7-km Hike Contextual Comments

G.5.2.1.2.1 USMC Hike Standards

The 7-km hike is a task that the following MOSs completed during the experiment: 0311, 0331, 0341, 0351 and 0352, PI, Provisional Machine Gunners, and Combat Engineers. There are varying standards to which we can compare this result. The following sections define those standards as well as the one we choose as a comparison.

The Infantry T&R Manual (30 Aug 2013) identifies minimum standards that Marines must be able to perform in combat, to include standards for tactical marches. In Chapter 8 of the Infantry T&R Manual in the MOS 0300 Individual Events section, task “0300-COND-1001: March under an approach load” is applicable to all 03XX MOSs, ranks PVT – LtCol. The condition and standard established by this task is: “Given an assignment as a member of a squad, individual weapon, and an approach load, complete a 20 kilometer march in under 5 hours." The march pace required by this standard is 4 kilometers per hour (km/h). In Chapter 9 of the Infantry T&R Manual in the MOS 0302/0369 Individual Events section, tasks "0302-OPS-2001: Lead an approach march" and "0369-OPS-2501: Lead an approach march" are applicable to MOS 0302 and 0369, ranks SSgt – MGySgt and 2ndLt – LtCol. The condition and standard established by these tasks is: "Given a mission, time constraints, an approach march load, organic weapons, and a route, move 24.8 miles (40 km) in a time limit of 8 hours." The march pace required by this standard is 5 km/h. Appendix E of the Infantry T&R Manual (Load Terms and Definitions) states: "The approach march load will be such that the average infantry Marine will be able to conduct a 20 mile hike in 8-hours with the reasonable expectation of maintaining 90% combat effectiveness." The march pace required by this definition is 4.02 km/h.

Chapter 3 of Fleet Marine Force Reference Publication (FMFRP) 3-02A, Marine Physical Readiness Training for Combat (16 Jun 2004) states: "The normal pace is 30 inches. A pace of 30 inches and a cadence of 106 steps per minute result in a speed of 4.8 km/h or 3 mi/h and a rate of 4 km/h or 2.5 mi/h if a 10-minute rest halt per hour is taken." Common Infantry practice today is to hike for 50 minutes and take a 10-minute break, while maintaining an overall pace of 4 km/h (resulting in a hiking pace of 4.8 km/h).

Driven by the need to pick an evaluative reference standard, this report follows the T&R Manual's intent to establish minimum standards and uses the 4 km/h march pace for a 20 km march established by task 0300-COND-1001 and supported by the definition of an approach load. Further, while an established reference standard is required to anchor observed performance to the T&R program, more important information is provided by performance differences observed between gender integrated and non-gender integrated units.

G.5.2.1.2.2 PIMG 7-km Hike Pace

This result is relevant to the training and combat environments as it will take integrated squads more time to conduct foot movements. Per the tactical march standards noted above, the Marine Corps standard of hiking is 4.0 km/h. The HD groups failed to meet this standard. The C group average pace was 4.18 km/h; the HD group average pace was 3.35 km/h, finishing 25.62 minutes behind the C group. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), it would take the C group 4.78 hr and the HD group 5.97 hr to complete the movement, meaning the HD group would finish 71.4 minutes behind the C group.

Furthermore, on any given day (under the same environmental conditions), the C group was faster than the HD group 100% of the time (19 of 19 trial cycles).

G.5.2.1.3 7-km Hike Additional Insights

A 25.62-minute difference in the 7-km hike for machine-gun squads is statically significant and operationally relevant, based on the fact that the HD group failed to meet the USMC standard of 4 km/h. Based on the USMC standard of a 4 km/h pace over a 7.20-km route (which would result in a 108-minute hike completion time over the 7.20 km), the HD group was 20.96 minutes slower than that standard.

Marine Corps Doctrinal Publications (MCDP) consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to “Being Faster” and states, “Physical speed, moving more miles per hour, is a powerful weapon in itself.” MCDP-6 Command and Control also speaks to speed relative to the enemy and states, “The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results.” Further insights may be gleaned from the Appendix, which shows the difference in speed by kilometer. In general, the difference in performance increased as the movement got longer.

G.5.2.1.4 7-km Hike Subjective Comments

For subjective comments relating to this task, see the Appendix.

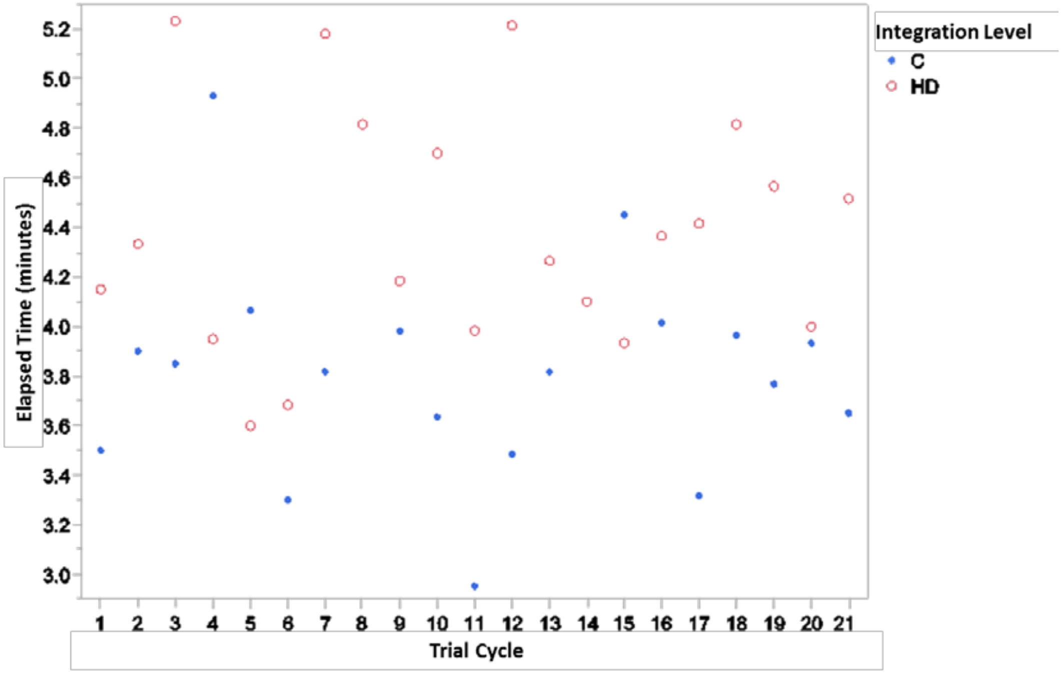
G.5.2.2 M240B Movement & Emplacement

G.5.2.2.1 M240B Movement & Emplacement Overview

This experimental task assessed the time for a 3-Marine squad to move approximately 100 meters to a SBF position and prepare the M240B medium machine gun for firing. This task was conducted immediately after completing a 1-km movement. The recorded time started immediately upon completing the 1-km movement and stopped when the gunner yelled “Gun up,” indicating that it was ready to fire. During emplacement, the squad assembled the M240, checked headspace and timing, and loaded the source of ammunition. A direct-assignment squad leader verified all procedures.

Figure G-2 displays all PIMG M240B Movement & Emplacement data. All data on the scatter plot are valid for analysis.

Figure G-2. M240B Movement & Emplacement



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.34 for the C group and 0.40 for the HD group.

The C group had a mean time of 3.81 minutes. This time is statistically significantly faster than the HD mean time of 4.38 minutes. This difference results in a 15.10%, or 0.57-minute, degradation in time. See Table G-2 and Table G-3 for detailed analytical results.

G.5.2.2.2 M240 Movement & Emplacement Contextual Comments

This result is relevant to the combat environment, as it will take integrated squads more time to move and initiate supporting fires. On average, it took the HD group 34 seconds longer to conduct this movement and emplacement. On any given day (under the same environmental conditions), the C group was faster than the HD group 84.2% of the time (16 of 19 trial cycles).

G.5.2.2.3 M20B Movement & Emplacement Additional Insights

A purely objective evaluation of 34 seconds is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated squad. Considering the 650 rounds-per-minute sustained rate of fire for the Russian RPK-47 machine gun, a single enemy machine-gun squad would have the opportunity fire 368 rounds against the pinned down squad prior to the initiation of friendly suppression by an integrated squad. The resultant trade in casualty exchange could be significant.

G.5.2.2.4 M240B Movement & Emplacement Subjective Comments

For subjective comments relating to this task, see the Appendix.

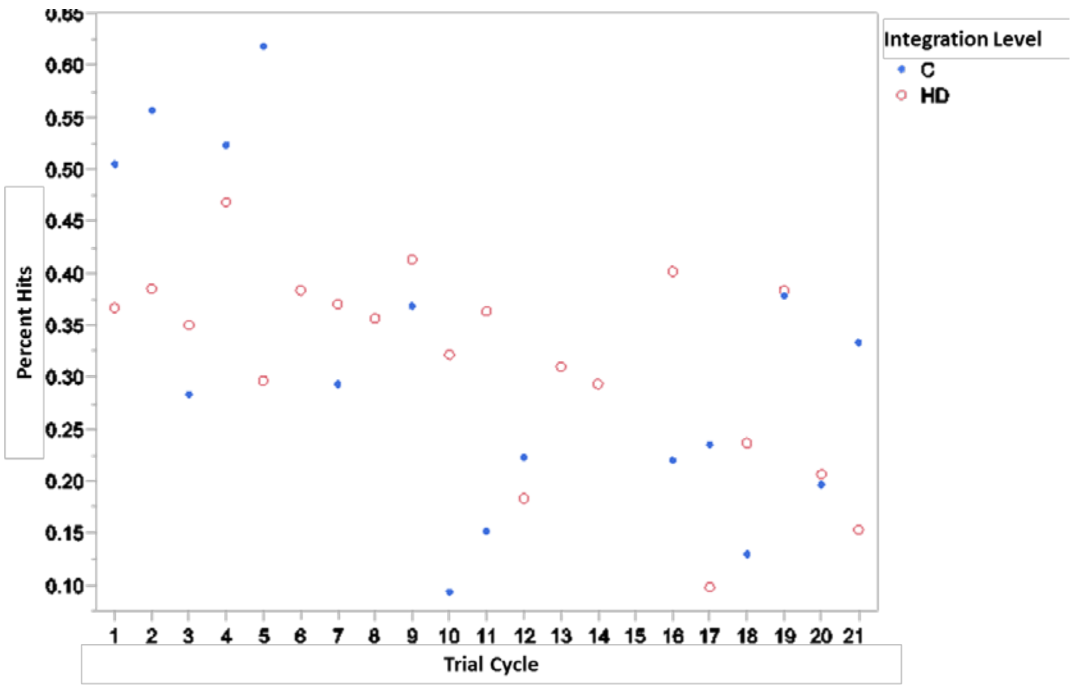
G.5.2.3 M240B Engagement

G.5.2.3.1 M240B Engagement Overview

This experimental task assessed the accuracy of a medium machine-gun squad engaging GTSs with 400 rounds during an attack and 200 rounds during a counterattack. Machine-gun squads engaged three GTSs during the attack and two during the counterattack. Each GTS captured the precise location of a round that passes within 3 meters of the location of hit and miss (LOMAH) sensor. A significant limitation of the GTS was that it did not capture any data that impacted the low (the berm of the target). The accuracy was determined by dividing the number of rounds detected on each target by the total amount of ammunition expended by each squad.

Figure G-3 displays all PIMG M240B Engagement data. All data on the scatter plot are valid for analysis.

Figure G-3. M240B Engagement



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.61 for the C group and 0.11 for the HD group. See Table G-2 and Table G-3 for detailed analytical results.

The C group had an average percent hit of 32.31%. This percentage is higher (but not statistically significantly) than the HD average of 31.71%. This difference in probability results in a 1.87%, or 1-percentage point, degradation in accuracy. The HD group was less variable than the C group, as shown by the 6.01-percentage point decrease in SD (15.61% for the C group and 9.60% for the HD group).

G.5.2.3.2 M240B Engagement Contextual Comments

In combat operations, accuracy is highly desirable in destroying or effectively suppressing an enemy position. On average, the HD group was only 0.6% less accurate than the C group.

G.5.2.3.3 M240B Engagement Additional Insights

The overall accuracy of each group had a downward (negative) trend over the course of the experiment. One would have expected the accuracy to improve over time. This can be attributed to a loss in interest due to the repetitive nature of the task and a lack of feedback to the volunteers.

G.5.2.3.4 M240B Engagement Subjective Comments

For subjective comments relating to this task, see the Appendix.

G.5.2.4 CASEVAC

G.5.2.4.1 CASEVAC Overview

This experimental task assessed the machine-gun squad's ability to move a 220-lb. dummy a distance of 100 meters to a CCP while wearing an assault load, individual weapon (M4), and the M240B crew-served load. Squads could use a variety of techniques, but they had to move all personnel and gear the entire distance, as well as carry the dummy off the ground. The machine-gun squad conducted this task at the conclusion of the counterattack engagement. Techniques varied between the 1-Marine (fireman) and 2-Marine carry. The recorded time started when Marines touched the dummy and stopped when the dummy and all members of the squad arrived at the CCP.

Figure G-4 displays all PIMG CASEVAC data. All data on the scatter plot are valid for analysis.

Figure G-4. PIMG CASEVAC

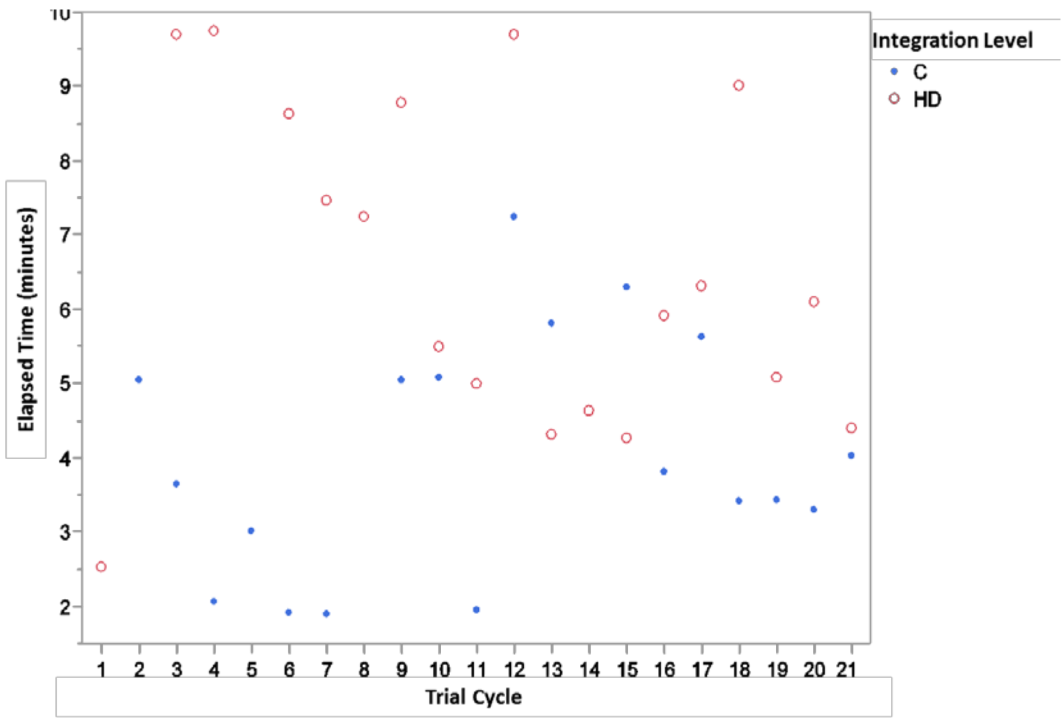


Figure 4. PIMG CASEVAC Plot

The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.34 for the C group and 0.18 for the HD group.

The C group had a mean time of 4.04 minutes. This time is statistically significantly faster than the HD mean time of 6.55 minutes. This difference results in a 62.10%, or 2.51-minute, degradation in CASEVAC time. Additionally, the HD had more variability, as shown by the 0.60-minute increase in SD (1.61 minutes for the C group and 2.21 minutes for the HD group). See Table G-2 and Table G-3 for detailed analytical results.

G.5.2.4.2 CASEVAC Contextual Comments

The implications of this task contain relevance to training and combat environments as a casualty must be moved expediently to a higher echelon of medical care. The data demonstrates that the HD group took 2.5 minutes longer on average than the C group. On any given day (under the same environmental conditions), the C group was faster than the HD group 87.5% of the time (14 of 16 trial cycles).

G.5.2.4.3 CASEVAC Additional Insights

While the "Golden Hour" is a common medical planning construct for C2 and logistical support, medical literature supports a "Platinum Ten" philosophy of first response. The U.S. National Library of Medicine references a French article that espouses, "on the battlefield, the majority of casualties die within ten minutes of the trauma*." (Wounded in Action: The Platinum Ten Minutes and the Golden Hour*, Daban) The fundamental principle is that a patient needs to be correctly triaged and moved to medical care as fast as possible.

G.5.2.4.4 CASEVAC Subjective Comments

For subjective comments relating to this task, see the Appendix.

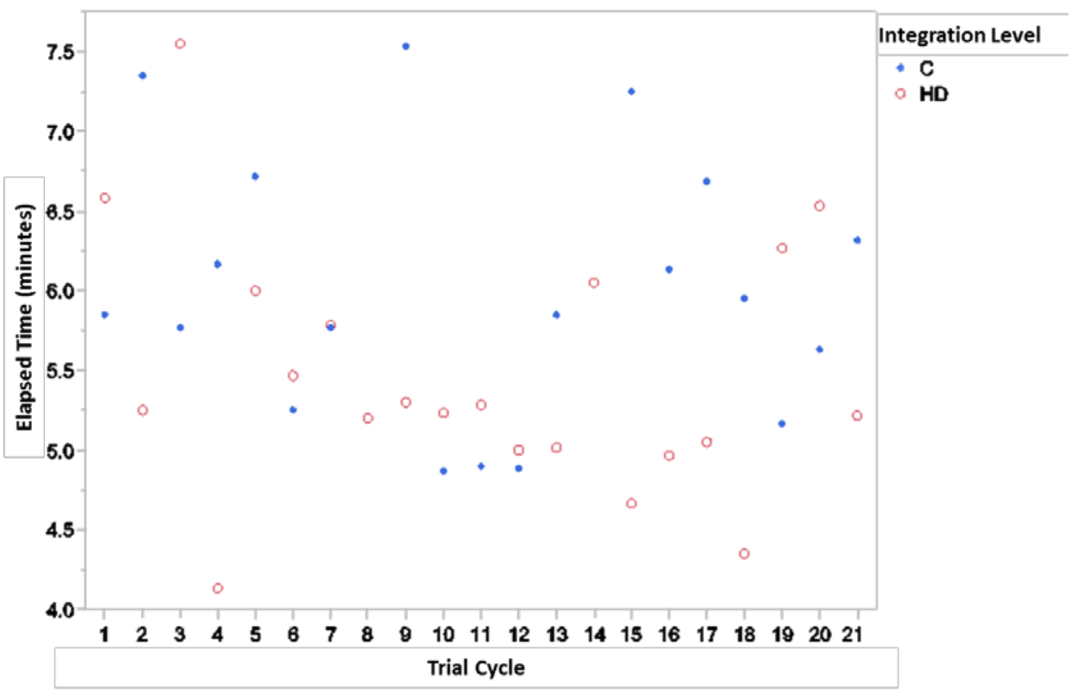
G.5.2.5 M2 Mount/Dismount

G.5.2.5.1 M2 Mount/Dismount Overview

This experimental task assessed the machine-gun squad's ability to mount an M2 heavy machine gun onto a tactical vehicle (HMMWV) and then dismount it. A brief, non-assessed administrative pause was conducted between the mounting and dismounting sub-tasks. This task was conducted immediately following the M2 live-fire portion of the experiment. The recorded time started when Marines touched a component of the M2 and it stopped when the last component of the M2 was back on the deck.

Figure G-5 displays all PIMG M2 Mount/Dismount data. All data on the scatter plot are valid for analysis.

Figure G-5. M2 Mount/Dismount



The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.30 for the C group and 0.22 for the HD group.

The C group had a mean time of 6 minutes. This time is statistically significantly slower than the HD mean time of 5.47 minutes. This difference results in a 8.84%, or

0.53-minute, improvement in mount/dismount time. See Table G-2 and Table G-3 for detailed analytical results.

G.5.2.5.2 M2 Mount/Dismount Contextual Comments

The challenging aspect of mounting and dismounting the M2 heavy machine gun is the weight of the components. On average, it took the HD group 32 seconds longer than the C group. On any given day (under the same environmental conditions), the HD group was faster than the C group 52.6% of the time (10 of 19 trial cycles).

G.5.2.5.3 M2 Mount/Dismount Additional Insights

This task began with all the equipment prestaged on the deck within 2 meters of the HMMWV. The Marines were not required to move the components to/from a secured location (armory). For an indication of the performance when moving heavy objects, see the TOW Engagement data, which involved moving a 50-lb missile a distance of 100 meters. An interesting comparison can be made to the 0331 squads. For unexplainable reasons, the PIMG C group was, on average, 2.05 minutes slower than the 0331 C group.

G.5.2.5.4 M2 Mount/Dismount Subjective Comments

For subjective comments relating to this task, see the Appendix.

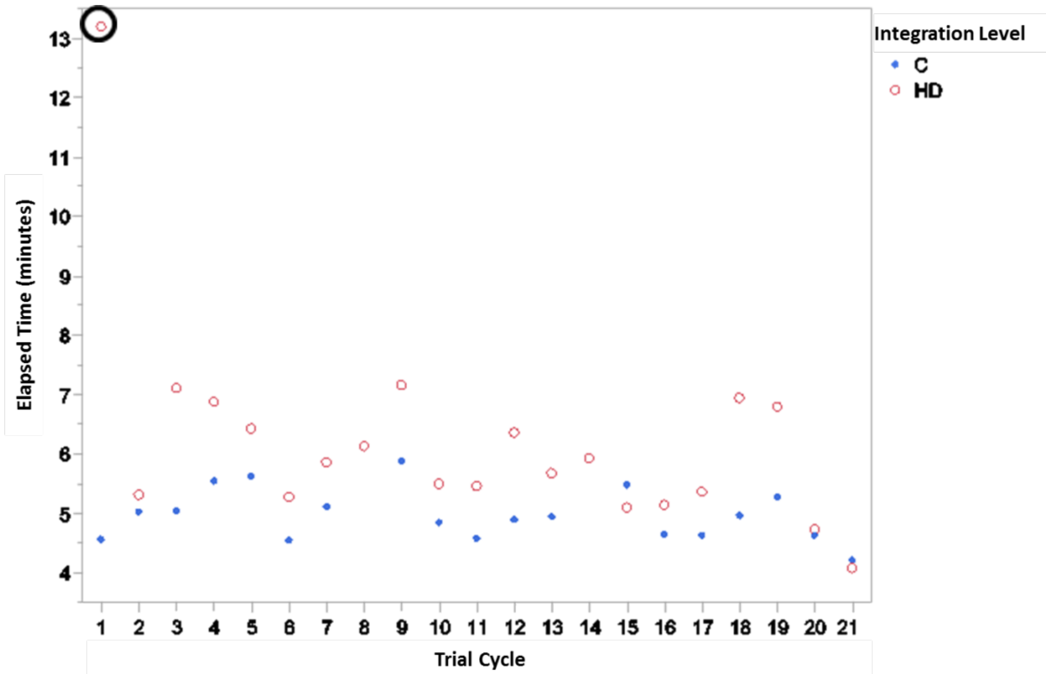
G.5.2.6 M240B Displace to LOA

G.5.2.6.1 M240B Displace to LOA Overview

This experimental task assessed a medium machine-gun squad moving approximately 300 meters to an LOA immediately after engaging targets from a SBF position. The recorded time started when the squad was told to displace and stopped when the tripod was down at the LOA.

Figure G-6 displays all PIMG M240B Displace to LOA data. The HD data point on trial cycle one was a high outlier based on a trial anomaly and removed from analysis. All other data on the scatter plot are valid for analysis.

Figure G-6. M240B Displace to LOA



The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.53 for the C group and 0.55 for the HD group.

The C group had a mean time of 4.98 minutes. This time is statistically significantly faster than the HD mean time of 5.87 minutes. This difference results in a 17.97%, or 0.89-minute, degradation in movement time. Additionally, the HD had more variability as shown by the 0.41-minute increase in SD (0.44 minutes for the C group and 0.85 minutes for the HD group). See Table G-2 and Table G-3 for detailed analytical results.

G.5.2.6.2 M240B Displace to LOA Contextual Comments

The ability to close with the objective after having conducted an attack is a crucial aspect to maintaining the momentum during offensive operations. On average, the HD group took 53.4 seconds longer, which results in that much less time to prepare for a counterattack. Based on the standard deviations, the variation in performance of the HD group is nearly than twice as much as the variation in performance of the C group. This inconsistency in the performance of the integrated squads leads to greater uncertainty and less confidence in their future performance from the mean.

G.5.2.6.3 M240B Displace to LOA Additional Insights

A purely objective evaluation of 53-second average delay is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated squad. Considering the 650 rounds-per-minute sustained rate of fire for the Russian RPK-47 machine gun, a single enemy machine-gun squad would have the opportunity fire 578 rounds against an integrated squad while moving to the LOA and exposed to enemy fire. Similarly, the integrated squad would have that much less time to provide reinforcing fires against an enemy counterattack. The resultant trade in casualty exchange could be significant.

G.5.2.6.4 M240B Displace to LOA Subjective Comments

For subjective comments relating to this task, see the Appendix.

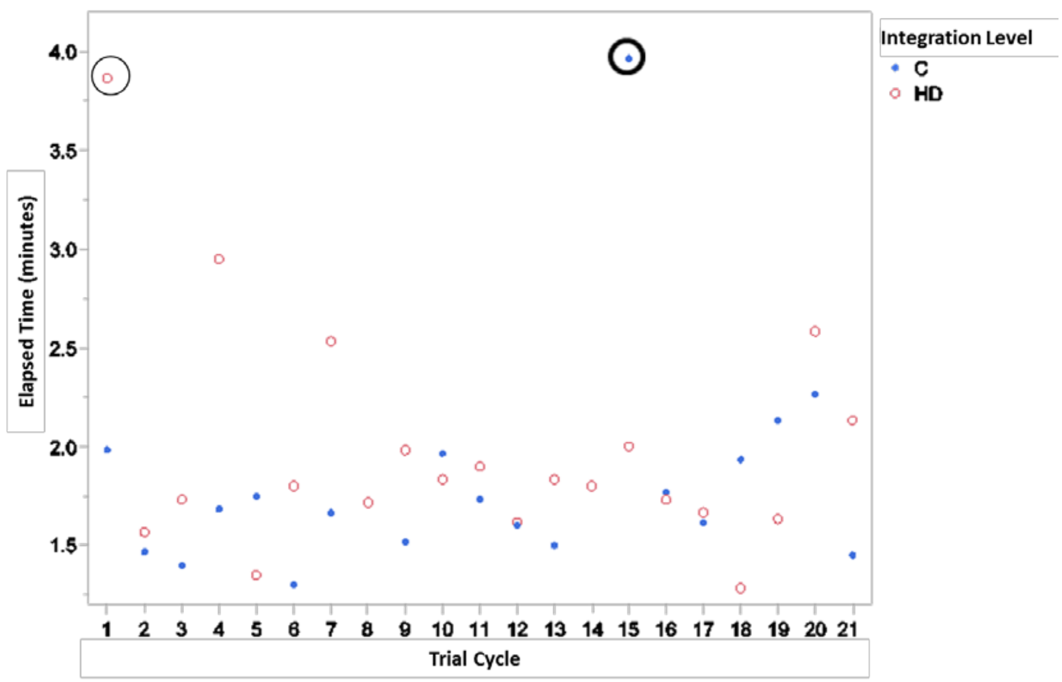
G.5.2.7 M2 Displacement

G.5.2.7.1 M2 Displacement Overview

This experimental task assessed the machine-gun squad's ability to move an M2 heavy machine gun approximately 100 meters from a firing position to a rally point. The recorded time started when the machine-gun squad yelled "Out of action" and the time stopped when the entire squad arrived at the rally point.

Figure G-7 displays all PIMG M2 displacement data. There was one influential point: the HD on trial cycle 1. Because the impact of this point is unknown, we perform all analysis with and without this point. The C on trial cycle 15 is an outlier due to unit attrition and excluded for analysis. The data point with the thin circle is the influential point, and the data point with the thicker circle is the outlier. With the exception of the C group data point on trial cycle 15, all data on the scatter plot are valid for analysis.

Figure G-7. M2 Displacement with Potential Influential Point Circled



The inclusion of the potential influential points does not change the statistical significance between groups. It does change the SD and percent differences between the integration levels. Once we remove the potential influential points, the percent difference between the C group and HD group decreases. The SD for the HD groups decreases without the potential influential point. Additionally, the removal of the potential influential points results in the data for HD group being normally distributed. The following sections discuss results with and with the potential influential points.

G.5.2.7.1.1 M2 Displacement Descriptive Statistics with Potential Influential Points

The data for the C group are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.64 but not normally distributed for the HD group, which had a p-value of <0.01. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a two-sided p-value of 0.10 (0.05 for a one-sided).

The C group had a mean time of 1.71 minutes. This time is statistically significantly faster than the HD mean time of 1.98 minutes. This difference results in a 15.79%, or 0.27-minute, degradation in displacement time. Additionally, the HD had more variability as shown by the 0.33-minute increase in SD (0.26 minutes for the C group and 0.59 minutes for the HD group). See Table G-2 and Table G-3 for detailed analytical results.

G.5.2.7.1.2 M2 Displacement Descriptive Statistics without Potential Influential Points

The data are normally distributed as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.64 for the C group and 0.03 for the HD group.

Without the influential point, the HD group mean time was 1.88 minutes. The C group is statistically significantly faster in a one-sided test. This difference results in a 10.25%, or 0.18-minute, degradation in displacement time. Additionally, the HD group had more variability as shown by the 0.15-minute increase in SD (0.26 minutes for the C group and 0.41 minutes for the HD group). See Table G-2 and Table G-3 for detailed analytical results.

G.5.2.7.2 M2 Displacement Contextual Comments

After an engagement, a machine-gun squad must be able to breakdown their weapon system and move to a position of cover as rapidly as possible. On average, the HD group took 16 seconds longer, which results in being exposed to the enemy that much longer. Based on the standard deviations, the variation in performance of the HD group is nearly than twice as much as the variation in performance of the C group. This inconsistency in the performance of the integrated squads leads to greater uncertainty and less confidence in their future performance from the mean.

G.5.2.7.3 M2 Displacement Additional Comments

One interesting comparison can be made between this task and the PIMG M240B Displacement. The PIMG M2 Displacement involved a shorter movement of 100 meters with a heavier load versus the PIMG M240B Displacement that involved a longer movement of 300 meters with a lighter load. The data reveals roughly the exact same result: a 15.79% difference and 17.91% difference, respectively, between the two different tasks.

G.5.2.7.4 M2 Displacement Subjective Comments

For subjective comments relating to this task, see the Appendix.

G.6 Statistical Modeling Results

G.6.1 Statistical Modeling Results Overview

The previous section discussed results only as they pertain to differences due to integration level alone. The goal of statistical modeling, as applied here, is to estimate, simultaneously, the effect of gender-integration levels and other relevant variables on squad performance. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For the same seven selected tasks described in the previous section, this section presents an overview of the analysis and results, and then presents the modeling results for each of the tasks.

For each task, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates an increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time but not a desired outcome for the percent hits outcome. The results report where certain patch numbers are significant for a given variable. The experiment tracked Marines within the machine-gun squad by a patch number that associated their random position within the squad to a specific billet. Table G-4 displays the patch numbers and associated billet titles for the machine-gun squad.

Table G-4. Patch Numbers and Billet Titles for the Machine-gun Squad

Patch Number	Billet Title
1	Gunner
2	Assistant Gunner
3	Ammo Man

G.6.2 PIMG Selected Tasks Method of Analysis

Due to the small number of trials, a mixed effects model with all machine-gun squad members and all types of personnel data does not work for the PIMG data set. Thus, we model each personnel variable with integration level separately with a random effect for who filled each position within the machine-gun squad. For example, age for each member of the machine-gun squad (three variables), a random effect for who filled each billet, and integration level are modeled with the result (response time or percentage hits) as the response variable. Where maximum likelihood estimation converged, AIC was used for variable selection. Otherwise, we comment on the significance of individual variables in the full model. Variables reported as significant are concluded to be significant based on at least a one-sided test.

G.6.3 PIMG Selected Tasks Overall Modeling Results

The weight for the entire machine-gun squad appears in the final reduced mixed effects model for the CASEVAC and M2 displacement with the potential influential point included. Weight is negatively correlated with the CASEVAC time but positively correlated with the M2 displacement time. The weight of each member of the machine-gun squad is highly correlated with integration level.

Rifle score for the entire machine-gun squad appears in the final reduced mixed effects model for the 7-km hike and M240B movement and emplacement and is positively correlated with both of these results. The rifle score for each patch number is highly correlated with integration level. The positive correlation with the response variable does not make sense from a theoretical perspective. Use caution when interpreting these results.

Integration level is significant for the M2 mount/dismount and M240B displace to LOA. For each task, modeling the random effects for the individuals participating in the task resulted in changes from the initial results in the descriptive statistics. Each respective task paragraph describes these changes.

The M240 engagement and M2 displacement without the potential influential point did not have a final model with any significant variables. Refer to Section G.5.2.3 and Section G.5.2.7.1.2, respectively, for these tasks' ANOVA results.

G.6.3.1 7-km Hike

We model elapsed time for the 7-km hike as a function of each personnel variable and integration level in a separate mixed effects model with a random effect for who filled each position within the machine-gun squad. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader,

- Age,
- Height,
- AFQT score,
- CFT MTC,
- CFT MANUF,
- PFT crunches,
- PFT 3-mile run.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 7-km hike time:

- GT score of patches 2 and 3,
- PFT crunches of patch 1,
- PFT 3-mile run of patches 1 and 2,
- Rifle score of patches 1, 2 and 3.

The following personnel variables are significant in their respective models and are negatively correlated with the 7-km hike time:

- Age of patch 1,
- Height of patches 1 and 3,
- Weight of patches 1, 2, and 3,
- CFT MTC of patch 1,
- CFT MANUF of patch 1.

The following personnel variables have no significant variables in their respective models:

- None

Because weight and rifle score were significant for all Marines, we fit a final mixed effects model with integration level, a random effect for who filled each billet, weight and rifle score. Table G-5 displays the coefficients and standard errors for the final model based on AIC results. All three weights are highly correlated (~0.76) with integration level, and all three rifle scores are highly correlated (~0.75) with integration level. Given the highly significant ANOVA results when treating integration level as the lone factor and considering the limited sample size of integrated squads, it calls into question the validity of the reduced model that includes the rifle score of all individuals. The positive relationship between rifle score and 7-km hike time does not make sense from a theoretical perspective. We report results for completion; however, use caution when interpreting the results.

Table G-5. Coefficients and Standard Errors for the Final 7-km Hike Model

Effect	Estimate	Std. Error
Rifle score patch 1	0.55	0.13
Rifle score patch 2	0.42	0.14
Rifle score patch 3	0.47	0.13

G.6.3.2 M240B Movement & Emplacement

We model elapsed time for the M240B movement and emplacement as a function of each personnel variable and integration level in a separate model in a mixed effects model where the random effect is who filled each position within the machine-gun squad. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader,
- Age,
- Height,
- Weight,
- AFQT score,
- PFT crunches,
- PFT 3-mile run.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the M240 movement and emplacement time:

- Age of patch 3,
- Weight of patch 3,
- CFT MTC of patch 3,
- Rifle score of patches 1, 2, and 3.

The following personnel variables are significant in their respective models and are negatively correlated with the M240B movement and emplacement time:

- None.

The following personnel variables have no significant variables in their respective models:

- GT score,
- CFT MUF.

Because rifle score was significant for all Marines, we fit a final mixed effects model with integration level, a random effect for who fills each billet, and rifle score. Table G-6 displays the coefficients and standard errors for the final model based on AIC results. All three rifle scores are highly correlated (~0.75) with integration level. Given the highly significant ANOVA results when treating integration level as the lone factor and considering the limited sample size of integrated squads, it calls into question the validity of the reduced model that includes the rifle score of all individuals. The positive relationship between rifle score and M240 movement and engagement time does not make sense from a theoretical perspective. We report results for completion; however, use caution when interpreting the results.

Table G-6. Coefficients and Standard Errors for the Final M240B Movement & Emplacement Model

Effect	Estimate	Std. Error
Rifle score patch 1	0.01	0.005
Rifle score patch 2	0.01	0.005
Rifle score patch 3	0.01	0.006

G.6.3.3 M240B Engagement

We model percent hits for the M240B engagement as a function of each personnel variable and integration level in a separate model in a mixed effects model where the random effect is who filled each position within the machine-gun squad. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the M240B engagement percent hits:

- Squad leader,
- GT score of patch 2,
- PFT 3-mile run of patch 2.

The following personnel variables are significant in their respective models and are negatively correlated with the M240B engagement percent hits:

- None.

The following personnel variables have no significant variables in their respective models:

- Age,
- Height,
- Weight,
- AFQT score,
- CFT MTC,
- CFT MUF,
- PFT crunches,
- Rifle score.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes only the intercept, meaning that we could not determine a mixed model that fits the data well. Refer to Section G.5.2.3 for this task to see the ANOVA result for differences between integration groups.

G.6.3.4 CASEVAC

We model elapsed time for the CASEVAC as a function of each personnel variable and integration level in a separate model in a mixed effects model where the random effect is who filled each position within the machine-gun squad. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- CFT MTC,
- CFT MUF.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader,
- Age,
- Height,
- AFQT score,
- PFT crunches,
- PFT 3-mile run,
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- GT score.

The following personnel variables are significant in their respective models and are positively correlated with the CASEVAC time:

- AFQT score of patch 1,
- GT score of patch 1, 2, and 3,
- Rifle score of patch 1.

The following personnel variables are significant in their respective models and are negatively correlated with the CASEVAC time:

- Weight of patches 1, 2, and 3.

The following personnel variables have no significant variables in their respective models:

- None.

Weight and GT score were significant for all Marines. Ideally, we would fit a final mixed effects model with integration level, GT and weight. Because GT score is missing for several Marines, AIC did not work for this variable. GT score is also highly correlated (~0.92) with integration level. Therefore, the final model includes integration level, a random effect for who filled each billet, and weight. Table G-7 displays the coefficients and standard errors for the final model based on AIC results. Weight is highly correlated (~0.90) with integration level. Given the highly significant ANOVA results when treating integration level as the lone factor and considering the limited sample size of integrated squads, it calls into question the validity of the reduced model that includes the weight of all individuals.

Table G-7. Coefficients and Standard Errors for the Final CASEVAC Model

Effect	Estimate	Std. Error
Weight patch 1	-0.05	0.01
Weight patch 2	-0.04	0.01
Weight patch 3	-0.03	0.02

G.6.3.5 M2 Mount/Dismount

We model elapsed time for the M2 mount/dismount as a function of each personnel variable and integration level in a separate model in a mixed effects model where the random effect is who filled each position within the machine-gun squad. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- Squad leader,
- Age,
- Height,

- AFQT score,
- PFT crunches.

The following personnel variables are significant in their respective models and are positively correlated with the M2 mount/dismount time:

- Weight of patches 2 and 3.

The following personnel variables are significant in their respective models and are negatively correlated with the M2 mount/dismount time:

- Height of patch 2,
- PFT 3-mile run of patch 1,
- Rifle score of patch 1.

The following personnel variables have no significant variables in their respective models:

- GT score,
- CFT MTC,
- CFT MUF.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes integration level only where HD has a difference of -0.50 minutes when compared to a C group and a statistically significant p-value of 0.06 in a one-sided test. This difference is a decrease from the -0.53-minute difference identified in the descriptive statistics, which is a 5.66% change.

G.6.3.6 M240B Displace to LOA

We model elapsed time for the M240B displace to LOA as a function of each personnel variable and integration level in a separate model in a mixed effects model where the random effect is who filled each position within the machine-gun squad. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader,
- Height,

- Weight,
- AFQT score,
- PFT crunches,
- PFT 3-mile run.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the M240B displace to LOA time:

- Age of patch 3,
- Height of patch 3,
- Weight of patch 3,
- CFT MTC of patch 3,
- CFT MUF of patch 3,
- PFT 3-mile run of patch 3.

The following personnel variables are significant in their respective models and are negatively correlated with the M240B displace to LOA time:

- None.

The following personnel variables have no significant variables in their respective models:

- GT score.

Because the effects of the personnel variables do not have any patters and their effects are often negligible, our final model includes integration level only where HD has a difference of 0.92 minutes when compared to a C group and a statistically significant p-value of 0.09 in a one-sided test. This difference is an increase from the 0.89-minute difference identified in the descriptive statistics, which is a 3.37% change.

G.6.3.7 M2 Displacement

G.6.3.7.1 M2 Displacement Overview

We model elapsed time for the M2 displacement as a function of each personnel variable and integration level in a separate model in a mixed effects model where the random effect is who filled each position within the machine-gun squad. The covariates in each model are the values of each personnel variable for each patch number. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

There was one influential point (HD group on trial cycle 1), and we model the M2 displacement time with and without this point. For this task, the influential point largely drives the conclusion that we draw about which variables affect the M2 displacement time. Including this point results in a model where height, weight and integration level are significantly correlated with the result. Excluding this point results in a model where no variables are consistently good predictors. The data point is valid, and there is no reason to exclude it; however, use caution when interpreting the results.

G.6.3.7.2 M2 Displacement with Influential Point

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader,
- Age,
- Height,
- Weight,
- AFQT score,
- GT score,
- PFT crunches,
- PFT 3-mile run,
- Rifle score.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- CFT MUF.

The following personnel variables are significant in their respective models and are positively correlated with the M2 displacement time:

- Squad leader,
- Age of patch 2,
- Height of patches 1, 2, and 3,
- Weight of patches 1, 2, and 3,

- CFT MTC of patch 1,
- CFT MUF of patch 1.

The following personnel variables are significant in their respective models and are negatively correlated with the M2 displacement time:

- PFT 3-mile run of patch 2,
- Rifle score of patch 2.

The following personnel variables have no significant variables in their respective models:

- None.

Because height and weight were significant for all Marines, we fit a final mixed effects model with integration level, a random effect for who filled each billet, height, and weight. Table G-8 displays the coefficients and standard errors for the final model based on AIC results. All three weights are highly correlated (~0.76) with integration level. Given the highly significant ANOVA results when treating integration level as the lone factor and considering the limited sample size of integrated squads, it calls into question the validity of the reduced model that includes the weights of all individuals.

Table G-8. Coefficients and Standard Errors for the Final M2 Displacement Model with Influential Point

Effect	Estimate	Std. Error
Integration level	0.72	0.25
Height patch 1	0.05	0.03
Weight patch 1	0.03	0.01
Weight patch 2	0.02	0.01
Weight patch 3	0.03	0.01

G.6.3.7.3 M2 Displacement without Influential Point

The models for the following variables do not run due to missing values:

- None.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader,
- Height,
- PFT 3-mile run.

The HD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- CFT MUF.

The following personnel variables are significant in their respective models and are positively correlated with the M2 displacement time:

- Age of patches 1 and 2,
- Height of patch 1,
- AFQT score of patch 1,
- CFT MTC of patch 1,
- CFT MUF of patch 1.

The following personnel variables are significant in their respective models and are negatively correlated with the M2 displacement time:

- PFT crunches of patch 2,
- PFT 3-mile run of patch 2.

The following personnel variables have no significant variables in their respective models:

- Weight,
- GT score,
- Rifle score.

We fit a final model with just integration level because there are no personnel variables significant for the whole machine-gun squad. Integration level is not significant in this model. We recommend referring back to ANOVA results in Section G.5.2.7 for differences between integration levels.

Appendix to Annex G
PIMG Supplemental Information

This appendix provides supplemental information for the PIMG portion of the GCEITF experiment. It provides additional descriptive and basic inferential statistics not described in Annex G.

Section 1. GCEITF Leadership Subjective Comments

The GCEITF leadership provided comments on their observations of the experiment throughout its execution. Table G A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table G A – Summary of GCEITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement			Requesting extra breaks			Requires extra assistance			Needs no assistance			Compensating for another Marine			Gear pass off			Other			No category			Total
		C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	
7-km Hike	M	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3
	F	0	6	6	0	3	3	0	3	3	0	0	0	0	0	0	0	6	6	0	0	0	0	1	1	19
	Unit	0	3	3	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	6
M240B Movement & Emplacement	M	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	F	0	3	3	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	5
	Unit	0	2	2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
M240B Engagement	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	3	3	6	2	2	4	0	0	0	0	0	0	0	0	0	1	0	1	11
CASEVAC to CPP	M	0	0	0	0	0	0	0	0	0	6	0	6	0	0	0	0	0	0	0	0	0	0	0	0	6
	F	0	2	2	0	0	0	0	0	0	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	6
	Unit	0	1	1	0	0	0	0	1	1	0	2	2	0	0	0	0	0	0	0	0	0	0	1	1	5
M2 Mount/Dismount	M	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Displace to LOA	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	2	2	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	3
	Unit	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
M2 Displacment	M	0	0	0	0	0	0	2	0	2	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	3
	F	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	Unit	0	1	1	0	0	0	5	3	8	3	2	5	0	0	0	0	0	0	0	0	0	1	3	4	18
1-km Hike	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Section 2. Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for 15 additional PIMG tasks. Annex G contains the descriptive statistics for the remainder of the PIMG tasks. The words “metric” and “task” are used interchangeably throughout this Appendix; they both refer to the experimental task.

The two tables below display the results for 15 additional PIMG metrics. Table G B displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent differences between integration levels.

Table G C displays ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare the two groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the HD group is different from that in the C group. We present basic inferential statistics for eight additional tasks.

Table G B – PIMG Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
1-km Hike (minutes)	C	19	10.49	0.60	7.66%
	HD	21	11.29	0.64	
7-km Hike; first km (minutes)	C	19	11.48	1.13	8.73%
	HD	17	12.48	1.43	
7-km Hike; second km (minutes)	C	19	11.99	0.97	24.23%
	HD	17	14.90	2.84	
7-km Hike; third km (minutes)	C	17	11.40	0.35	30.56%
	HD	16	14.89	2.83	
7-km Hike; fourth km (minutes)	C	18	12.91	2.55	40.19%
	HD	18	18.09	4.20	
7-km Hike; fifth km (minutes)	C	19	18.36	3.95	14.33%
	HD	19	21.00	3.61	
7-km Hike; sixth km (minutes)	C	19	15.16	3.21	42.80%
	HD	19	21.66	3.95	
7-km Hike; seventh km (minutes)	C	19	16.35	3.09	19.92%
	HD	20	19.60	3.47	

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
M240B Movement (minutes)	C	19	2.20	0.24	26.77%
	HD	21	2.79	0.31	
M240B Emplacement (minutes)	C	19	1.61	0.38	3.93%
	HD	20	1.67	0.44	
M240B Engagement Attack (hits on target)	C	17	99.76	64.98	-14.60%
	HD	20	85.20	42.20	
M240B Engagement C-Atk (hits on target)	C	16	94.81	47.60	10.22%
	HD	20	104.50	43.98	
M2 Mount (minutes)	C	19	4.80	0.71	-8.73%
	HD	21	4.38	0.73	
M2 Dismount (minutes)	C	19	0.98	0.21	-0.13%
	HD	21	0.98	0.28	
M2 hits on target (% hits)	C	19	25.53%	15.22%	59.06%
	HD	21	40.60%	14.25%	

Table G C – PIMG ANOVA and T-Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	2-Sided P-Value	1-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
1-km Hike (minutes)	16.56 (1, 38)	< 0.01*	HD-C	< 0.01*	< 0.01*	0.55	1.06	0.47	1.14
7-km Hike; first km (minutes)	5.47 (1, 34)	0.03*	HD-C	0.03*	0.01*	0.43	1.57	0.27	1.74
7-km Hike; second km (minutes)	17.61 (1, 34)	< 0.01*	HD-C	< 0.01*	< 0.01*	1.94	3.87	1.65	4.16
7-km Hike; third km (minutes)	25.48 (1, 31)	< 0.01*	HD-C	< 0.01*	< 0.01*	2.53	4.44	2.24	4.73
7-km Hike; fourth km (minutes)	20.06 (1, 34)	< 0.01*	HD-C	< 0.01*	< 0.01*	3.67	6.71	3.22	7.16
7-km Hike; fifth km (minutes)	4.59 (1, 36)	0.04*	HD-C	0.04*	0.02*	1.03	4.24	0.56	4.71
7-km Hike; sixth km (minutes)	30.86 (1, 36)	< 0.01*	HD-C	< 0.01*	< 0.01*	4.96	8.02	4.52	8.47
7-km Hike; seventh km (minutes)	9.52 (1, 37)	< 0.01*	HD-C	< 0.01*	< 0.01*	1.88	4.63	1.48	5.03
M240B Movement (minutes)	43.42 (1, 38)	< 0.01*	HD-C	< 0.01*	< 0.01*	0.47	0.70	0.44	0.74

Metric	F Statistic (df)	F Test P-Value	Comparison	2-Sided P-Value	1-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
M240B Emplacement (minutes)	0.23 (1, 38)	0.64	HD-C	0.64	0.32	-0.11	0.24	-0.16	0.29
M240B Engagement Attack (hits on target)	0.76 (1, 35)	0.42	HD-C	0.43	0.22	-38.70	9.57	-45.87	16.74
M240B Engagement C-Atk (hits on target)	0.40 (1, 34)	0.53	HD-C	0.53	0.27	5.74	13.63	-16.49	35.86
M2 Mount (minutes)	3.39 (1, 38)	0.07*	HD-C	0.07*	0.04*	-0.71	-0.12	-0.80	-0.04
M2 Dismount (minutes)	0.00 (1, 38)	0.99	HD-C	0.99	0.51	-0.10	0.10	-0.13	0.13
M2 hits on target (% hits)	10.47 (1, 38)	< 0.01*	HD-C	<0.01*	< 0.01*	0.09	0.22	0.07	0.23

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA.

Additional Task Results:

7-km Hike by km. The general trend for the 7-km hike was that the difference between the LD and C groups, the HD and C groups, and the HD and LD groups increased over the course of the hike.

1-km Hike. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.17 for the C group and 0.84 for the HD group.

The C group had a mean time of 10.49 minutes. This time is statistically significantly faster than the HD group mean time of 11.29 minutes. The HD group was 7.66% slower than the C group.

M240B Movement. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.18 for the C group and 0.81 for the HD group.

The C group had a mean time of 2.20 minutes. This time is statistically significantly faster than the HD group mean time of 2.79 minute. The HD group was 26.77% slower than the C group.

M240B Emplacement. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.50 for the C group and 0.48 for the HD group.

The C group had a mean time of 1.61 minutes. This time is faster (but not statistically significantly) than the HD group mean time of 1.67 minutes. The HD group was 3.93% slower than the C group.

M240B Engagement Attack. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.66 for the C group and 0.32 for the HD group.

The C group had a mean of 99.76 hits on target. This number is higher (but not statistically significantly) than the HD group mean of 85.2 hits. The HD group produced 14.60% fewer hits than the C group and the difference is not statistically significant in a t-test.

M240B Engagement C-atk. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.66 for the C group and 0.32 for the HD group.

The C group had a mean of 94.81 hits on target. This number is lower (but not statistically significantly) than the HD group mean of 104.5 hits. The HD group produced 10.22% more hits than the C group.

Mount M2 on Vehicle. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.94 for the C group and 0.07 for the HD group.

The C group had a mean time of 4.8 minutes. This time is statistically significantly slower than the HD group mean time of 4.38 minutes. The HD group was 8.73% faster than the C group.

Dismount M2 from Vehicle. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.68 for the C group and 0.01 for the HD group.

The C group had a mean time of 0.9798 minutes. This time is slower (but not statistically significantly) than the HD group mean time of 0.9786 minutes. The HD group was 0.13% faster than the C group.

M2 hits on target. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.04 for the C group and 0.90 for the HD group.

The C group had a mean percentage hit of 25.53%. This average is statistically significantly lower than the HD group mean of 40.60%. The HD group was 59.06% more accurate than the C group.

Annex H.
Field Artillery Cannoneer (MOS 0811)

This annex details the Field Artillery Cannoneer (MOS 0811) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 3 March – 11 April 2015 at training area Quakenbush, the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, California. The following sections outline the Field Artillery Cannoneer Scheme of Maneuver, Limitations, Deviations, Dataset Description, Descriptive and Basic Inferential Statistics, and Modeling Results.

H.1 Scheme of Maneuver

H.1.1 Experimental Cycle Overview

The GCEITF volunteers executed a variety of subtasks over the course of a 1-day trial cycle at the MCAGCC in Twentynine Palms, California. A record trial consisted of randomly selected gun sections with zero, 1, or 2 females integrated into the gun crews for the entire SOM. Volunteers were assigned to one of six billets on each gun: Assistant Gunner/#1, #2, #3, #4/5, Recorder, or Driver.

The daily SOM consisted of four phases. Phase one was a preparatory phase, where volunteers received their randomized gun and billet assignments. Phase two was Reconnaissance, Selection, and Occupation of Position (RSOP). This was the Section Chief's opportunity to familiarize each volunteer with his leadership style, the howitzer, and terrain surrounding the gun position. Phase three was the live-fire trial, as described in the following paragraphs. After the trial was complete, the battery entered phase four, where the battery prepared for the next day by sanitizing the position and conducting sustainment, resupply, and bivouac.

H.1.2 Experimental Details

The experimental daily SOM was designed around the mission of Marine Artillery and its function in a tactical environment. The mission of Marine Artillery is to furnish close and continuous fire support by neutralizing, destroying, or suppressing targets that threaten the success of the supported unit. The design of the experiment was to mirror an artillery section providing timely, accurate, and continuous-fire support. The evaluated subtasks were the most physically demanding and operationally relevant tasks that a cannoneer executes when operating the M777A2 howitzer. Initial Fatigue surveys were given to each volunteer at the beginning of each experimental day, followed by a final Fatigue survey at the completion of that particular day's tasks. Workload surveys were administered immediately following specified tasks. Finally, volunteers completed cohesion surveys at the end of each trial cycle. For survey instruments, see the GCEITF EAP Annexes D and M.

Objective and subjective measures of performance were captured for each task. The objective measures of performance included the time required to perform the task and the heart rate of volunteers while performing the task (reported separately). Subjective measures of performance and ability included Fatigue, Workload, and Cohesion Surveys conducted at various times during each trial cycle.

H.1.3 Additional Context

H.1.3.1 Training and Readiness Standards

Marine Corps artillery training and readiness standards are defined in NAVMC 3500.7A. It is critical to understand that the role of the T&R standard is not to determine whether the unit is performing adequately; rather, per the T&R Manual, “it identifies the minimum standards that Marines must be able to perform in combat.” Particularly in the realm of Fire Support, given the safe and proper operation of the howitzer and ancillary equipment, this standard should not be construed as “fast enough.” It is merely a tool by which commanders can train and evaluate a unit.

H.1.3.2 Individual Loads and Battery Timeline

Volunteers carried a standard 40-lb. load, which included a flak jacket with full set of SAPI plates, Kevlar helmet, six magazines, and individual first-aid kit. The artillery SOM included 6 days of trials followed by 1 rest day. Between trial days, the battery bivouacked in the training area. Personnel stayed in two-man tents with the standard-issue sleeping system. The Battery redeployed to Camp Wilson on the seventh day and returned the following day to execute follow-on 6-day cycles. All normal battery operations, such as logistics resupply and dunnage removal, were completed by the volunteers.

H.1.3.3 Fire Commands Standards

The artillery experiment conducted five trials concurrently in 1 day. All howitzers conducted each task at the same time, much like they would in a real training exercise. This method of execution was the safest and fastest way to complete the trials. If it were possible to have more time in the training area and to subject the volunteers to the same environmental variables and rest schedule, performing the SOM one howitzer section at a time would have been preferred. By completing the trials simultaneously, one howitzer had the ability to hamstring the entire gun-line’s SOM. To combat this effect, each task was queued manually by the Battery Gunnery Sergeant or the Fire Direction Center (FDC) with regimented 5-to-10-minute breaks after certain tasks. The FDC used a parallel sheaf, e.g., every howitzer on the gun-line used the same firing data and “At My Command” fire commands to ensure each howitzer began firing at the same time. The result was safe, as close to operational realism as the experiment could be, and with easily identifiable start and stop timelines.

H.1.3.4 Digital vs Manual

The Digital Fire Control System (DFCS) is a computer-based indirect-fire control system for the M777A2 howitzer. The DFCS provides gun location, navigation, and digital communications, and displays the direction and elevation of the howitzer. DFCS enhances a cannoneer's ability to perform many tasks rapidly. However, introducing the DFCS into the experiment would increase the likelihood of technical difficulties requiring a pause in the trial to troubleshoot. The DFCS requires many cables to be connected and disconnected rapidly. The cables and digital-communications equipment require a tempered and methodical procedure for employment. Because the crews would be randomized each day for all howitzers, the trials were executed without the aid of the DFCS.

H.1.4 Experimental Tasks

H.1.4.1 Emplacing the Howitzer

Emplacement of the howitzer involves bursts of speed, heavy lifting, entrenching, and safe, meticulous transitioning from subtask to subtask. It is the method by which a howitzer is changed from its towed configuration (attached to a Medium Tactical Vehicle Replacement (MTVR), also known as a prime mover) and seated on the deck, capable of firing. This task involves dismounting the prime mover, placing the trident bar on the deck, manually manipulating the trail arms, entrenching the spades, and setting up the aiming stakes 100 meters away. Batteries can expect to emplace three to four positions a day during a normal field-training evolution. The experiment captured two emplacement times during each trial.

H.1.4.2 Ammunition Handling

A significant portion of a cannoneer's time in the field is spent moving ammunition. The Artillery projectiles used during the experiment were approximately 40 inches long and weighed 105 pounds. The Marine Corps uses a separate, 7-ton truck to carry munitions. Additional stowage for complete rounds, which include the fuse, primer, and propellant, can be located in the prime mover using the Loose Projectile Restraint System (LPRS). The experiment required volunteers to offload and upload projectiles from both the ammunition truck and the LPRS.

In the operational environment, load ammunition can be distributed among howitzers dispersed hundreds of meters apart. A battery under resupply by an air-delivered sling-loaded pallet requires a howitzer section to move ammunition by foot over a long distance. To simulate these situations, the experiment required each individual to carry two projectiles 100 meters each.

H.1.4.3 Indirect-Fire Missions

The howitzer section is the final link between requested support and delivery of fires. When a unit requests artillery support in the operating forces, timeliness is of utmost importance. The task is physically demanding, specifically for the ammunition-handling billets, where they must manually pick up projectiles, place them on the loading tray, and ram the round into the breech while the tube is elevated. Each trial consisted of firing 21 projectiles in five fire missions throughout the day, at high and low angle.

H.1.4.4 Out-of-Traverse Limits

The left and right traverse limit on an emplaced howitzer is a cone of 45 degrees. To engage targets outside this cone, the crew pumps up the suspension and pivots the howitzer on its wheels by pushing the tube left or right as a team. Once aligned in the direction of the target, the crew will emplace the howitzer by digging in the trail arms and preparing for the fire mission. This task is common with deployed batteries supporting multiple units throughout a battlespace.

H.1.4.5 Gun Nets

To simulate concealing the howitzer’s position from enemy aerial observation, volunteers emplaced and displaced camouflaged nets using a sledgehammer and stakes. Volunteers were administered a workload survey during emplacement and displacement of the nets.

H.1.4.6 Displacement

Timely displacement enables a battery to deliver fire support to maneuver units and move to the next firing point before the enemy has the opportunity to conduct counter-battery fire. During this task, which occurred twice per day, volunteers dismantled the ammunition pit, attached the trident bar, pumped the howitzer to ride height, and dislodged the howitzer’s dug-in trail arms, transitioning the howitzer from a firing configuration to a towed configuration. Volunteers were administered workload and fatigue surveys during this event.

H.1.5 Loading Events

These loading events ensured, to the greatest extent possible, that all volunteers were equally fatigued, and simulated the intense physical and mental workload present in combat or in a combat-focused training environment. A separate howitzer was used for loading volunteers not randomly selected on any given day’s assessment. A regular full-day’s SOM was executed on this additional howitzer to accomplish proper loading.

H.1.6 Scheme of Maneuver Summary

Table H-1 displays number of trial cycles planned, executed, and analyzed by task. The artillery experiment executed 143 of 144 planned record trials over a 40-day period. A recorded trial occurred over the course of 1 day, during which a randomly selected 6-Marine gun section with zero, 1, or 2 females completed the entire SOM. The howitzer sections completed the SOM simultaneously to maximize time, safety, and similar conditions for each trial. The Marines remained in the field 6 days a week and completed all required tasks to sustain the battery outside the actual assessment. A trial consisted of repeatable tasks that a battery expects to conduct during field training or in the operational environment. Key experimental tasks included emplacing and displacing the howitzer, conducting indirect-fire missions, defending the battery, and moving and handling ammunition. When not selected to participate in a trial, volunteers completed the SOM on a loaded gun so they were equally fatigued.

H.2 Limitations

H.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a similar manner to those in an operational environment; however, artificial limitations or interruptions were introduced that changed or altered the normal performance of a task. Although these limitations represent a degree of artificiality, they do not detract significantly from the data-collection plan. The following limitations were observed for the 0811 Field Artillery Cannoneer assessment.

H.2.2 Relative Difficulty of Record Test

Though the tasks included in the artillery assessment were among the more physically strenuous experienced by 0811s in training and combat, the overall difficulty of the assessment was significantly less challenging than a comparatively long training exercise in the operating forces, much less combat. The entirety of the daily assessment, from billet assignment to final displacement, averaged approximately 5 hours. Following the assessment, volunteers conducted post-firing maintenance, as well as dunnage and resupply activities. One day a week was a rest day, during which the volunteers were trucked out of the training area for showers and laundry. In a regimental- or battalion-level exercise, Marines generally conduct battery activities, to include firing, moving, and conducting ancillary tasks, throughout the day and into the night. A cannoneer during these training exercises can expect to average 18-20 hours of work a day.

H.2.3 Qualified Section Chiefs

Leadership played a key role in task results. By controlling for leadership, performance differences between experimental groups can be more readily attributed to the volunteers. Section Chiefs slated by Manpower to the Battery were of disparate quality and, in some cases, uncertified (all Section Chiefs attained certification prior to the execution of the assessment). These direct-assignment Marines varied in their fundamental and implicit understanding of the howitzer, how to train junior Marines, and how to run the section at peak efficiency. Some Section Chiefs were noticeably less experienced than others, imparting performance disparity not attributable to the volunteers into the data. From a statistical standpoint, this did not hinder the ability to draw conclusions from results; given the spread of data collected over the variety of Section Chiefs, performance levels often were of sufficient consistency to attain statistical significance. Rather, the effect of this reality was imparting noise in the data that may obscure some performance trends, reducing their visibility and statistical significance. As discussed in the previous paragraph, these data should not be construed as effective benchmarks for timeliness or performance of tasks in regards to fire support standards.

H.2.4 Quantity of Howitzers in Training

The experiment was designed to assess an eight-Marine (six volunteers per gun) howitzer section in the executions of the SOM. Six firing howitzers and one backup provided the workup phase's approximately 40 volunteers the opportunity to maximize training time and experimental reality. In training, only three howitzers were used, with a fourth in reserve. For the first few months of training, sections comprised 15+ Marines per gun and, in the latter months, 8 Marines per gun, with half of the volunteers not training on a howitzer at any given time. This caused reduced realism in training and forfeited training opportunities, hobbling the effectiveness of the battery during the experiment.

H.2.5 Consistency of SOP in training

To compare task performance between sections and experimental groups, it is necessary that operating procedures are constant between sections and across the battery. During the pilot test, lack of a common SOP became apparent during the performance of tasks not well defined by the M777 technical manual. The Guns Platoon Sergeant was aggressive in attempting to create consistency across guns throughout the assessment as discrepancies arose. However, minute variations in operating procedures were frequent and difficult to detect.

H.2.6 Equipment

The howitzers used by the battery were of varied quality, but universally aged. Although the failure rate of the guns was not significantly higher than counterpart guns, frequent mechanical problems remained. These issues ranged from scavenge system failures to hydraulic fluid leaks to fire-linkage failures. Captured incidents of impacts are noted, but, in many cases, the nuances of each gun created uncaptured variations in section performance.

H.2.7 Terrain

The assessment called for two emplacements and displacements per day. In training and in combat, the result of the movement is, generally speaking, a new firing position. For the experiment, to allow for fluidity of communications and avoid the requirement to move the entire battery, only the guns emplaced and displaced. This saved time and minimized interruptions, but created unrealistic conditions. As time passed, the initial emplacement gun position for each howitzer became worn and uneven. The assessment team mitigated this by backblading each position with heavy equipment and making minor adjustments to gun placements, but constant digging quickly changed the consistency of the soil and the experience of digging, particularly from howitzer to howitzer.

H.3 Deviations

Deviations to the execution of the Field Artillery Cannoneer SOM were made and can be found in the Experimental Data Report signed May 2015; however, no deviations affected the analysis methodology outlined in the EAP.

H.4 Data Set Description

H.4.1 Data Set Overview

The Artillery experiment timeline included 4 pilot trial days and 29 record trial days. Pilot trial data was not used for analysis. Record trials were conducted as planned from 8 March through 11 April 2015.

H.4.2 Record Test Volunteer Participants

Forty Marine volunteers participated in the Artillery experiment, including 28 males and 12 females. One female volunteer dropped from the experiment before record trials began, leaving a total of 39 volunteers, 28 males and 11 females.

H.4.3 Planned, Executed, and Analyzed Trial Cycles

The original plan called for 6 trials to be captured in each cycle, equating to 144 planned trials for each task. Only 5 trials could be completed each day due to the volunteer population. In order to reach the 144 planned trials, all make-up days were used. Two trials were discarded because of a howitzer recoil malfunction on 9 March and an ankle injury on 19 March. Part of a trial on 3 April was also discarded because the gunner's sights were not secured to the howitzer. Any other lost trial data is due to data collector or Toughbook program error, terrain, improper procedure, or equipment malfunction that led to an invalid data point.

Table H-1. 0811 Planned, Executed, and Analyzed Trial Cycles[1]

Task	Number of planned observations	Number of executed observations	Number of observations used in analysis
3-Round Low-Angle WP; 1st Round Response	144	136	121
3-Round Low-Angle WP; Indiv. Subsequent Rounds	144	140	125
3-Round Low-Angle WP; Total Fire Mission	144	136	121
3-Round High-Angle WP; 1st Round Response	144	139	122
3-Round High-Angle WP; Indiv. Subsequent Rounds	144	141	132
3-Round High-Angle WP; Total Fire Mission	144	139	123
3-Round Low-Angle HE from FP2; 1st Round Response	144	132	105
3-Round Low-Angle HE from FP2; Indiv. Subsequent Rounds	144	140	121
3-Round Low-Angle HE from FP2; Total Fire Mission	144	132	103
3-Round Low-Angle HE; 1st Round Response	144	137	121
3-Round Low-Angle HE; Indiv. Subsequent Rounds	144	142	133
3-Round Low-Angle HE; Total Fire Mission	144	137	120
9-Round Low-Angle HE; 1st Round Response	144	137	110
9-Round Low-Angle HE; Indiv. Subsequent Rounds	144	140	128
9-Round Low-Angle HE; Total Fire Mission	144	137	107
Ammo Prep Ammo; Truck Offload	144	142	141
Ammo Prep; LPRS Offload	144	142	140
Ammo Resupply; Ammo Movement	144	143	141
Displacement of Howitzer; Ammo Truck Upload	144	140	136
Displacement of Howitzer; LPRS Upload	144	142	138
Displacement of Howitzer; Total Displacement	144	139	126
Displacement of Howitzer; Trail Arms	144	138	132
Displacement of Howitzer; Trident Bar	144	142	137
Emplacing Howitzer; Dismount	144	143	136
Emplacing Howitzer; Total Emplacement	144	143	135
Emplacing Howitzer;Trident Bar Drop	144	143	137
Final Displacement of Howitzer; Mount Prime Mover	144	143	104
Final Displacement of Howitzer; Total Displacement	144	143	121
Final Displacement of Howitzer; Trail Arms	144	143	131
Final Displacement of Howitzer; Trident Bar	144	143	133
Gun Nets	144	141	139

Hasty Ammo Prep; LPRS Offload	144	141	140
Hasty Emplacement of Howitzer; Dismount	144	142	136
Hasty Emplacement of Howitzer; Total Emplacement	144	141	132
Hasty Emplacement of Howitzer; Trident Bar Drop	144	142	141
Speed Shift #1; First Drop	144	141	139
Speed Shift #2; First Drop	144	141	132
Strike Nets	144	139	136

[1]Some data was not captured or captured incorrectly due to human (Data Collector - DC) error or data processing equipment (Toughbook) error. Other data points were classified as outliers or potential influential points and were excluded from the analysis as described in the methodology section.

H.5 Descriptive and Basic Inferential Statistics

H.5.1 Descriptive Statistics Overview

This report accounts for various Artillery Section combinations when integrating MOS qualified female Marines with MOS qualified male Marines. It is important to quantify the billets within a Howitzer Section and to understand which billets a Marine could be expected to fill. These billets will be mentioned throughout the Annex. A Howitzer Section often consists of eight Marines. The section positions include Section Chief, Gunner, #1 Cannoneer/A-Gunner, #2 Cannoneer, #3 Cannoneer, #4/5 Cannoneer, Recorder, and Driver. The Section Chief is responsible for everything that happens on his Howitzer. His responsibilities are technical and require extensive MOS experience. The Gunner position also requires technical knowledge. He is responsible for giving commands and using the instruments to aim the howitzer. Because the nature of these positions is supervisory, requires leadership and experience, and is not very physical, these two billets were filled by direct assignments and were not moved to different howitzers during the experiment. Each of the other six billets was filled by a volunteer and randomized daily. For a complete list of each of the duties of the cannoneer, read the Marine Corps TM 10407C-OR/1. This experiment uses the 7:1 configuration, meaning 7 cannoneers, Gunner included, and 1 Section Chief on each gun. This is a common configuration used by operational units, which are generally constrained by the number of cannoneers available.

This report also refers to various concentrations of Howitzer Sections in terms of integration levels. A control group (C group) refers to an all-male crew filling any combination of the six volunteer billets; this type of crew was the control for the experiment. The low-density crew (LD group) was made up of five males and one female filling any combination of the six billets. The high-density crew (HD group) included four males and two females filling any combination of the six billets. The Marines were rotated among the billets by random selection without respect to MOS knowledge or experience. This was done in an effort to examine the effects of integration at all crew levels and to allow Marines to step in and fill any of the six billets at any given time.

In addition to integration level comparisons, there are certain tasks that were analyzed by what is referred to as “critical billet.” Though most tasks are to be completed as a section, where Marines can assist other Marines, it is important to understand that certain tasks, such as the conduct of fire missions, require each cannoneer to perform his or her duties. Compensation for a Marine failing to do his or her tasks results in slower performance and unresponsive fires. A critical billet was identified for tasks where a physically demanding duty was the bottleneck for a section. For example, the #4 cannoneer is the loader and must carry each round to the howitzer. The rest of the section relies on the #4 to complete his or her task before they can proceed with their own. Each critical billet is explained in the results that follow.

In every case, the integration concentration level was evaluated for statistical differences in performance, but in only select cases were critical billets considered. In these cases, the bottleneck in task performance was such that, were it not identified as a variable in success, the impact of integration might be masked. For instance, in the case of the speed shift, or out of traverse, two Marines are executing the preponderance of physical labor: numbers 1 and 2 cannoneers. These Marines are responsible for pumping the suspension of the howitzer, a particularly demanding task. Were analysis to be conducted based purely on integration level, the contribution of, as an example, the recorder could be misinterpreted as providing an inordinate impact on mission accomplishment.

All the tasks selected for this experiment were based on Training and Readiness (T&R) Manual requirements for Artillery Training NAVMC 3500.7, Marine Corps Warfighting Publication 3-16.3 TTP for the field Artillery Cannon Gunnery, Marine Corps TM 10407C-OR/1, and guidance from Plans Policies & Operations (PP&O), Training and Education Command (TECOM), and Operating Forces Units. In addition, the tasks selected for the experiment were deemed physically demanding, repeatable, and commonly performed under time-constrained conditions in training and potentially in a combat environment.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA), Tukey Tests (or non-parametric equivalent as necessary), and scatter plots. The first table displays the metric, integration levels, sample sizes, means, and standard deviations. The second table shows ANOVA and Tukey test results including, but not limited to, metrics, p-values suggesting statistical significance, integration level elapsed time differences, and percentage differences between integration levels. If non-parametric tests were needed, the second table displays these results instead of ANOVA and Tukey test results. Subsequent subsections cover each task in detail, along with scatterplots of the data. If p-values are less than the a-priori significance level of 0.10, we conclude that there is statistical evidence that the mean elapsed time for the experimental groups, LD and HD, are different from the C group

H.5.1.1 Emplacement of Howitzer by Integration Level

The emplacement task is the link between tactical march and the battery being fire capable. During a tactical march, the howitzer is in a towed configuration and is moved by a prime mover over long distances. In order for the howitzer to provide fire support to friendly units in an area of operations, the howitzer must emplace.

To assess the proficiency of the cannon crewman sections conducting an emplacement, time hacks for certain subtasks were captured and used in the metric calculations below. These subtasks included the howitzer being disconnected from the prime mover, the command “ready drop” given by the gunner, the #4 cannoneer removing the trident bar, and the entrenchment of the spades. The emplacement time started when the Section Chief gave the order to begin and finished when the section chief ordered the offload of ammunition. The total time for this task did not include the time it took for the direct assignment Section Chief and battery staff to lay the howitzer with the M2A2 aiming circles. Because the laying process does not involve physical labor or the volunteers, and it was subtracted from the total times.

The experiment captured two separate emplacements during each trial. The first emplacement began at the first firing position with volunteers seated in the back of the prime mover and the howitzer in a towed configuration. The second emplacement time began with the volunteers moving 300 meters by foot to a supplemental firing position where the howitzer halted in the towed configuration. The only difference between the two emplacements was the position where volunteers began the task. The results for total emplacement times for both positions are displayed in the tables below. See the Artillery Appendix for dismount, repositioning, and trident bar calculations. Statistical analysis included analysis of variance (ANOVA) and Tukey tests, which is a multiple comparison statistical test.

H.5.1.2 Emplacement of Howitzer by Integration Level Scatterplots

The scatterplots in figures H-1 and H-2 display the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling.

Figure H-1. Emplacement of Howitzer by Integration Level

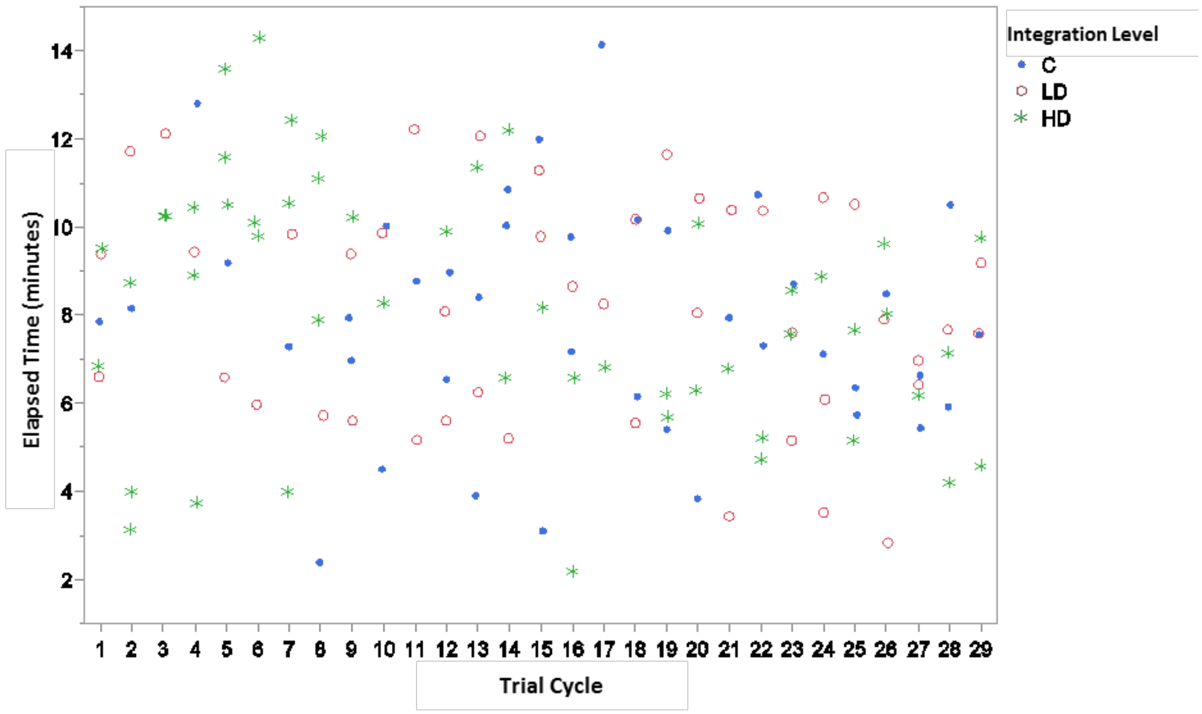
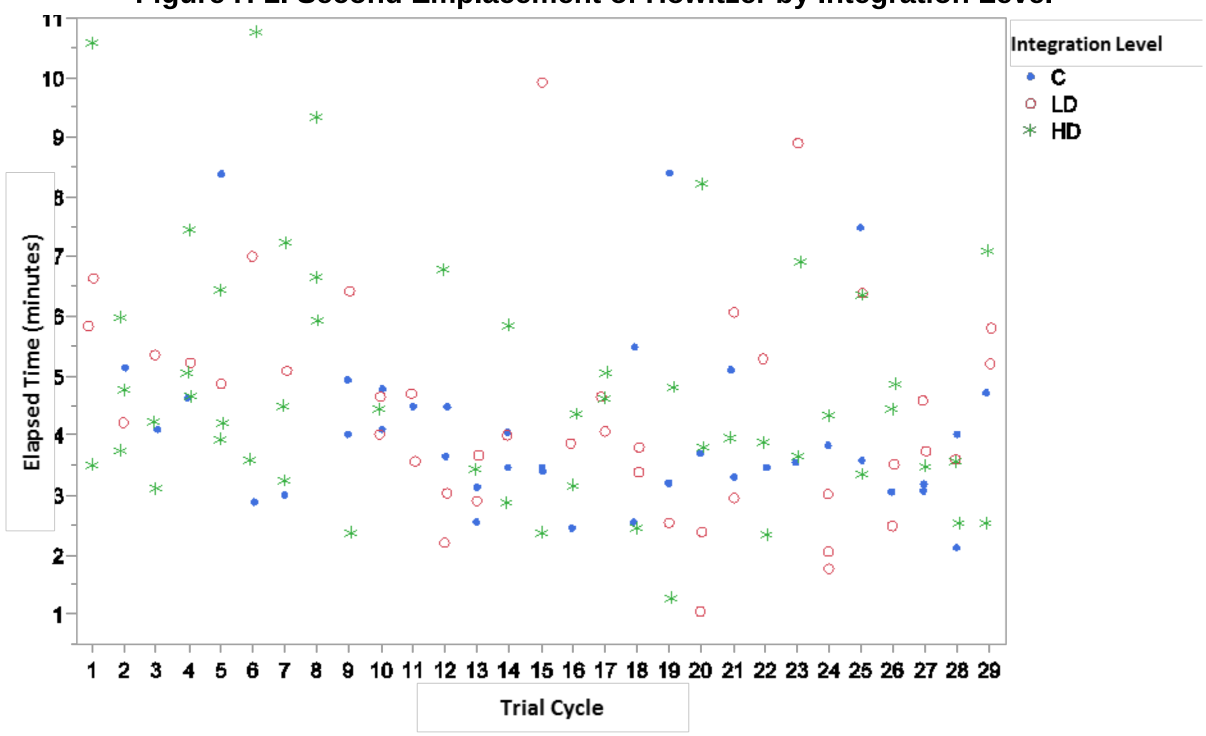


Figure H-2. Second Emplacement of Howitzer by Integration Level



H.5.1.3 Emplacement of Howitzer Data Tables and Analysis

The tables below summarize the results of the task Emplacement of Howitzer. Table H-2 compares means across metrics and integration levels. Table H-3 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-2. Emplacement of Howitzer by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
First Emplacement of Howitzer; Total Emplacement (minutes)	C	40	7.86	2.58
	LD	44	8.12	2.53
	HD	51	8.19	2.83
Second Emplacement of Howitzer; Total Emplacement (minutes)	C	38	4.08	1.44
	LD	42	4.39	1.81
	HD	52	4.76	2.06

Table H-3. Emplacement of Howitzer by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
First Emplacement Howitzer; Total Emplacement (minutes)	0.18 (2, 132)	0.83	LD-C	0.25	3.21%	0.90	-0.75	1.26	-0.95	1.46
			HD-C	0.33	4.20%	0.83	-0.64	1.30	-0.84	1.49
			HD-LD	0.08	0.95%	0.99	-0.87	1.02	-1.06	1.21
Second Emplacement of Howitzer; Total Emplacement (minutes)	1.60 (2, 129)	0.21	LD-C	0.31	7.67%	0.72	-0.39	1.01	-0.53	1.16
			HD-C	0.69	16.87%	0.18	0.02	1.36	-0.12	1.49
			HD-LD	0.38	8.55%	0.58	-0.28	1.03	-0.41	1.16

H.5.1.3.1 First Emplacement of Howitzer

The data are normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of 1.00 for the C group, 0.19 for the LD group, and 0.77 for the HD group. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

For the first emplacement, the tables above show that, on average, the C group had a mean time of 7.86 minutes, compared to the LD group and HD group, who had a mean time of 8.12 minutes (15 seconds slower than the C group) and 8.19 minutes (15 seconds slower than the C Group), respectively. The C group was faster, on average, than both LD and HD groups. Neither of these comparisons was deemed statistically significant.

H.5.1.3.2 Second Emplacement of Howitzer

The C group data and HD group are not normally distributed, as evidenced by the Shapiro-Wilk Test that results in a p-value of < 0.01, but LD group data is normally distributed with a p-value of 0.10. We proceed with presenting ANOVA results because sample sizes are sufficiently large ($n > 30$) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

For the second emplacement, the tables above show, on average, that the C group had a mean time of 4.08 minutes compared to the LD group and HD group, which had mean times of 4.39 minutes (18 seconds slower than the C Group) and 4.7 minutes (40 seconds slower than the C Group), respectively. The C group executed the task faster, on average, than both LD and HD groups, but neither comparison was statistically significant.

H.5.1.4 Emplacement of Howitzer by Integration Level Contextual Comments

Although, in both emplacements, the differences between the integrated groups and the control groups were not statistically significant, speed and consistency in the execution of this task is critical to the success of a battery. Although the task of howitzer emplacement is expected to be executed expeditiously, safety and precision must also be present to guarantee the well-being of Marines and the accuracy of supporting fires. Additional time to complete an emplacement task leads to non-responsive fires, which has the potential to give the enemy a significant advantage in fire superiority. Speed directly affects the battery's mission to be fire capable in order to achieve fire superiority over the enemy.

H.5.2 Ammunition Handling by Integration Level

As stated in the Artillery SOM, a significant portion of a cannoneer's time in the field is spent moving ammunition. This task is physically demanding, as it requires the Marines to lift a 155-mm round, weighing 106 lbs., from the ground to their shoulder and maintain positive control while transporting the round 100 meters. Ammunition offloading in this experiment was regimented in order to capture the differences among groups. In an operational environment, ammunition is constantly being shifted, redistributed, and organized. In addition, a common practice is to spread-load munitions among howitzers, moving them from howitzer to howitzer as necessary. The ammunition-handling tasks were broken down into five subtasks: Ammo Truck Offload, Ammo Truck Upload, LPRS Offload, LPRS Upload, and Ammo Resupply.

H.5.2.1 Ammunition Truck Offload/Upload by Integration Level Scatterplots

The scatterplots in figures H-3 and H-4 display the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling. Influential points are identified by dark solid black circles.

Figure H-3. Ammunition Truck Offload by Integration Level

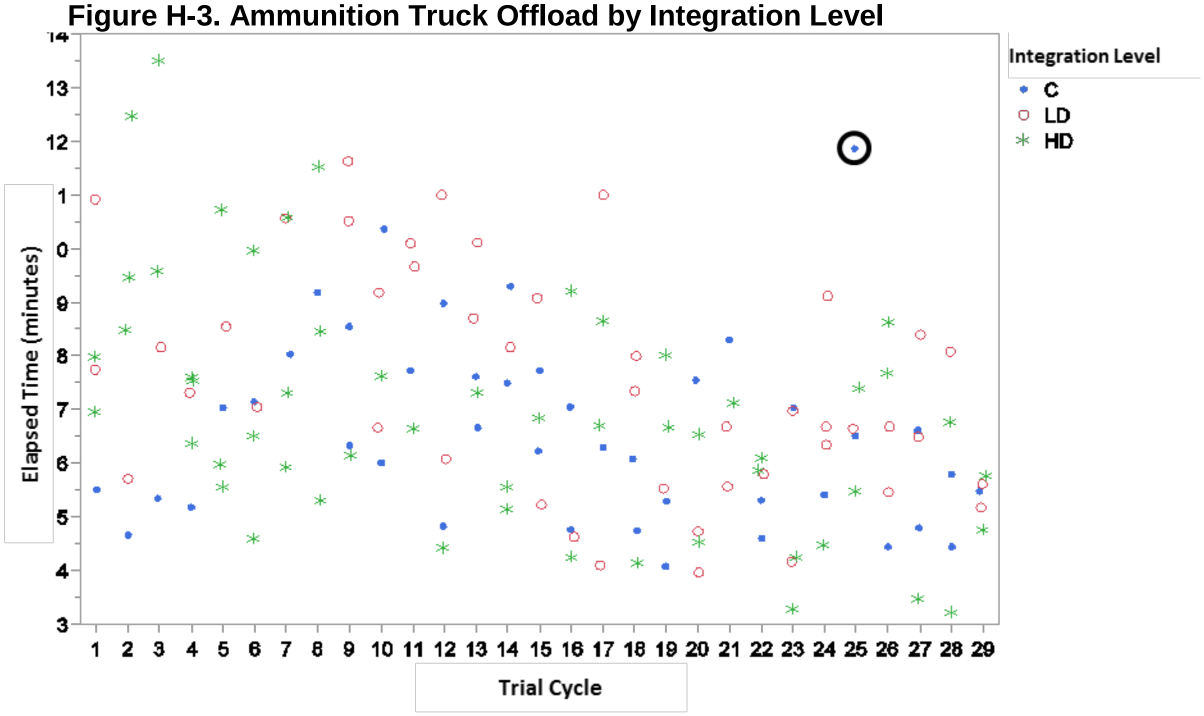
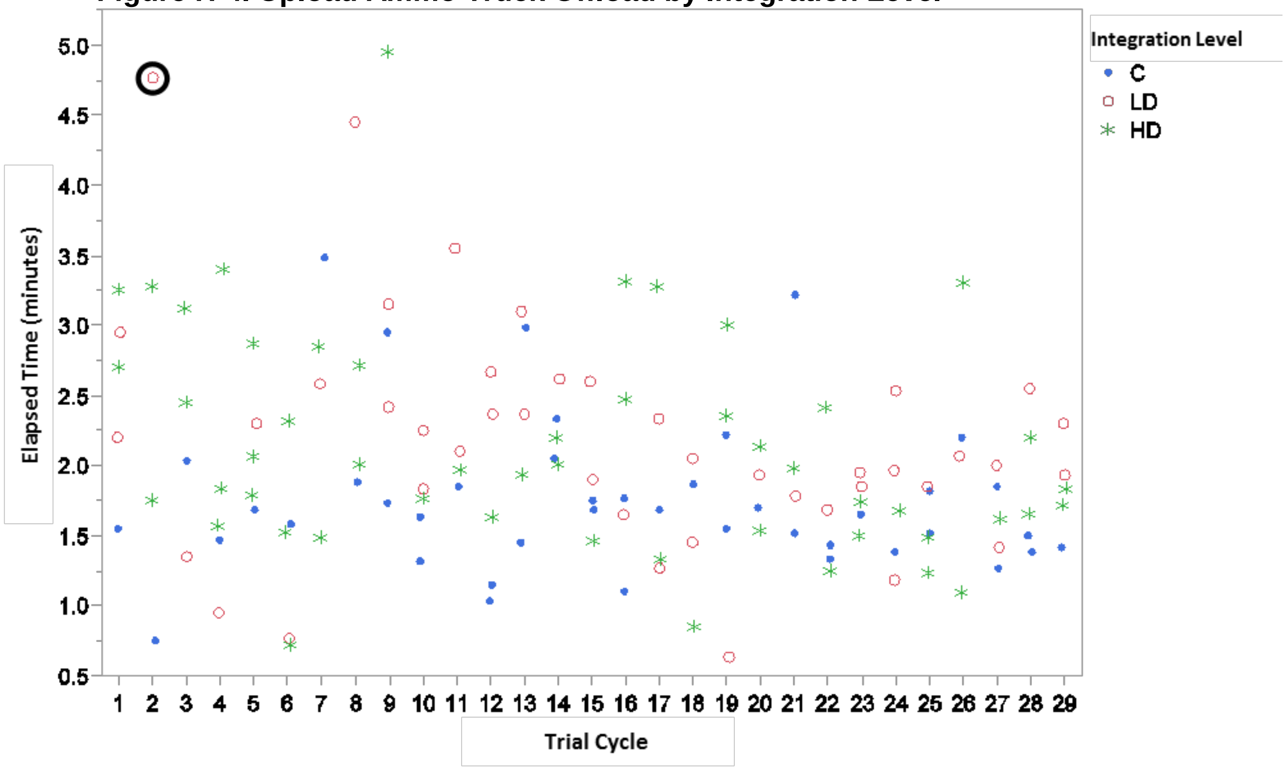


Figure H-4. Upload Ammo Truck Offload by Integration Level



H.5.2.2 Ammunition Truck Offload/Upload by Integration Level Data Tables and Analysis

The tables below summarize the results of the subtasks, Ammo Truck Offload and Ammo Truck Upload. Table H-4 compares means across metrics and integration levels. Table H-5 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-4. Ammunition Truck Offload/Upload by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Truck Offload (minutes)	C	42	6.58	1.75
	LD	45	7.45	2.10
	HD	54	6.94	2.27
Truck Offload [Excluding Influential Points] (minutes)*	C	41	6.45	1.55
	LD	45	7.45	2.10
	HD	54	6.94	2.27
Ammo Truck Upload (minutes)*	C	42	1.76	0.56
	LD	43	2.18	0.82
	HD	51	2.13	0.80

Ammo Truck Upload [Excluding Influential Points] (minutes)*	C	42	1.76	0.56
	LD	41	2.06	0.63
	HD	50	2.07	0.69

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table H-5. Ammunition Truck Offload/Upload by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Truck Offload (minutes)	1.95 (2, 138)	0.15	LD-C	0.87	13.22%	0.13	0.10	1.64	-0.05	1.79
			HD-C	0.36	5.46%	0.68	-0.38	1.09	-0.52	1.24
			HD-LD	-0.51	-6.86%	0.44	-1.23	0.21	-1.38	0.35
Truck Offload [Excluding Influential Points] (minutes)	2.61 (2, 137)	0.08*	LD-C	1.00	15.48%	0.06*	0.24	1.75	0.09	1.91
			HD-C	0.49	7.56%	0.48	-0.24	1.21	-0.38	1.36
			HD-LD	-0.51	-6.86%	0.43	-1.22	0.19	-1.36	0.34
Ammo Truck Upload (minutes)	4.19 (2, 133)	0.02*	LD-C	0.42	24.04%	0.03*	0.14	0.70	0.09	0.75
			HD-C	0.37	21.25%	0.04*	0.11	0.64	0.05	0.69
			HD-LD	-0.05	-2.25%	0.95	-0.31	0.22	-0.37	0.27
Ammo Truck Upload [Excluding Influential Points] (minutes)	3.45 (2, 130)	0.03*	LD-C	0.30	17.28%	0.08*	0.06	0.54	0.01	0.59
			HD-C	0.32	18.03%	0.05*	0.09	0.55	0.04	0.59
			HD-LD	0.01	0.64%	0.99	-0.22	0.24	-0.26	0.29

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

The first subtask, Ammunition Truck Offload, consisted of four cannoneers offloading 20 M795 High Explosive (HE) rounds and six M825 Smoke Rounds from the bed of an MTVR to an ammunition pit on the ground 10 yards away. The M795 and M825 are ballistically and dimensionally similar and offer the same experience in moving the projectile manually. This task occurred once during each trial. The start of the event was recorded by the Section Chief's command to offload, and the stop cue was recorded after all rounds were on the deck. The #3 and #4 Cannoneers were not involved in this task as they were concurrently completing the LPRS task.

The data are normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of 0.03 for the C group, 0.16 for the LD group, and 0.08 for the HD group. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

On average, the C group had a mean time of 6.58 minutes, compared to the LD group and HD group, which had a mean time of 7.45 minutes and 6.94 minutes, respectively. Neither comparison is statistically significant. However, the C group was faster, on average, than both the LD and HD groups.

After excluding influential points, the average of the C group drops to 6.45 minutes. This comparison with the LD group becomes significant; however, the HD group remains not statistically significant.

H.5.2.2.1 Ammunition Truck Upload by Integration Level Overview and Results

The second subtask, Ammunition Truck Upload, consisted of moving eight HE rounds from the ammunition pit back onto the bed of the truck. This task occurred once during each trial. The start of the event was recorded by the Section Chief's command to upload the ammunition, and the stop cue was recorded after all rounds were on the truck. The #3 and #4 Cannoneers were not involved in this task as they were concurrently completing the LPRS task.

The C group data and HD group are not normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of < 0.01, but the LD group data is normally distributed with a p-value of 0.01. We proceed with presenting ANOVA results because sample sizes are sufficiently large ($n > 30$) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

On average, the C group had a mean time of 1.76 minutes; the LD group, 2.18 minutes; and the HD group, 2.13 minutes. Both the LD and HD groups are statistically significant when compared to the C group. The C group is faster, on average, than both the LD and HD groups.

Excluding Influential Points reveals the LD group having a mean time of 2.06 minutes and the HD group having a mean time of 2.07 minutes. The C group remains the same, but is still faster and is still statistically significant when compared to the LD and HD groups.

H.5.2.2.2 Ammunition Truck Contextual Comments

Unloading or moving ammunition slower than a howitzer is firing has the potential to result in untimely fires. Cannoneers will rarely have all the time they need to keep ammunition orderly, remove dunnage, and keep the ready board fully stocked. Keeping up with these tasks is continuous and must be accomplished as quickly as possible so as to not delay a fire mission.

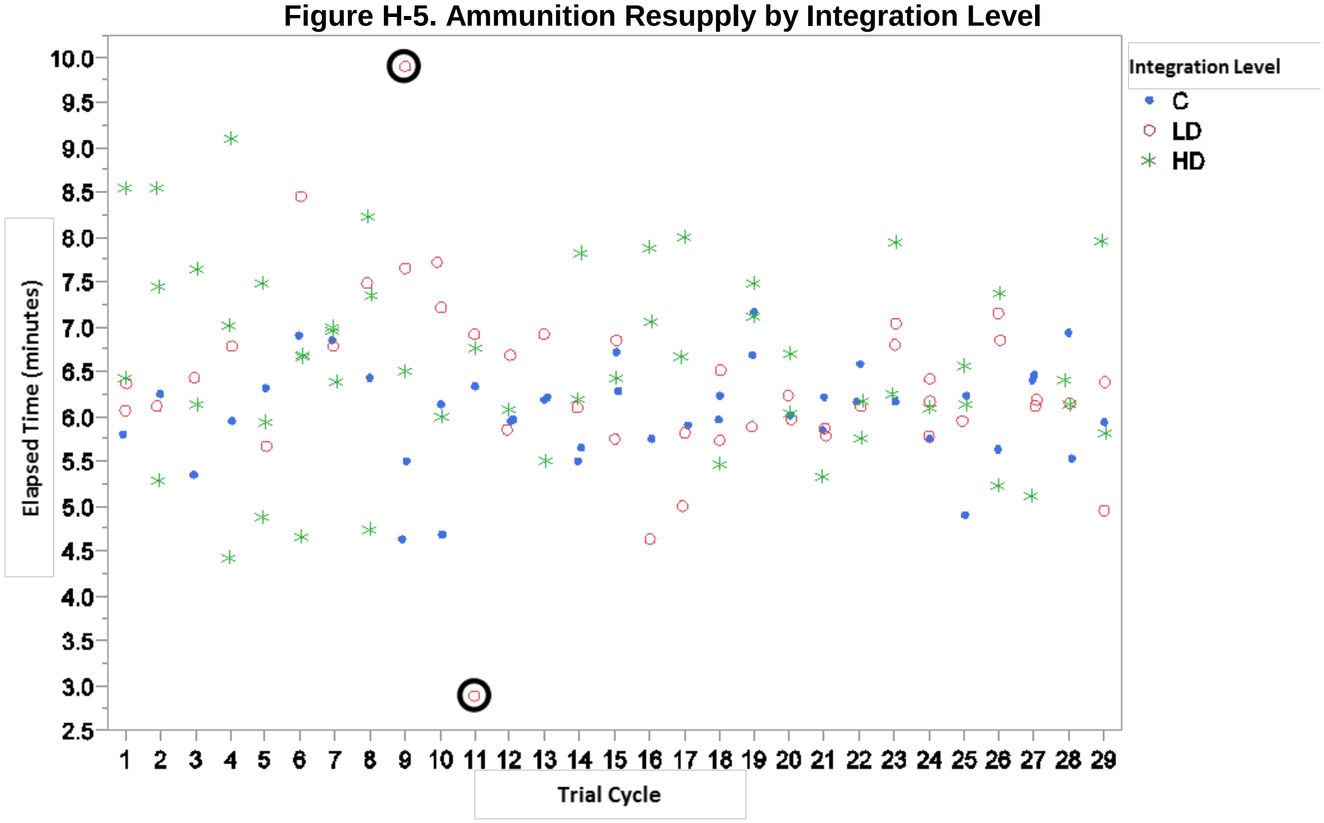
H.5.2.3 Ammunition Resupply by Integration Level

In the ammunition resupply task, all volunteers were required to carry two individual projectiles for 100 meters. This was completed by moving to a turnaround point 50 meters away. The start time began when all six volunteers stepped off, and it stopped when they returned with the projectile.

H.5.2.4 Ammunition Resupply by Integration Level Scatterplot

The scatterplot in Figure H-5 displays the data used in the analysis of the results. All data points shown in the scatterplot were determined to be valid and used in the analysis and modeling.

Figure H-5. Ammunition Resupply by Integration Level



H.5.2.5 Ammunition Resupply by Integration Level Data Table and Analysis

The tables below summarize the results of the task Ammunition Resupply. Table H-6 compares means across metrics and integration levels. Table H-7 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-6. Ammunition Resupply by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Ammo Resupply; Ammo Movement (minutes)*	C	42	6.05	0.56
	LD	45	6.36	1.04
	HD	54	6.58	1.06
Ammo Resupply; Ammo Movement [Excluding Influential Points] (minutes)*	C	42	6.05	0.56
	LD	43	6.36	0.74
	HD	54	6.58	1.06

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table H-7. Ammunition Resupply by Integration Level

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Ammo Resupply; Ammo Movement (minutes)	3.83 (2, 138)	0.02*	LD-C	0.31	5.08%	0.28	-0.04	0.65	-0.11	0.72
			HD-C	0.53	8.78%	0.02*	0.20	0.86	0.13	0.93
			HD-LD	0.22	3.52%	0.46	-0.10	0.55	-0.17	0.61
Ammo Resupply; Ammo Movement [Excluding Influential Points]	4.74 (2, 136)	0.01*	LD-C	0.31	5.06%	0.22	-0.01	0.62	-0.07	0.68
			HD-C	0.53	8.78%	< 0.01*	0.23	0.83	0.17	0.89
			HD-LD	0.23	3.55%	0.39	-0.07	0.50	-0.13	0.58

(minutes)										

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

The C group and HD group data are normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of 0.20 and 0.88, but the LD group data is not normally distributed with a p-value of < 0.01. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The tables above show that, on average, the C group had a mean time of 6.05 minutes; the LD group, 6.36 minutes; and the HD group, 6.58 minutes. The C group is faster, on average, than both the LD and HD groups. The difference between the C and HD groups is statistically significant.
Excluding Influential Points reveals the LD group's mean time remains the same; however, the standard deviation is reduced. The effect is minimal on the comparisons and remains not statistically significant.

H.5.2.5.1 Ammunition Resupply by Integration Level Contextual Comments

When ammunition must be moved over a long distance, such as 100 meters or longer, every Marine in the position is expected to assist in the task. In training and combat, movement of ammunition is usually done when there is a pause in firing or not when it can be done without hindering fire mission response time. However, immediate resupply being conducted while requests for fires are being called to the battery can require this task be done as swiftly as possible. This task is physically demanding and requires Marines to share the burden equally.

H.5.2.6 Loose Projectile Restraint System (LPRS) Offload / Upload by Integration Level

Two LPRS offloads and one LPRS upload were completed during the cycles. The offload task is physically demanding because it requires #3 and #4 Cannoneers to move three M795 High Explosive Rounds from the LPRS on the bed of the MTVR to the ammunition pit on the ground 10 yards away. The upload is the reverse of the offload, i.e., stowing rounds from the ammunition pit into the truck. During this task, #3 was located in the prime mover while #4 was on the ground. These two Marines were the only Cannoneers participating in this task and therefore were deemed as critical billets.

The data collected from this task was analyzed in two separate ways. The first method of comparison looked at performance differences between the control group (C), the low-density group (LD), and a high-density group. The second method of comparison looked at performance differences between gun crews grouped only by the gender of the two critical billets—Cannoneer #3 and #4 (regardless of the crew composition).

H.5.2.6.1 LPRS Offload / Upload by Integration Level Scatterplots

The scatterplots in figures H-6, H-7, and H-8 display the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling.

Figure H-6. LPRS Offload by Integration Level

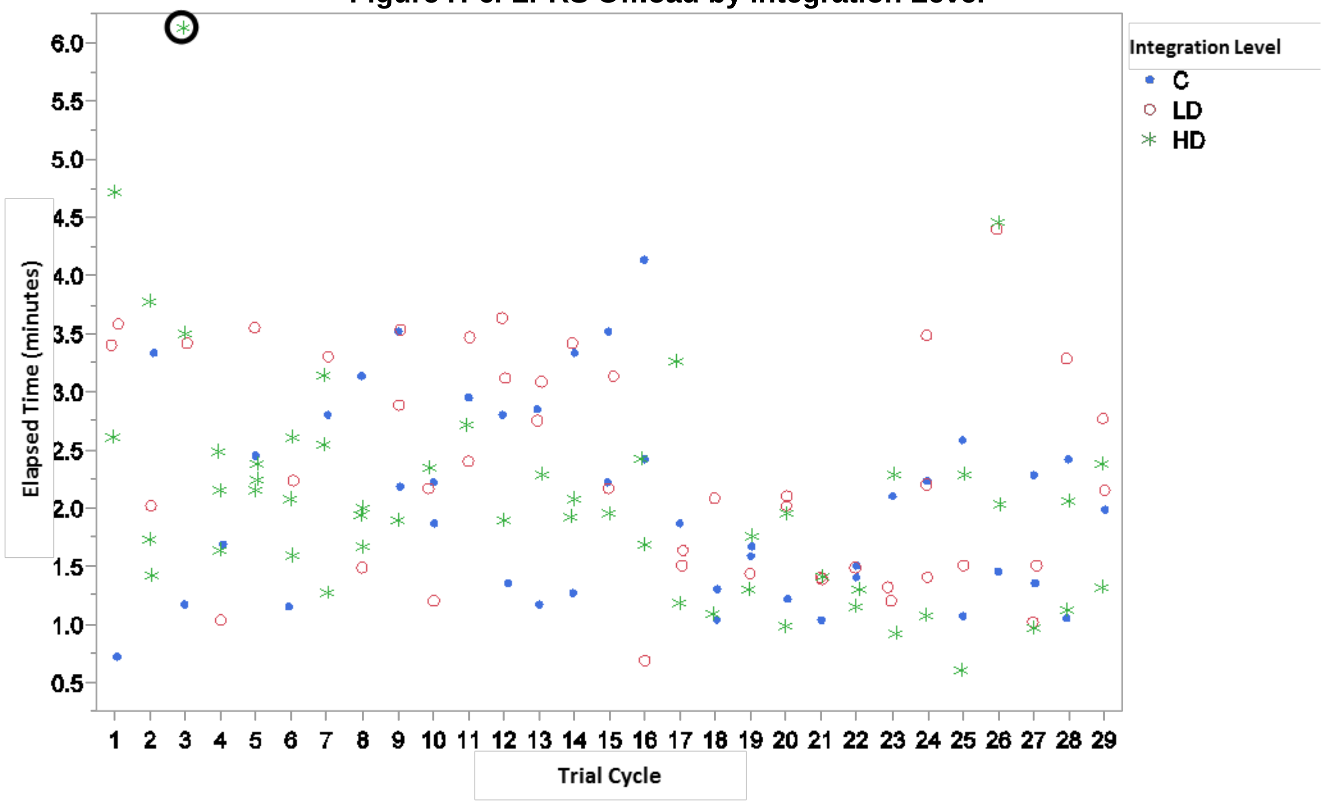


Figure H-7. Upload by Integration Level

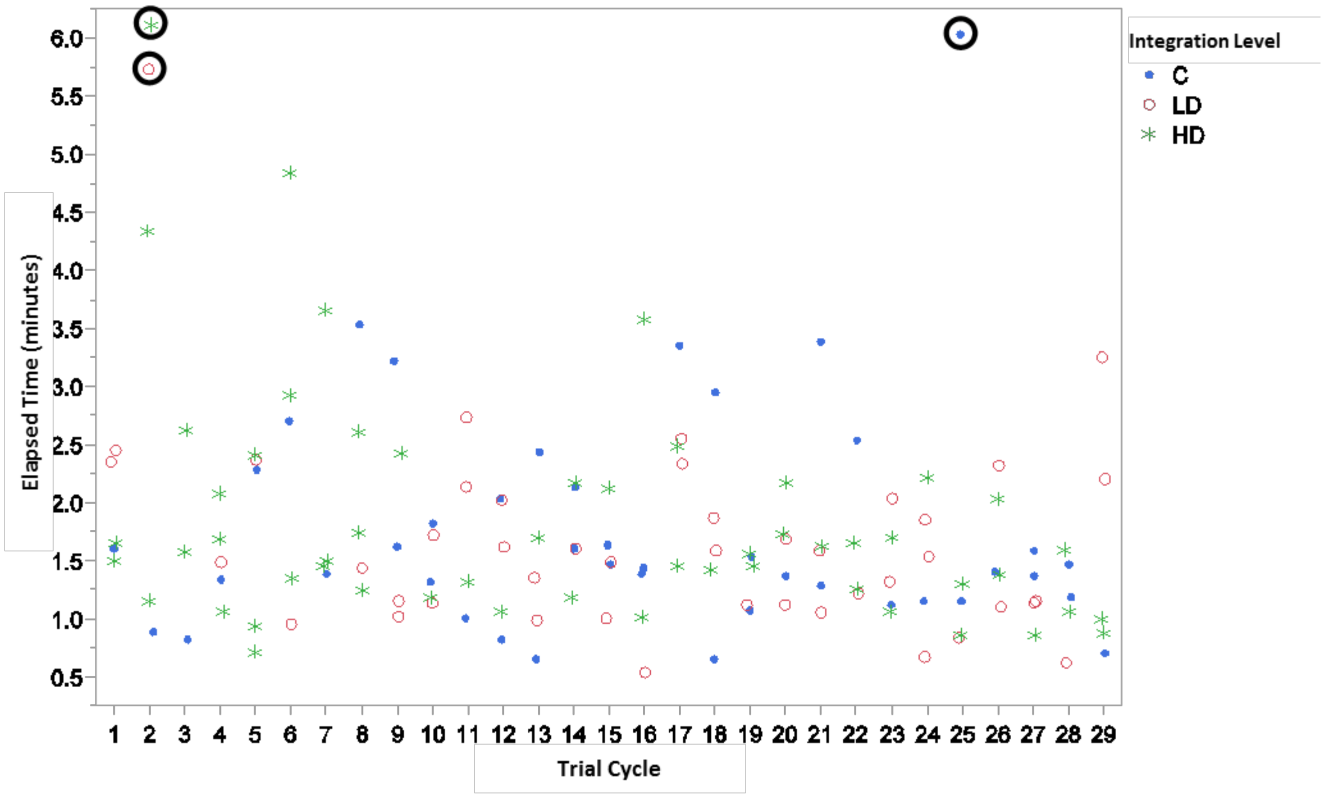
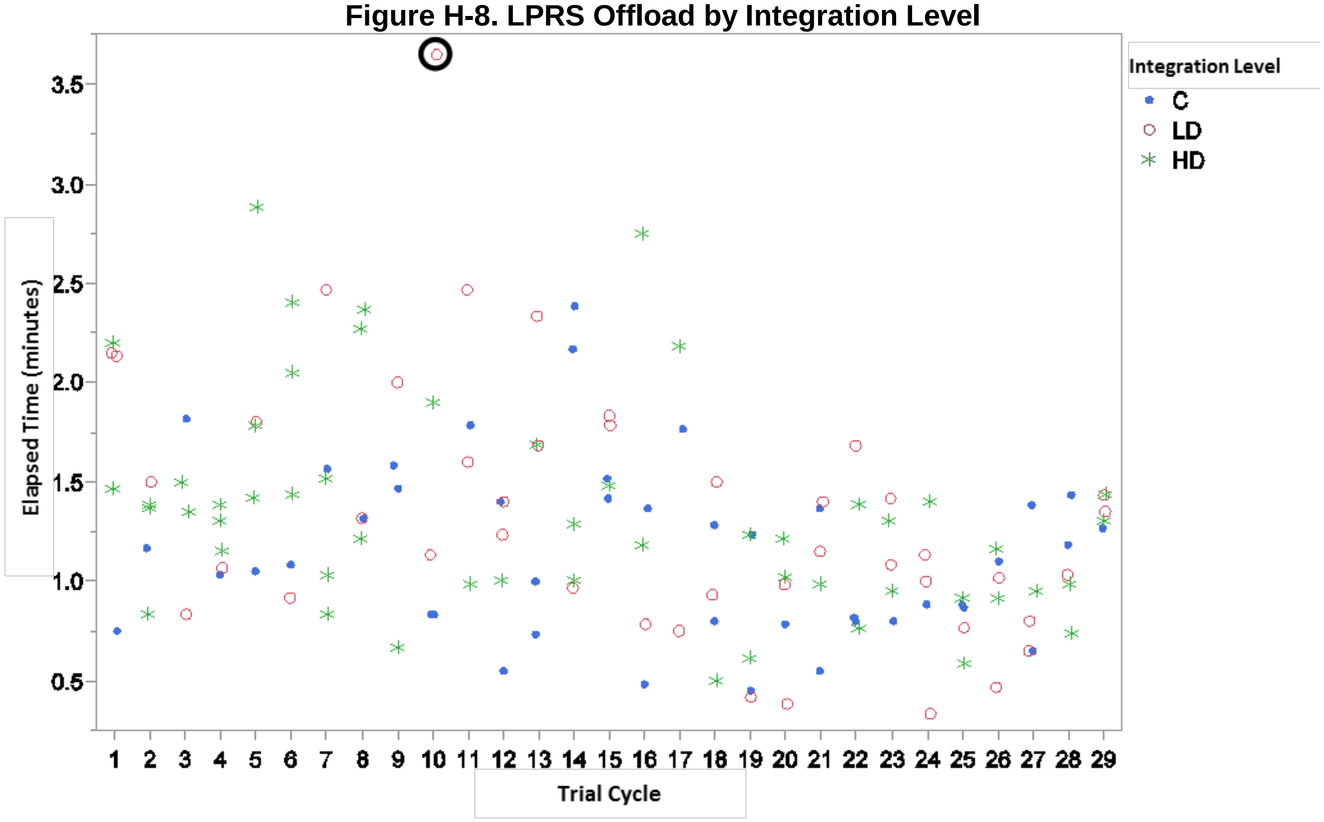


Figure H-8. LPRS Offload by Integration Level



H.5.2.7 LPRS Offload / Upload by Integration Level Data Table and Analysis

The tables below summarize the results of the subtasks, LPRS Offload/Upload. Table H-8 compares means across metrics and integration levels. Table H-9 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-8. LPRS Offload / Upload by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
First Emplacement LPRS Offload (minutes)	C	42	2.03	0.84
	LD	44	2.34	0.95
	HD	54	2.10	1.00
First Emplacement LPRS Offload [Excluding Influential Points] (minutes)	C	42	2.03	0.84
	LD	44	2.34	0.95
	HD	53	2.03	0.84
Second Emplacement LPRS Offload (minutes)	C	42	1.77	1.02
	LD	44	1.67	0.88
	HD	54	1.84	1.04
Second Emplacement LPRS Offload [Excluding Influential Points] (minutes)	C	41	1.67	0.78
	LD	43	1.57	0.62
	HD	53	1.76	0.86
LPRS Upload (minutes)	C	43	1.15	0.44
	LD	44	1.33	0.65
	HD	53	1.35	0.54
LPRS Upload [Excluding Influential Points] (minutes)	C	43	1.15	0.44
	LD	43	1.28	0.55
	HD	53	1.35	0.54

Table H-9. LPRS Offload / Upload by Integration Level ANOVA and Test Results

Metric	F	F Test	Comparison	Difference	%	P-	80 %	80%	90%	90%

	Statistic (df)	P-Value			Difference	Value	LCB	UCB	LCB	UCB
First Emplacement LPRS Offload (minutes)	1.28 (2, 137)	0.28	LD-C	0.31	15.10%	0.29	-0.04	0.66	-0.11	0.73
			HD-C	0.07	3.59%	0.92	-0.26	0.41	-0.33	0.47
			HD-LD	-0.23	-10.01%	0.44	-0.56	0.10	-0.63	0.16
First Emplacement LPRS Offload [Excluding Influential Points] (minutes)	1.86 (2, 136)	0.16	LD-C	0.31	15.10%	0.24	-0.02	0.63	-0.08	0.70
			HD-C	0.00	-0.15%	1.00	-0.32	0.31	-0.38	0.37
			HD-LD	-0.31	-13.26%	0.20	-0.62	0.00	-0.68	0.06
Second Emplacement LPRS Offload (minutes)	0.38 (2, 137)	0.69	LD-C	-0.10	-5.85%	0.88	-0.47	0.26	-0.54	0.34
			HD-C	0.07	3.97%	0.94	-0.28	0.42	-0.35	0.49
			HD-LD	0.17	10.43%	0.66	-0.17	0.52	-0.24	0.59
Second Emplacement LPRS Offload [Excluding Influential Points] (minutes)	0.71 (2, 134)	0.49	LD-C	-0.09	-5.65%	0.84	-0.38	0.19	-0.44	0.25
			HD-C	0.09	5.61%	0.83	-0.18	0.37	-0.24	0.42
			HD-LD	0.19	11.94%	0.46	-0.08	0.46	-0.14	0.51
LPRS Upload (minutes)	1.99 (2, 135)	0.14	LD-C	0.21	17.90%	0.20	0.00	0.41	-0.04	0.45
			HD-C	0.20	17.12%	0.19	0.00	0.39	-0.04	0.43
			HD-LD	-0.01	-0.66%	1.00	-	0.19	-	0.23

							0.20		0.24	
LPRS Upload [Excluding Influential Points] (minutes)	1.84 (2, 134)	0.16	LD-C	0.15	13.05%	0.38	-0.04	0.34	-0.08	0.38
			HD-C	0.20	17.12%	0.15	0.02	0.38	-0.02	0.42
			HD-LD	0.05	3.59%	0.90	-0.14	0.23	-0.17	0.27

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

H.5.2.7.1 First LPRS Offload by Integration Level

The C group data and LD group data are normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of 0.05 and 0.01, but the HD group data is not normally distributed with a p-value of < 0.01. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The tables above show that, on average, the C group had a mean time of 2.03 minutes; the LD group, 2.34 minutes; and the HD group, 2.10 minutes. The C group is faster but not statistically significant.

Excluding Influential Points shows the HD group to be as fast as the C group; however, it remains statistically insignificant.

H.5.2.7.2 Second LPRS Offload by Integration Level

No groups' data are normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The tables above show that, on average, the C group had a mean time of 1.77 minutes; the LD group, 1.67 minutes; and the HD group, 1.84 minutes. The LD group was the fastest group, but no comparison is statistically significant.

Excluding Influential Points shows the C group mean time dropping to 1.67 minutes, the LD group dropping to 1.57 minutes, and the HD group dropping to 1.76 minutes. Ultimately, the comparisons remain similar with no statistical significance in comparison.

H.5.2.7.2.1 LPRS Upload by Integration Level

The C group and LD group data are normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of 0.13 and 0.01, but the HD group data is not normally distributed with a p-value of < 0.01. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The tables above show that, on average, the C group had a mean time of 1.15 minutes; the LD group; 1.36 minutes; and the HD group, 1.35 minutes. The C group was faster than both the LD and HD groups, but this is not statistically significant.

Excluding Influential Points, the LD group drops in mean time to 1.30 minutes. The C group remains the fastest group, but this is not statistically significant.

H.5.2.7.3 Loose Projectile Restraint System (LPRS) Offload / Upload by Critical Billet

This second method of comparison for LPRS Offload / Upload looks at performance differences between gun crews grouped only by the gender of the Marines serving in two critical billets. In this analysis, the first critical billet is the Cannoneer #3, and it is denoted by either a (M) for male or a (F) for female. The second critical billet is the Cannoneer #4 and, again, is marked by a (M) or (F) to denote the gender of the Marine. In the tables below, the gender of the Cannoneer #3 and #4 billets are represented in two letters, with the first letter corresponding to the Cannoneer #3 billet and the second letter to the Cannoneer #4 billet.

H.5.2.7.4 LPRS Offload / Upload by Critical Billet Scatterplot

The scatterplots in figures H-9, H-10, and H-11 display the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling.

Figure H-9. First Emplacement LPRS Offload by Critical Billet

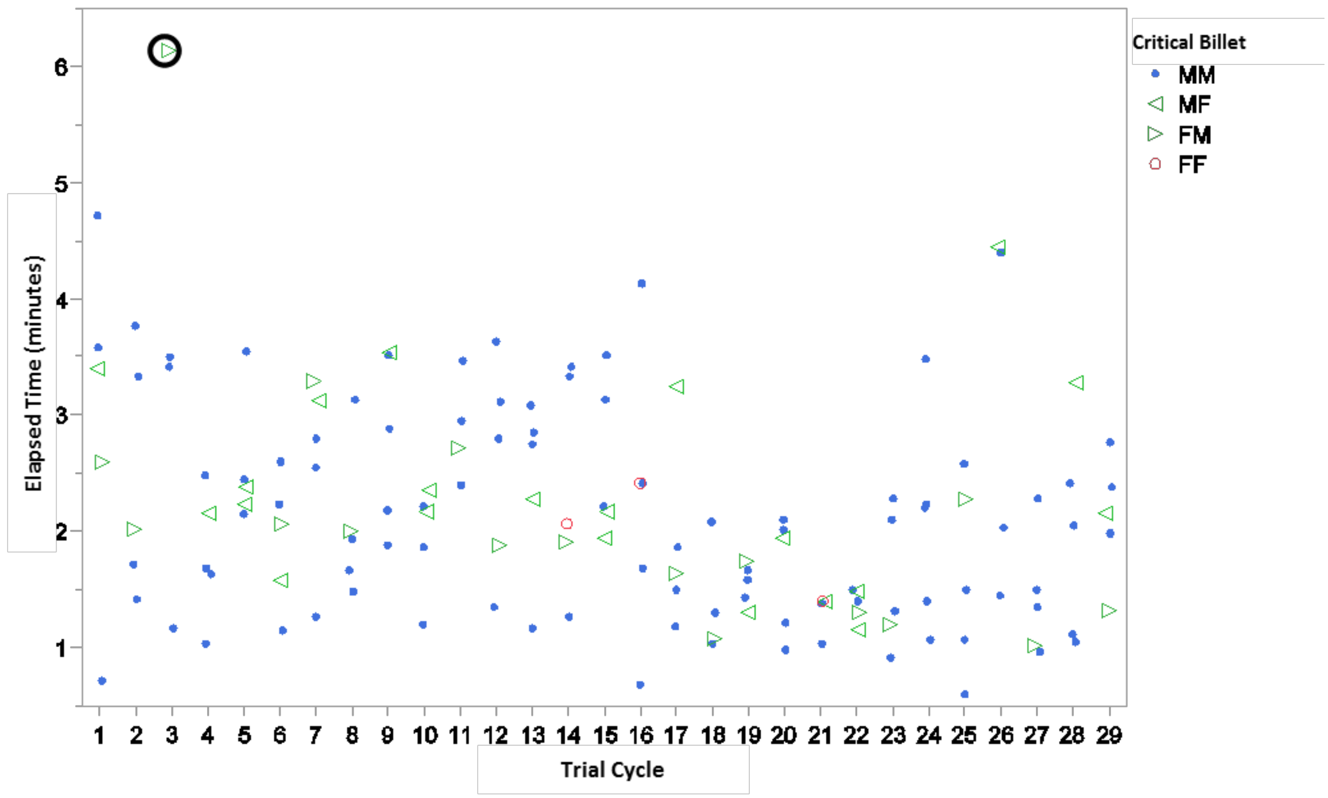


Figure H-10. Second Emplacement LPRS Offload by Critical Billet

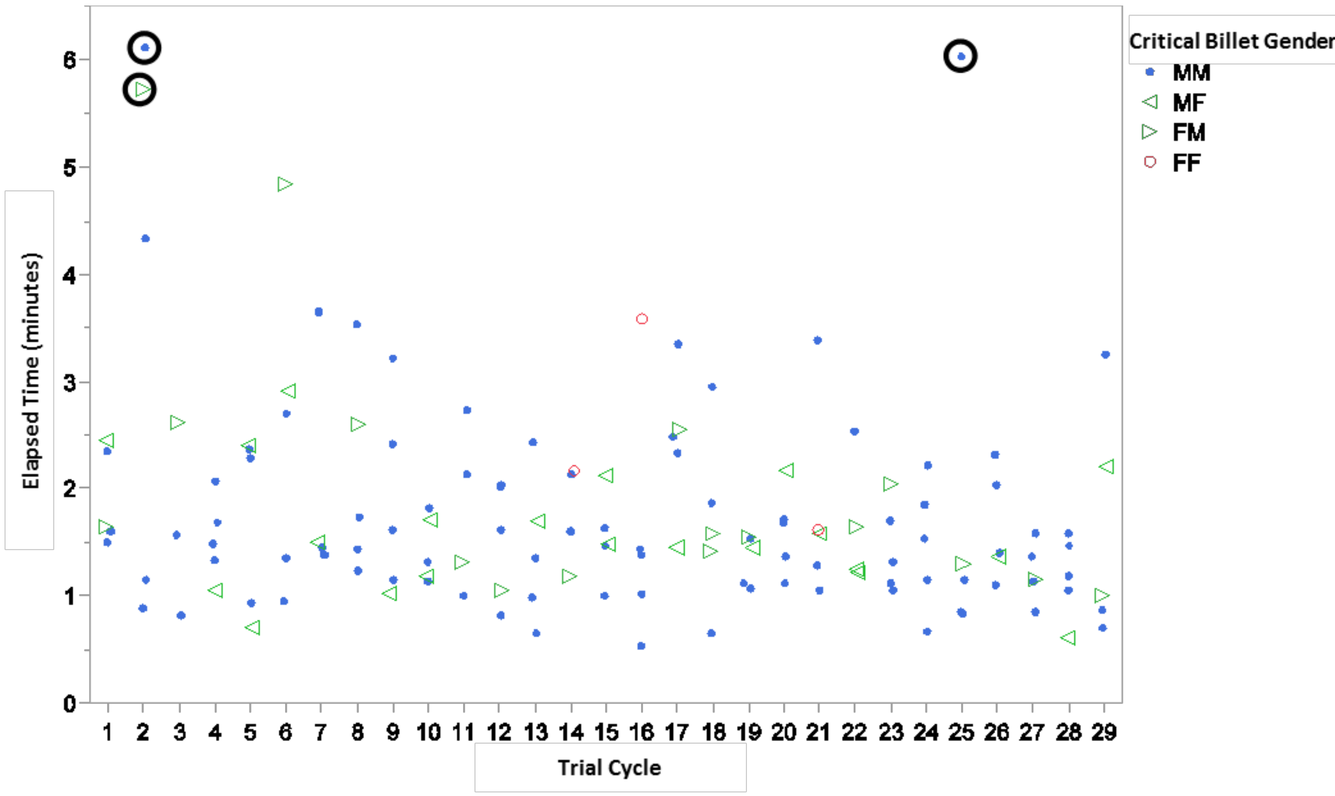
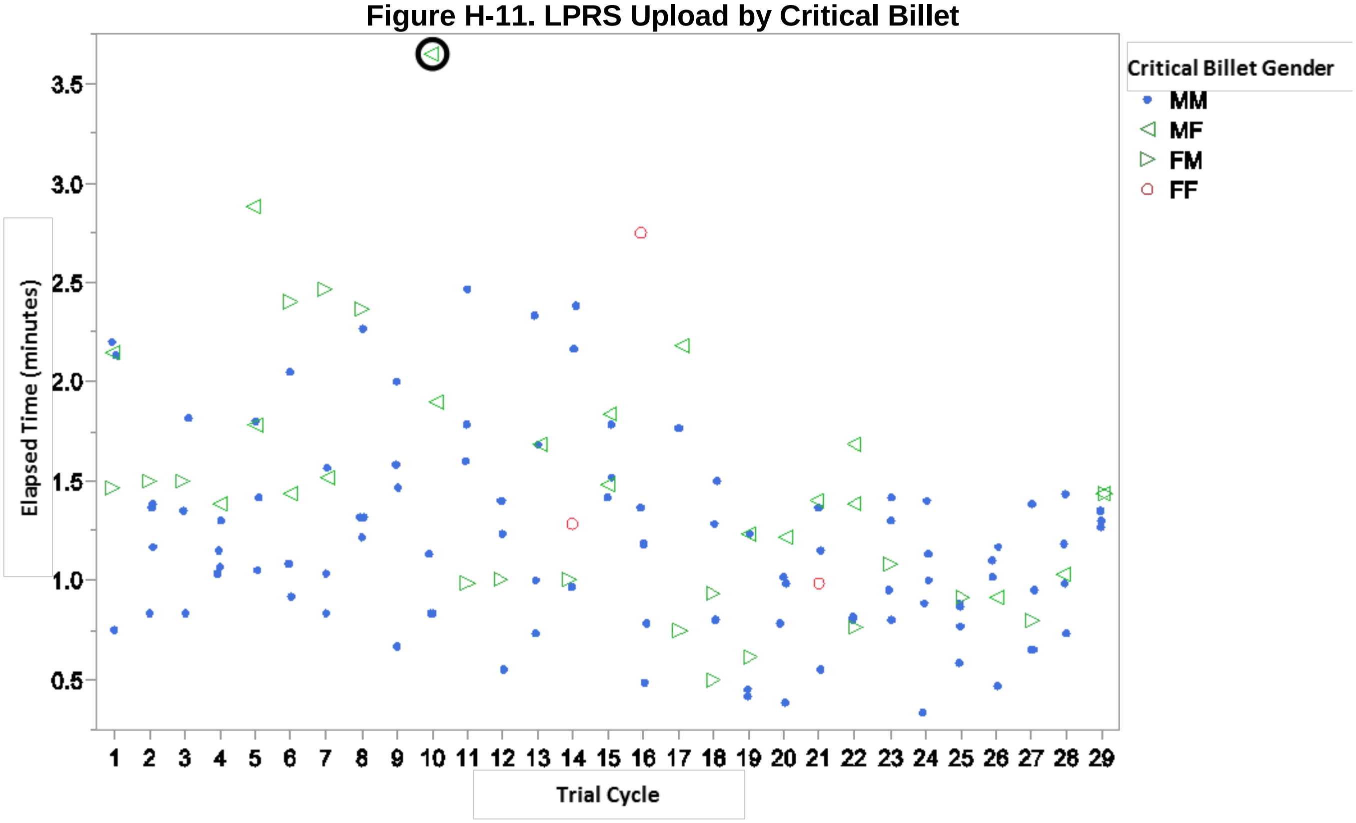


Figure H-11. LPRS Upload by Critical Billet



H.5.2.7.5 Loose Projectile Restraint System (LPRS) Offload / Upload by Critical Billet

The tables below summarize the results of the task LPRS Offload/Upload analysis by critical billet. Table H-10 compares means across metrics and integration levels. Table H-11 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-10. LPRS Results by critical billet

Metric	Integration Level	Sample Size	Mean	SD
First Emplacement LPRS Offload (minutes)	MM	99	2.12	0.93
	MF	21	2.37	0.85
	FM	17	2.13	1.20
	FF	3	1.96	0.52
First Emplacement LPRS Offload [Excluding Influential Points] (minutes)	MM	99	2.12	0.93
	MF	21	2.37	0.85
	FM	16	1.88	0.63
	FF	3	1.96	0.52
Second Emplacement LPRS Offload (minutes)	MM	99	1.73	0.97
	MF	21	1.60	0.59
	FM	17	2.07	1.33
	FF	3	2.46	1.01
Second Emplacement LPRS Offload [Excluding Influential Points] (minutes)	MM	97	1.64	0.75
	MF	21	1.60	0.59
	FM	16	1.84	0.96
	FF	3	2.46	1.01
LPRS Upload (minutes)*	MM	97	1.20	0.47
	MF	20	1.71	0.64
	FM	18	1.25	0.61

	FF	3	1.67	0.95
LPRS Upload [Excluding Influential Points] (minutes)*	MM	97	1.20	0.47
	MF	19	1.61	0.46
	FM	18	1.25	0.61
	FF	3	1.67	0.95

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table H-11. LPRS Results by Critical Billet ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
First Emplacement LPRS Offload (minutes)††	2.53	0.47	MF-MM	0.25	11.67%	0.24	-0.02	0.52	-0.10	0.60
			FM-MM	0.01	0.42%	0.98	-0.40	0.41	-0.52	0.54
			FF-MM	-0.16	-7.56%	0.65	-0.71	0.39	-0.98	0.65
			MF-FM	0.25	11.20%	0.50	-0.21	0.69	-0.35	0.83
			FF-MF	-0.41	-17.22%	0.31	-0.95	0.14	-1.17	0.35
			FF-FM	-0.17	-7.95%	0.70	-0.76	0.42	-0.96	0.62
First Emplacement LPRS Offload [Excluding Influential Points] (minutes)	0.98 (3, 135)	0.40	MF-MM	0.25	11.67%	0.65	-0.17	0.67	-0.24	0.74
			FM-MM	-0.24	-11.38%	0.74	-0.71	0.23	-0.79	0.31
			FF-MM	-0.16	-7.56%	0.99	-1.19	0.86	-1.36	1.04
			MF-FM	0.49	26.00%	0.34	-0.09	1.07	-0.19	1.17
			FF-MF	-0.41	-17.22%	0.88	-1.49	0.67	-1.67	0.85
			FF-FM	0.08	4.30%	1.00	-1.02	1.18	-1.20	1.37
Second Emplacement LPRS Offload (minutes)††	3.88	0.28	MF-MM	-0.13	-7.53%	0.43	-0.34	0.08	-0.40	0.14
			FM-MM	0.34	19.97%	0.32	-0.10	0.79	-0.24	0.93
			FF-MM	0.73	42.21%	0.34	-0.37	1.82	-0.94	2.40
			MF-FM	-0.47	-22.92%	0.19	-0.93	-0.02	-1.07	0.12

			FF-MF	0.86	53.78%	0.28	-0.23	1.95	-0.79	2.51
			FF-FM	0.38	18.54%	0.60	-0.68	1.45	-1.12	1.89
Second Emplacement LPRS Offload [Excluding Influential Points] (minutes)	1.46 (3, 133)	0.23	MF-MM	-0.04	-2.46%	1.00	-0.40	0.32	-0.46	0.38
			FM-MM	0.21	12.56%	0.75	-0.20	0.61	-0.27	0.68
			FF-MM	0.82	49.99%	0.26	-0.07	1.70	-0.22	1.85
			MF-FM	-0.25	-13.34%	0.77	-0.75	0.26	-0.83	0.34
			FF-MF	0.86	53.78%	0.27	-0.07	1.79	-0.23	1.95
			FF-FM	0.61	33.26%	0.58	-0.34	1.56	-0.50	1.72
LPRS Upload (minutes)††	13.08	< 0.01*	MF-MM	0.51	42.15%	< 0.01*	0.31	0.71	0.25	0.76
			FM-MM	0.05	3.88%	0.76	-0.15	0.25	-0.21	0.31
			FF-MM	0.47	39.07%	0.48	-0.56	1.50	-1.11	2.05
			MF-FM	0.46	36.83%	0.03	0.02	0.72	0.12	0.80
			FF-MF	-0.04	-2.16%	0.95	-1.05	0.97	-1.55	1.48
			FF-FM	0.42	33.88%	0.52	-0.58	1.43	-1.09	1.94
LPRS Upload [Excluding Influential Points] (minutes)††	13.69	< 0.01*	MF-MM	0.40	33.65%	< 0.01*	0.25	0.65	0.21	0.60
			FM-MM	0.05	3.88%	0.76	-0.15	0.25	-0.21	0.31
			FF-MM	0.47	39.07%	0.48	-0.56	1.50	-1.11	2.05
			MF-FM	0.36	28.66%	0.05	0.12	0.59	0.06	0.66
			FF-MF	0.07	4.06%	0.92	-0.95	1.08	-1.48	1.61
			FF-FM	0.42	33.88%	0.52	-0.58	1.43	-1.09	1.94

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
††Indicates results presented are from Robust ANOVA and Welch's t-tests with p-values compared to 0.0167 for Bonferroni adjustment. F-statistics is a Chi-square statistic from Robust ANOVA, and the F-test Pp-value is from a Robust ANOVA. The P-values are two-sided p-values from Welch's t-tests and the CIs are from Welch's t-tests.

H.5.2.7.5.1 First and Second LPRS Offload by Critical Billet

The MM and FM groups' data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01, but the FF and MF groups' data are normally distributed with p-values of 0.66 and 0.13. In addition, the FM group standard deviation is more than twice that of the FF group standard deviation. As such, we recommend using the robust ANOVA results shown above. We proceed with presenting robust ANOVA results because they were confirmed by a Kruskal-Wallis Test with a p-value of 0.55.

The tables above show that, on average, the mean times for the MM, MF, FM, and FF groups showed no statistical significance. The results are similar when excluding Influential Points.

H.5.2.7.5.2 Second LPRS Offload by Critical Billet

The MM and FM groups' data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01, but the FF and MF groups' data are normally distributed with p-values of 0.52 and 0.61. In addition, the FM group standard deviation is more than twice that of the MF group standard deviation. As such, we recommend using the robust ANOVA results shown above. We proceed with presenting robust ANOVA results because they were confirmed by a Kruskal-Wallis Test with a p-value of 0.28.

The tables above show that, on average, the mean times for the MM, MF, FM, and FF groups showed no statistical significance. The results are similar when excluding Influential Points.

H.5.2.7.5.3 LPRS Upload by Critical Billet

The MM and FM groups' data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01, but the FF and MF groups' data are normally distributed with p-values of 0.52 and 0.61. In addition, the FF group standard deviation is more than twice that of the MM group standard deviation. As such, we recommend using the robust ANOVA results shown above. We proceed with presenting robust ANOVA results because they were confirmed by a Kruskal-Wallis Test with a p-value of < 0.01.

The tables above show that, on average, the MM group had a mean time of 1.20 minutes with a SD of 0.47 minutes. The MF group had a mean time of 1.71 minutes. The FM group had a mean time of 1.25 minutes. The FF group had a mean time of 1.67 minutes. The MF-MM and MF-FM group comparisons are statistically significant.

Excluding Influential Points results in a drop in the MF mean time to 1.61 minutes. The MF-FM group comparison is no longer statistically significant. The MF-MM group comparison remains statistically significant as MM are 33.65% faster.

H.5.2.7.5.4 LPRS Offload / Upload by Critical Billet Contextual Comments

In the case of weaker Marines, it was a constant effort to dissuade the practice of disassembling the LPRS in order to reduce the effort required to execute the task. Intuitively, this tendency makes logical sense, as it eliminates the requirement to lift the projectile straight out of the system. However, the disassembly of the system results in an unsafe condition, in that, during normal operations, there are likely more than three projectiles in the load. Furthermore, terrain can result in the (unsecured) projectiles falling. This can cause injuries and damage to projectiles and equipment. In the above data, the result of this practice masks the disparity in performance between stronger and weaker Marines, regardless of gender.

H.5.3 Indirect-Fire Missions Overview

Five fire missions were executed by a howitzer section during each 1-day cycle. The missions were: a 3-round low-angle HE, a 3-round high-angle white phosphorus (WP) smoke, a 9-round low-angle HE, a 3-round low-angle smoke out-of-traverse limits, and, finally, another 3-round low-angle HE from the supplemental position. The M795 (HE) and M825 (WP) are ballistically and dimensionally similar and offer the same experience in the handling of the projectile. There is no ostensible difference at the gun section level between a HE and WP mission.

Procedures for each fire mission were largely the same, regardless of shell/fuze combination or angle of fire. Each fire mission began with the Fire Direction Center (FDC) starting the event. As the gun was receiving the firing data from the FDC, they were simultaneously aiming the howitzer by traversing left or right and adjusting the quadrant up or down. The loader, or #4 Cannoneer, would pick up a projectile from the ammunition pit, hold it to allow the Section Chief to inspect the round, and then place the round onto the loading tray. Cannoneer #4 would then join the Driver and ram the round into the breech. The #2 Cannoneer would then load the powder increments into the breech powder chamber and close the breech. Data collectors recorded the time hacks for both the ram and the closing of the breech. Finally, the howitzer was fired by the #1 Cannoneer on the Section Chief's command. Every shot and misfire was captured and used to calculate shot rate (seconds per shot) and elapsed time.

In the case of a high-angle mission, the loading of the projectile would occur prior to fully elevating the howitzer, a minor departure in procedure from the low-angle missions. The tube would be depressed between shots in order to facilitate the loading of the howitzer.

For all five fire missions, the #4 Cannoneer was identified as a critical billet. This was due to the Marine's physically challenging position and place on the critical path of the fire mission. The #1 Cannoneer was identified as an additional critical billet in the case of the high-angle mission. This Marine's role in high-elevation missions is the elevating and depressing of the gun to facilitate loading. The manipulation of elevation is executed through the spinning of a hand wheel; each turn of the wheel results in a delta of approximately 10 mils of elevation. For the high-angle missions, projectiles were loaded at 800 mils, and the missions averaged approximately 1050 mils, requiring roughly 25 turns of the wheel per elevation or depression. This task does not require pure strength so much as short bursts of endurance.

The data collected from the indirect-fire missions were analyzed in two separate ways. The first method of comparison looked at performance differences between the control group (C), the low-density group (LD), and the high-density group (HD). The second method of comparison looked at performance differences between gun crews grouped only by the gender of the critical billets, regardless of the crew composition. The critical billets identified for the high-angle mission were the #1 and #4 Cannoneers. All other fire missions had #4 as a critical billet.

H.5.3.1 Indirect-Fire Missions by Integration Level

The scatterplots in figures F-12 through F-16 display the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling.

Figure H-12. 3-Round Low-Angle HE; Indiv. Subsequent Rounds by Integration Level

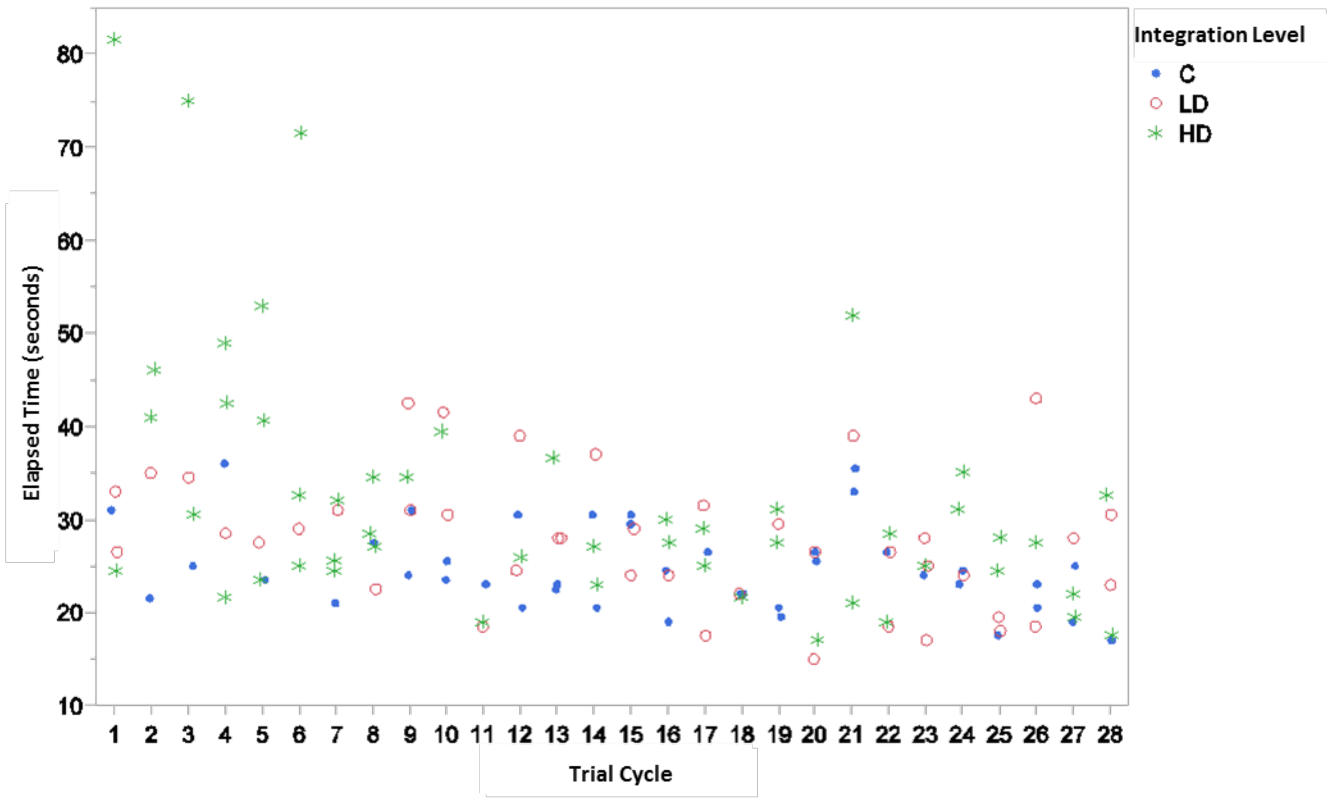


Figure H-13. 3-Round High-Angle WP; Indiv. Subsequent Rounds by Integration Level

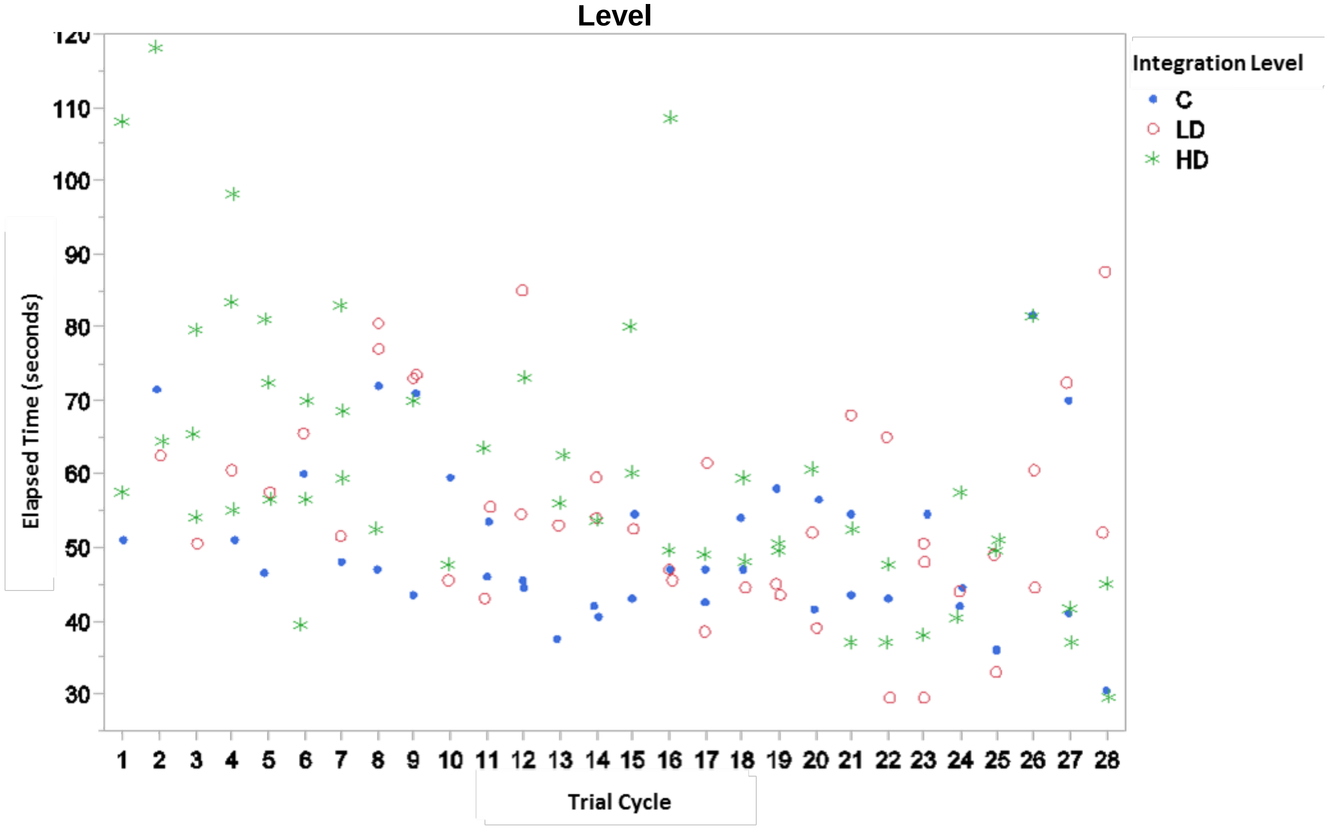


Figure H-14. 9-Round Low-Angle HE; Indiv. Subsequent Rounds by Integration Level

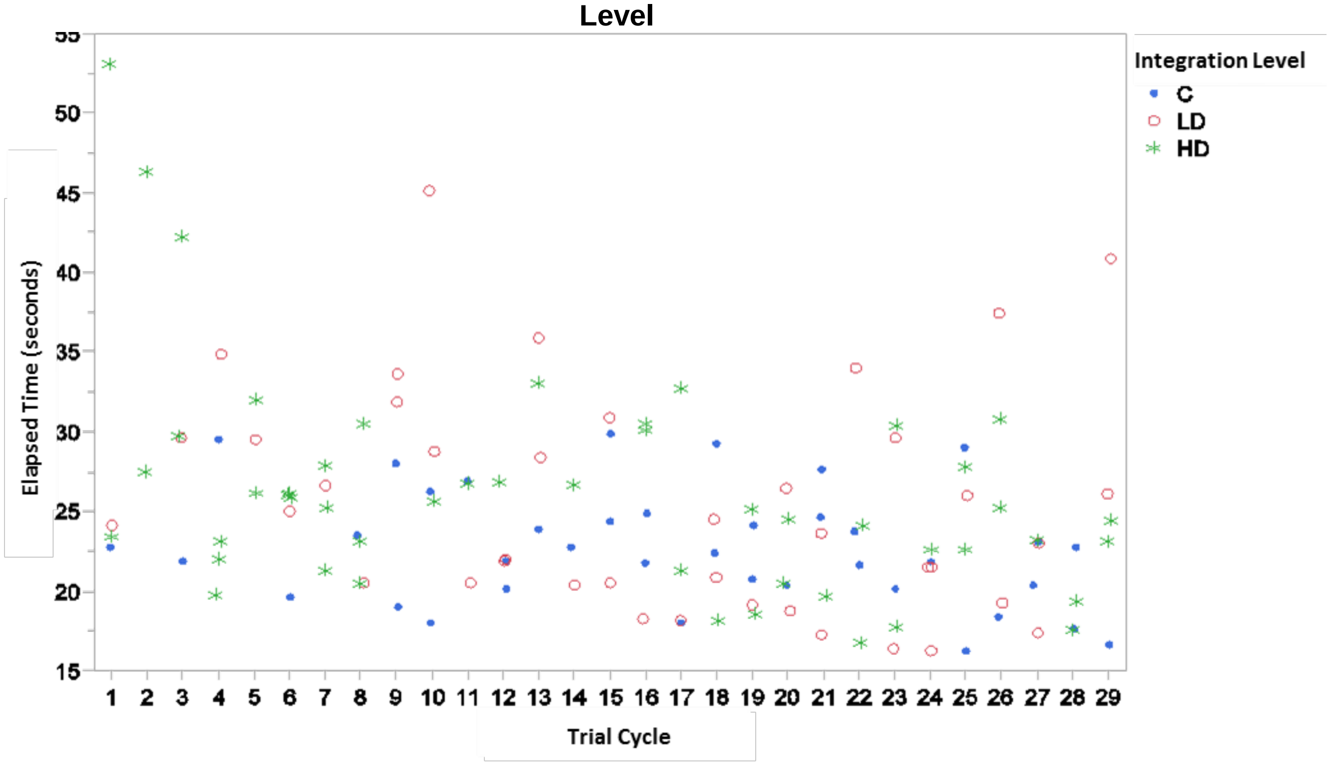


Figure H-15. 3-Round Low-Angle WP; Indiv. Subsequent Rounds by Integration Level

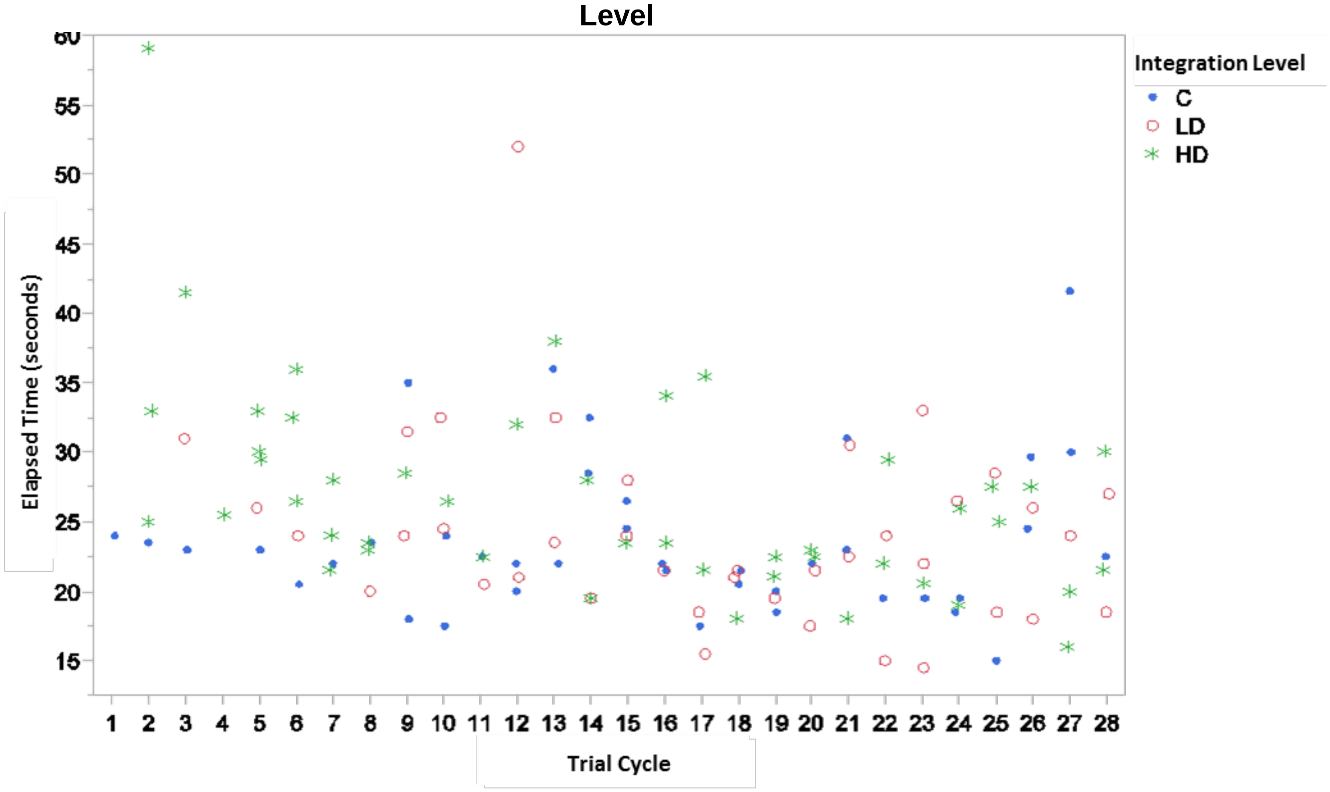
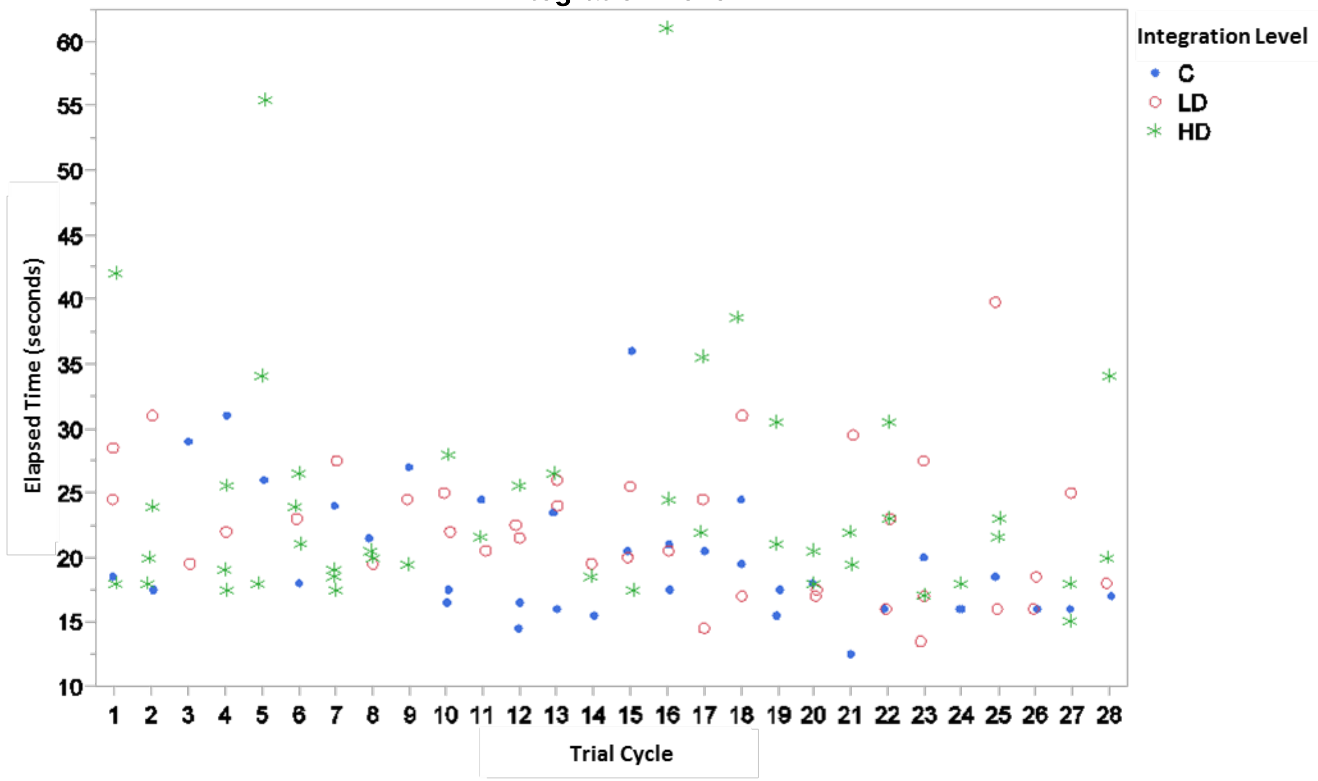


Figure H-16. 3-Round Low-Angle HE from FP2; Indiv. Subsequent Rounds by Integration Level



H.5.3.2 Indirect-Fire Missions by Integration Level Data Table and Analysis

The tables below summarize the results of the task Indirect-Fire Missions analysis by Integration Level. Table H-12 compares means across metrics and integration levels. Table H-13 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-12. Indirect-Fire Mission Results by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
3-Round Low-Angle HE; Indiv. Subsequent Rounds (average seconds per round)*	C	41	24.74	4.66
	LD	42	27.75	7.18
	HD	50	32.53	14.02
3-Round High-Angle WP; Indiv. Subsequent Rounds (average seconds per round)*	C	39	50.32	11.08
	LD	42	54.87	14.03
	HD	51	60.96	19.27

9-Round Low-Angle HE; Indiv. Subsequent Rounds (average seconds per round)*	C	38	22.72	3.68
	LD	41	25.52	7.01
	HD	49	26.06	6.96
3-Round Low-Angle WP; Indiv. Subsequent Rounds (average seconds per round)*	C	40	23.66	5.56
	LD	39	24.09	6.77
	HD	46	26.83	7.48
3-Round Low-Angle HE from FP2; Indiv. Subsequent Rounds (average seconds per round)*	C	36	19.88	5.10
	LD	38	22.32	5.43
	HD	47	24.44	9.41

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

Table H-13. Indirect-Fire Missions by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB**	80% UCB**	90% LCB**	90% UCB**
3-Round Low-Angle HE; Indiv. Subsequent Rounds (average second per round)††	12.70)	< 0.01*	LD-C	3.00	12.14%	0.03*	0.02	0.08	0.01	0.09
			HD-C	7.79	31.47%	< 0.01*	0.08	0.18	0.07	0.19
			HD-LD	4.78	17.24%	0.04	0.03	0.13	0.02	0.14
3-Round High-Angle WP; Indiv. Subsequent Rounds (average second per round)	5.28 (2, 129)	< 0.01*	LD-C	4.54	9.03%	0.39	-1.43	10.52	-2.63	11.71
			HD-C	10.64	21.14%	< 0.01*	4.92	16.35	3.78	17.50
			HD-LD	6.09	11.11%	0.15	0.50	11.69	-0.63	12.81
9-Round	3.42	0.04*	LD-C	2.80	12.31%	0.12	0.39	5.20	-0.09	5.68

Low-Angle HE; Indiv. Subsequent Rounds (average second per round)	(2, 125)		HD-C	3.34	14.72%	0.04*	1.03	5.65	0.57	6.12
			HD-LD	0.55	2.14%	0.91	-1.71	2.81	-2.17	3.26
3-Round Low-Angle WP; Indiv. Subsequent Rounds (average seconds per round)	2.88 (2, 122)	0.06*	LD-C	0.43	1.83%	0.96	-2.16	3.03	-2.69	3.55
			HD-C	3.17	13.40%	0.08*	0.67	5.66	0.17	6.17
			HD-LD	2.74	11.36%	0.15	0.22	5.25	-0.28	5.75
3-Round Low-Angle HE from FP2; Indiv. Subsequent Rounds (average second per round)	4.13 (2, 118)	0.02*	LD-C	2.45	12.32%	0.31	-0.43	5.33	-1.01	5.91
			HD-C	4.56	22.96%	0.01*	1.82	7.31	1.27	7.86
			HD-LD	2.11	9.47%	0.37	-0.52	4.82	-1.13	5.36

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.††Indicates results presented are from Robust ANOVA and Welch's t-tests with p-values compared to 0.033 for Bonferroni adjustment. F-statistics is a Chi-square statistic from Robust ANOVA, and the F-test Pp-value is from a Robust ANOVA. The P-values are two-sided p-values from Welch's t-tests and the CIs are from Welch's t-tests.

H.5.3.2.1.1 3-Round Low-Angle (HE) Results by Integration Level

The C and LD groups' data are normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of 0.08 and 0.25, but the HD group's data are not normally distributed with a p-value of < 0.01. In addition, the HD group standard deviation is more than twice that of the C group standard deviation. As such, we recommend using the robust ANOVA results shown above. We proceed with presenting robust ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

The first Indirect-Fire Mission is the 3-round low-angle HE Indiv. Subsequent Rounds fire mission. The tables above show that, on average, the C group had a mean time of 24.74 seconds per round; the LD group, 27.75 seconds per round; and the HD group, 32.53 seconds per round. On average, the C group was faster than the LD and HD groups, and the differences between the HD and C groups, and the HD and LD groups are statistically significant.

H.5.3.2.1.2 3-Round High-Angle (WP) Results by Integration Level

The C and HD groups' data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01, but the LD group's data is normally distributed with a p-value of 0.28. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The second Indirect-Fire Mission is the 3-round high-angle WP Indiv. Subsequent Rounds fire mission. The tables above show, on average, the C group had a mean time of 50.32 seconds per round; the LD group, 54.87 seconds per round; and the HD group, 60.96 seconds per round. On average, the C group was faster than the LD and HD groups, and the difference between the HD and C groups is statistically significant.

H.5.3.2.1.3 9-Round Low-Angle (HE) Results by Integration Level

The C and LD groups' data are normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of 0.30 and 0.02, but the HD group's data is not normally distributed with a p-value of < 0.01. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA

The third Indirect-Fire Mission is the 9-round low-angle HE Indiv. Subsequent Rounds fire mission. The tables above show that, on average, the C group had a mean time of 22.72 seconds per round, while the LD group recorded a mean time of 25.52 seconds per round and the HD group noted a mean time of 26.06 seconds per round. On average, the C group was faster than the LD and HD groups. The difference between the HD and C groups is statistically significant.

H.5.3.2.1.4 3-Round Low-Angle (WP) Results by Integration Level

The C, LD, and HD groups' data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The fourth Indirect-Fire Mission is the 3-round low-angle WP Indiv. Subsequent Rounds task. The tables above show that, on average, the C group had a mean time of 23.66

seconds per round. The LD and HD groups had mean times of 24.09 and 26.83 seconds per round, respectively. On average, both LD and HD groups were slower than the C group. The difference between the HD and C groups is statistically significant.

H.5.3.2.1.5 3-Round Low-Angle (HE) from FP2 Results (Analysis by Integration Level)

The C and HD groups' data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01, but the LD group's data is normally distributed with a p-value of 0.10. We proceed with presenting ANOVA results because sample sizes are sufficiently large ($n > 30$) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The fifth Indirect-Fire Mission is the 3-round low-angle HE from FP2 Indiv. Subsequent Rounds fire mission. The tables above show that, on average, the C group had a mean time of 19.88 seconds per round; the LD group, 22.32 seconds per round; and the HD group, 24.44 seconds per round. On average, the C group was faster than the LD and HD groups. The difference between the HD and C groups is statistically significant.

H.5.3.2.2 Indirect-Fire Missions by Critical Billet

This second method of comparison for Indirect-Fire Missions looks at performance differences between gun crews grouped only by the gender of the Marines serving in critical billets. In this analysis, the critical billet for all Low-Angle Fire Missions is the Cannoneer #4, and it is denoted by either an (M) for male or an (F) for female.

For the High-Angle Missions, there are two critical billets—the #1 Cannoneer and #4 Cannoneer. The tables below display the Cannoneer #1 and #4 billets as a two-letter representation, where the first letter represents the gender of the Marine occupying the Cannoneer #1 billet and the second letter represents the gender of the Marine occupying the Cannoneer #4 billet.

H.5.3.2.2.1 Low-Angle Fire Mission by Critical Billet—Cannoneer #4

Four out of the five fire missions are low-angle fire missions and are analyzed below by the critical billet—Cannoneer #4. The four fire missions are: a 3-round low-angle High Explosive (HE), a 9-round low-angle HE, a 3-round low-angle smoke out-of-traverse limits, and, finally, another 3-round low-angle HE from the supplemental position.

H.5.3.2.2.2 Low-Angle Fire Mission by Critical Billet—Cannoneer #4 Scatterplots

The scatterplots in figures H-17 through H-20 display the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling.

Figure H-17. 3-Round Low-Angle HE; Indiv. Subsequent Rounds by Critical Billet

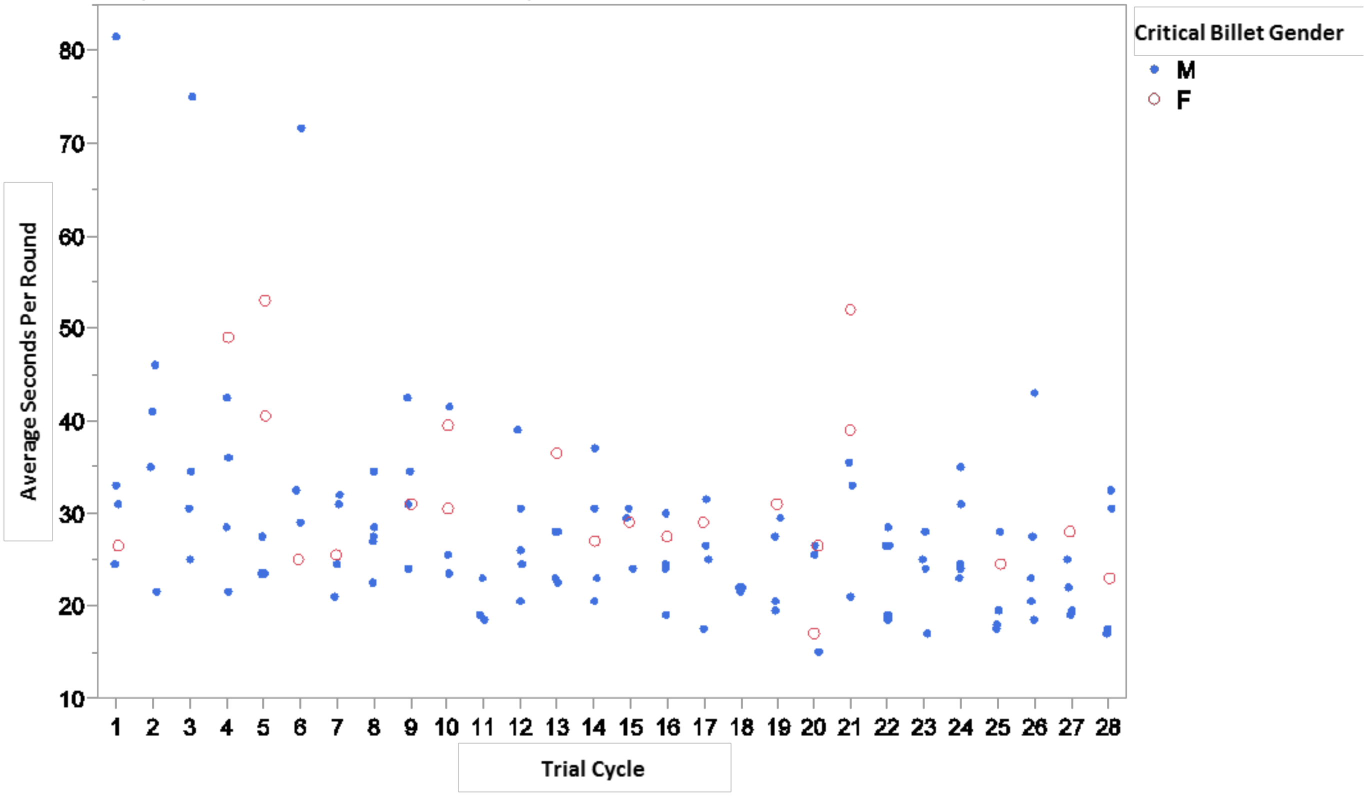


Figure H-18. 9-Round Low-Angle HE; Indiv. Subsequent Rounds by Critical Billet

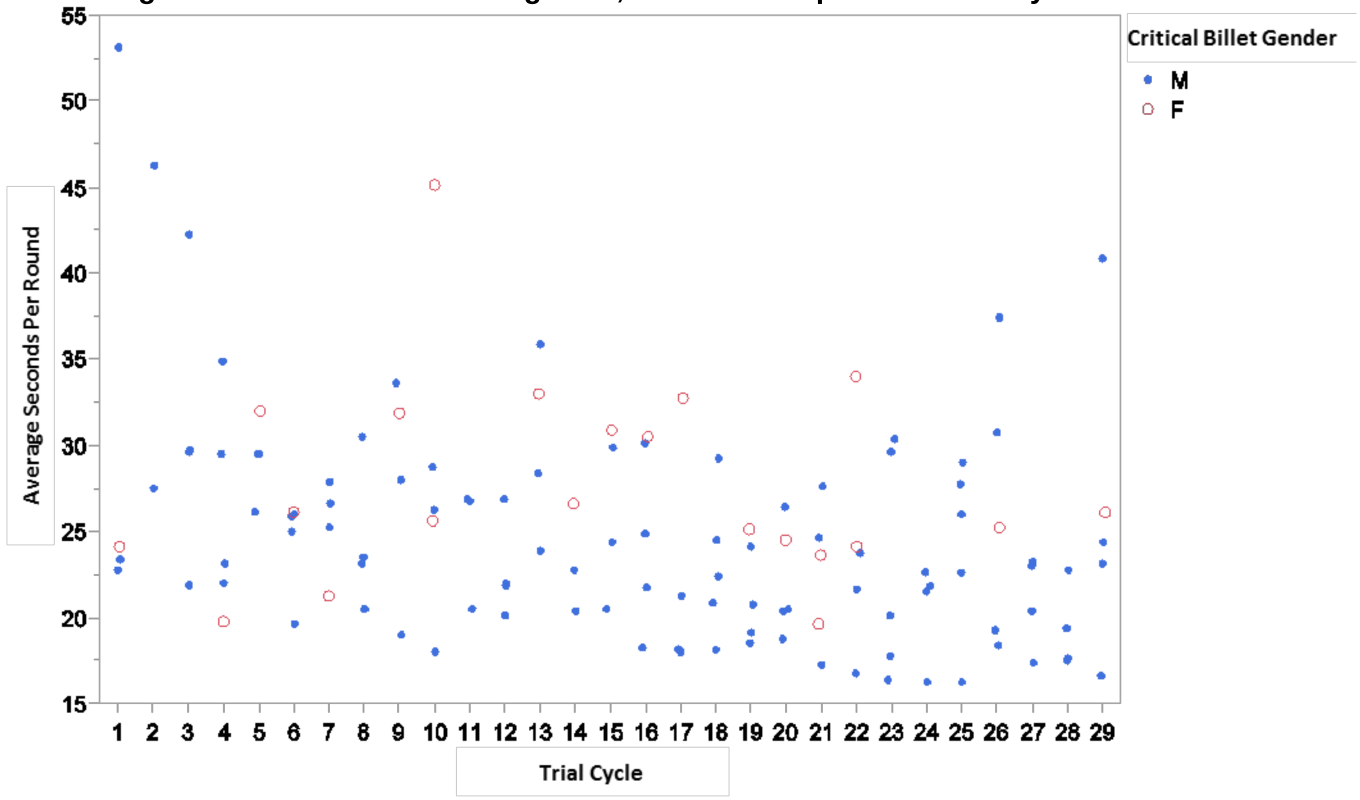


Figure H-19. 3-Round Low-Angle WP; Indiv. Subsequent Rounds by Critical Billet

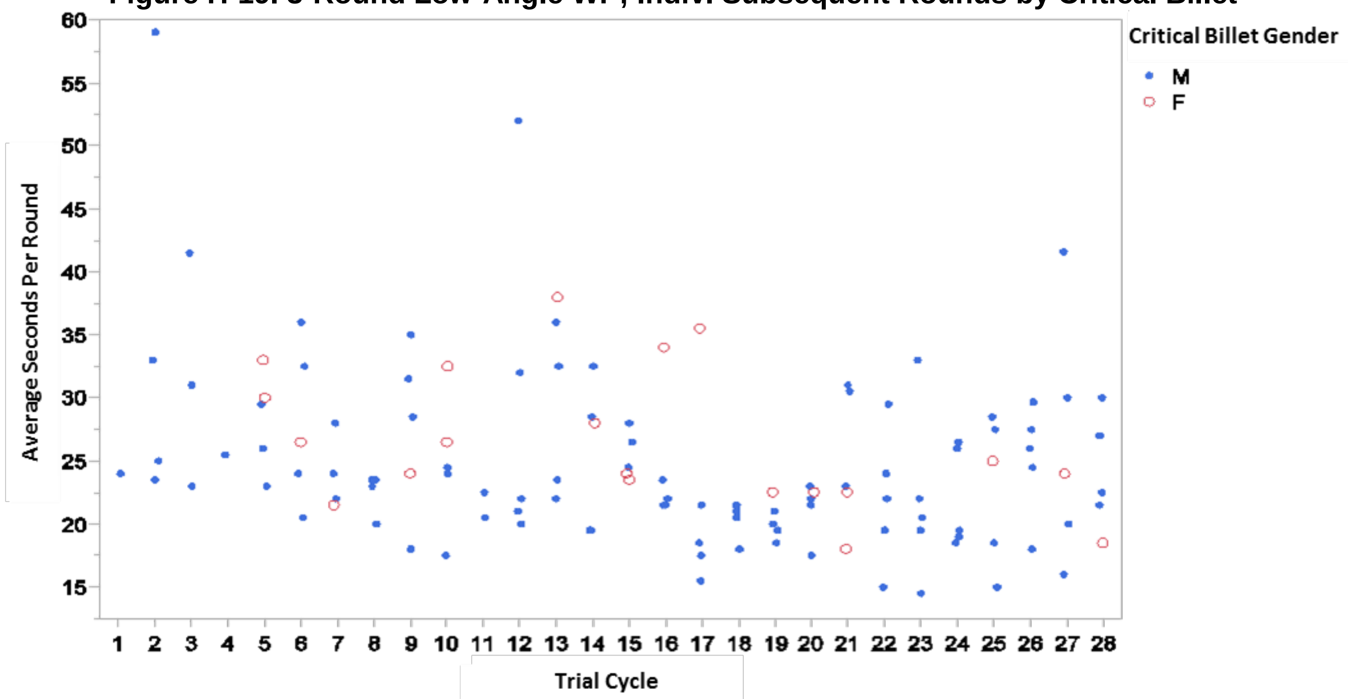
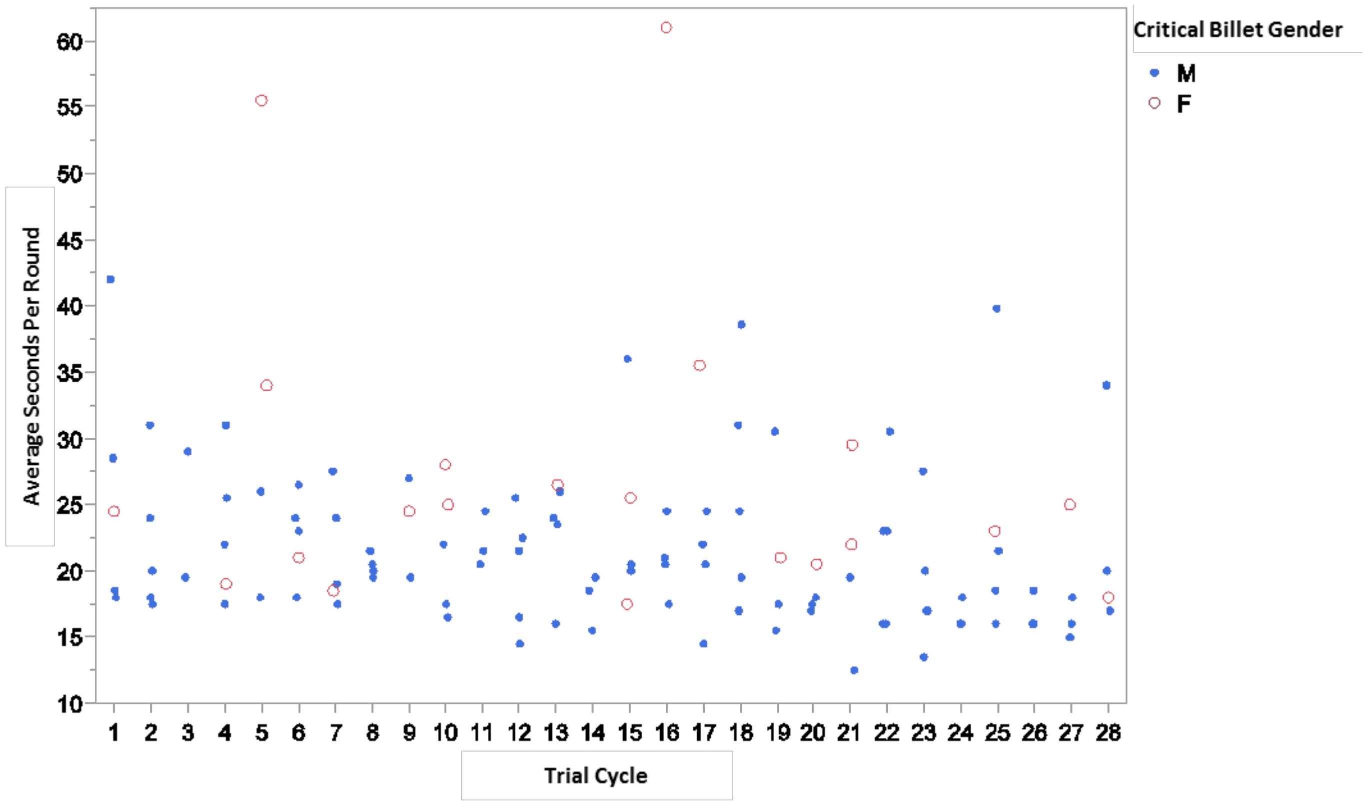


Figure H-20. 3-Round Low-Angle HE from FP2; Indiv. Subsequent Rounds by Critical Billet



H.5.3.2.2.3 Low-Angle Fire Mission by Critical Billet—Cannoneer #4 Data Table and Analysis

The tables below summarize the results of the task Indirect-Fire Missions analysis by Critical Billet (Low Angle Missions). Table H-14 compares means across metrics and integration levels. Table H-15 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences

Table H-14. Indirect-Fire Mission Results (Descriptive Statistics by critical billet)

Metric	Integration Level	Sample Size	Mean	SD
3-Round Low-Angle HE; Indiv. Subsequent Rounds (average seconds per round)*	M	111	27.89	10.33
	F	22	32.39	9.57
9-Round Low-Angle HE; Indiv. Subsequent Rounds (average seconds per round)*	M	107	24.34	6.27
	F	21	27.71	5.90

3-Round Low-Angle WP; Indiv. Subsequent Rounds (average seconds per round)†	M	105	24.66	6.98
	F	20	26.50	5.63
3-Round Low-Angle HE from FP 2; Indiv. Subsequent Rounds (average seconds per round)*	M	100	21.37	5.77
	F	21	27.38	11.36

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Indicates contradicting statistical significance results between ANOVA and a non-parametric equivalent test.

Table H-15. Indirect-Fire Missions ANOVA and T-test Results

Metric	F statistic (df)	F Test P-value	Comparison	Difference	% Difference	2-sided P-Value	1-sided P-Value	80% LCB
3-Round Low-Angle HE; Indiv. Subsequent Rounds (average seconds per round)	3.41 (1, 131)	0.07*	F-M	4.50	15.78%	0.06*	0.03* (0.06*)	1.44
9-Round Low-Angle HE; Indiv. Subsequent Rounds (average seconds per round)	5.16 (1, 126)	0.02*	F-M	3.37	13.83%	0.02*	0.01* (0.02*)	1.50
3-Round Low-Angle WP; Indiv. Subsequent Rounds (average seconds per round)†	1.23 (1, 123)	0.27	F-M	1.84	7.44%	0.07†	0.03†	0.01†
3-Round Low-Angle HE from FP 2; Indiv. Subsequent Rounds (average seconds per round)	12.68 (1, 119)	< 0.01*	F-M	6.01	28.11%	0.03*	0.01* (0.03*)	2.65

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
†Results presented are from a Mann-Whitney non-parametric test due to non-normality.

H.5.3.2.2.3.1 3-Round Low-Angle (HE) by Critical Billet—Cannoneer #4

The M group data is not normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of < 0.01, but the F group data is normally distributed with a p-value of 0.02. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a p-value of 0.01. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA

The tables above show that, on average, the M group had a mean time of 27.89 seconds per round, while the F group had a mean time of 32.39 seconds per round. On average, the F group was 15.78% slower than the M group, and the difference is statistically significant in one-sided and two-sided hypothesis tests

H.5.3.2.2.3.2 9-Round Low-Angle (HE) by Critical Billet—Cannoneer #4

The M group data is not normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of < 0.01, but the F group data is normally distributed with a p-value of 0.03. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a p-value of < 0.01. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA

The tables above show that, on average, the M group had a mean time of 24.34 seconds per round. The F group had a mean time of 27.71 seconds per round. On average, the F group was 13.83% slower than the M group, and the difference is statistically significant in one-sided and two-sided hypothesis tests.

H.5.3.2.2.3.3 3-Round Low-Angle (WP) by Critical Billet—Cannoneer #4

The M group data is not normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of < 0.01, but the F group data is normally distributed with a p-value of 0.23. Because of a lack of normality, we recommend using the Mann-Whitney test results because they disagree with ANOVA test results.

The tables above show that, on average, the M group had a mean time of 24.66 seconds per round. The F group had a mean time of 26.50 seconds per round. On average, the F group was 7.44% slower than the M group, and the difference is not statistically significant.

H.5.3.2.2.3.4 3-Round Low-Angle (HE) from FP2 by Critical Billet—Cannoneer #4

The M group and F group data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01. We proceed with presenting ANOVA results because they were confirmed by a Mann-Whitney Test with a p-value of < 0.01.

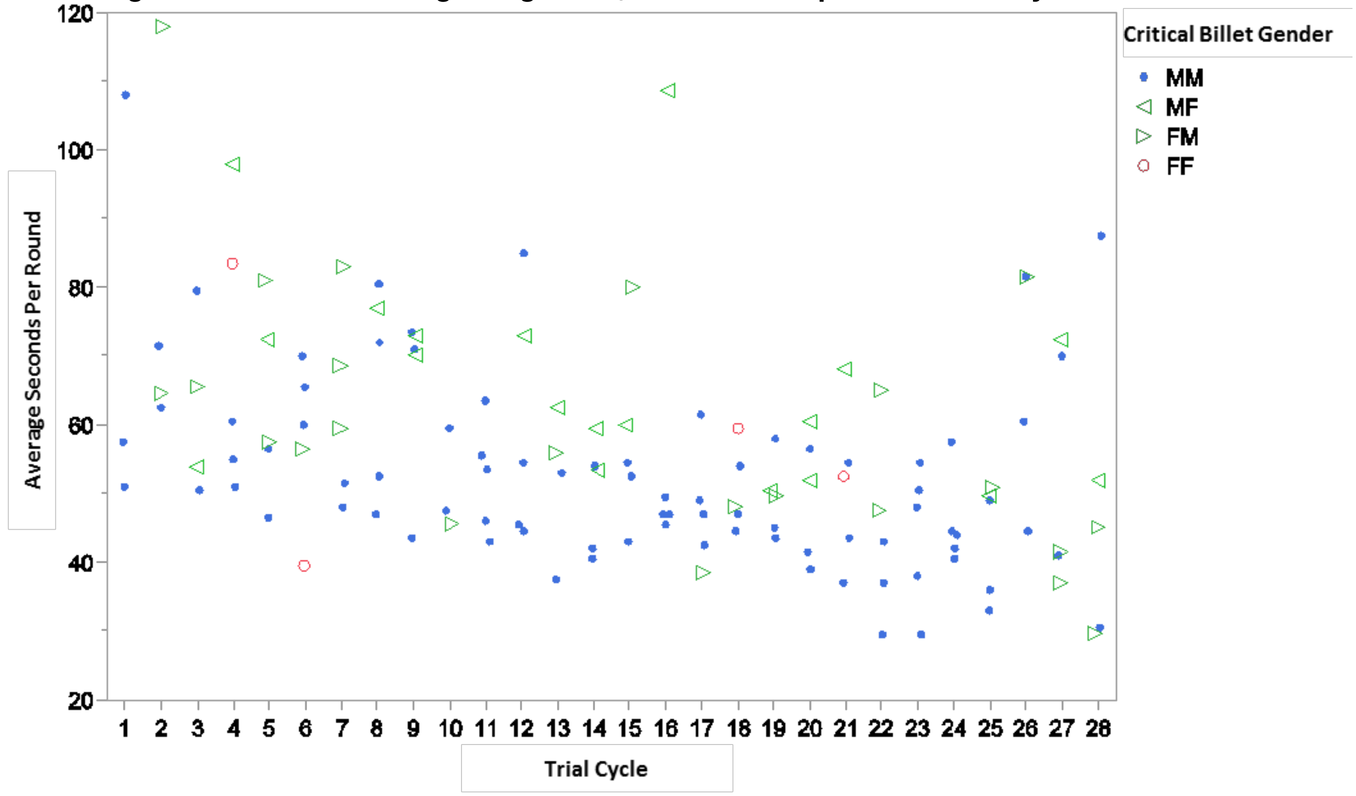
The tables above show that, on average, the M group had a mean time of 21.37 seconds per round. The F group had a mean time of 27.38 seconds per round. On average, the F group was 28.11% slower than the M group, and the difference is statistically significant in one-sided and two-sided hypothesis tests.

H.5.3.2.2.4 High-Angle Fire Mission by Critical Billet—Cannoneers #1 and #4

One out of the five fire missions is a High-Angle fire mission. High-Angle fire missions are analyzed below by two critical billets—Cannoneer #1 and Cannoneer #4. The only High-Angle fire mission is the 3-round high-angle (WP) fire mission.

H.5.3.2.2.5 High-Angle Fire Mission by Critical Billet—Cannoneer #1 and #4 Scatterplot
The scatterplot in Figure H-21 displays the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling.

Figure H-21. 3-Round High-Angle WP; Indiv. Subsequent Rounds by Critical Billet



H.5.3.2.2.6 High-Angle Fire Mission by Critical Billet—Cannoneer #1 and #4 Data Table and Analysis
The tables below summarize the results of the task Indirect-Fire Missions analysis by Critical Billet (High-Angle Mission). Table H-16 compares means across metrics and integration levels. Table H-17 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-16. High-Angle Fire Mission Results by critical billet

Metric	Integration Level	Sample Size	Mean	SD
3-Round High-Angle WP; Indiv. Subsequent Rounds (average seconds per round)	MM	86	52.39	13.81
	MF	19	66.65	15.76
	FM	23	59.54	19.73

	FF	4	58.75	18.46

Table H-17. Indirect-Fire Missions by Critical Billet ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
3-Round High-Angle WP; Indiv. Subsequent Rounds (average second per round)	5.07 (3, 128)	< 0.01*	MF-MM	14.27	27.24%	< 0.01*	6.65	21.99	5.24	23.29
			FM-MM	7.16	13.66%	0.20	0.00	14.31	-1.20	15.51
			FF-MM	6.36	12.15%	0.85	-9.23	21.96	-11.84	24.57
			MF-FM	7.11	11.94%	0.45	-2.34	16.56	-3.93	18.14
			FF-MF	-7.90	-11.86%	0.79	-24.67	8.87	-27.48	11.68
			FF-FM	-0.79	-1.33%	1.00	-17.31	15.72	-20.08	18.49

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

H.5.3.2.2.6.1 3-Round High-Angle (WP) by Critical Billet—Cannoneer #1 and Cannoneer #4

The MM group data is not normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of < 0.01, but the FF, FM, and MF groups' data are normally distributed with p-values of 0.81, 0.08, and 0.01. We proceed with presenting ANOVA results because they were confirmed by a Kruskal-Wallis Test with a p-value of < 0.01. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA

For the High-Angle critical billet analysis, the #1 Cannoneer and #4 Cannoneer were used to calculate the following average times. For the 3-Round High-Angle WP Indiv. Subsequent Rounds, the MM group had a mean time of 52.39 seconds per round; the MF group, 66.65 seconds per round; the FM group, 59.54 seconds per round; and the FF group, 58.75 seconds per round. On average, the MF group was 27.24% slower than the MM group, and the difference between the MF and MM groups is statistically significant. On average, the FM was 13.66% slower than the MM group, and the difference between the FM and MM groups is not statistically significant. On average, the FF group was 12.15% slower than the MM group, and the difference between the FF and MM groups is not statistically significant.

H.5.3.3 Indirect-Fire Missions Contextual Comments

The Fire Mission is the crux of Marine Artillery. The timely delivery of fires in support of maneuver units is top priority for every Artillery Unit, and all efforts inevitably lead to the pursuit of increased responsiveness and accuracy of fire support. Fire missions come in many varieties, from planned fires to immediate suppression. In kinetic operations, presuming the safe operation of the howitzer, speed is paramount, delays unacceptable, and missed timelines abject failures. The specific disparities between the times listed in the above tables are important, but less so than the disparities in the context of aggregate effect. One would expect the difference in responsiveness to become more pronounced as time and fatigue increase, particularly over the course of a deployment. Further, given the often reactionary nature of fires, any degradation of fire responsiveness is a detriment to the supported Marines.

H.5.4 Out-of-Traverse Limits Overview

When providing fire support in a sprawling battle space, it is not a guarantee that the enemy target will always be in line with the azimuth of fire. FDCs will routinely shift howitzers while battle tracking troops in contact. Aligning the howitzer on a potential target, known as “laying,” allows the section to get a head start if a fire mission is requested. Otherwise, and more often than not, the fire mission response will take as long as it takes the howitzer section to shift onto the target. The speed shift in the experiment’s scheme of maneuver was the latter, tied to a fire mission. The time at which the howitzer could initially respond to the target was dependent on how fast the howitzer could shift.

The out-of-traverse task, also known as a speed shift, happened in conjunction with the 3-round low-angle smoke mission. The FDC initiates the fire mission by passing the fire commands to the howitzer sections. Once the FDC stated “do not load, azimuth…,” the crew began pumping the suspension and uprooting the spades. Once the crew was balancing the howitzer on its wheels, the gunner gave the direction to shift. One of the wheels had its brake released, and the crew pushed the cannon tube in the direction the gunner stated. When the gunner was aligned with the new azimuth, he halted the crew and dropped the howitzer into position. A second speed shift was triggered after end of mission was given by the FDC. The calculations below show the elapsed time from the FDC start until the gunner’s command to drop the howitzer. Because terrain could sometimes require the Section Chief to reposition his gun, the calculations are computed based on the first drop.

This task is performed by all six Cannoneers. However, the two most physically demanding positions (and the critical billets) are the numbers 1 and 2 Cannoneers, doctrinally responsible for pumping the howitzer’s suspension. The remaining crew members are responsible for providing leverage to dislodge the spades and pushing the gun to the new azimuth.

The data collected from the out-of -traverse task was analyzed in two separate ways. The first method of comparison examined performance differences between the control group (C), the low-density group (LD), and a high-density group (HD). The second method compared performance differences between gun crews grouped by the gender of the critical billets, regardless of the crew composition. The critical billets identified for the out-of-traverse mission were the #1 and # 2 Cannoneers.

H.5.4.1 Out-of-Traverse by Integration Level Scatterplots

The scatterplots in figures H-22 and H-23 display the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling.

Figure H-22. Speed Shift #1; First Drop by Integration Level

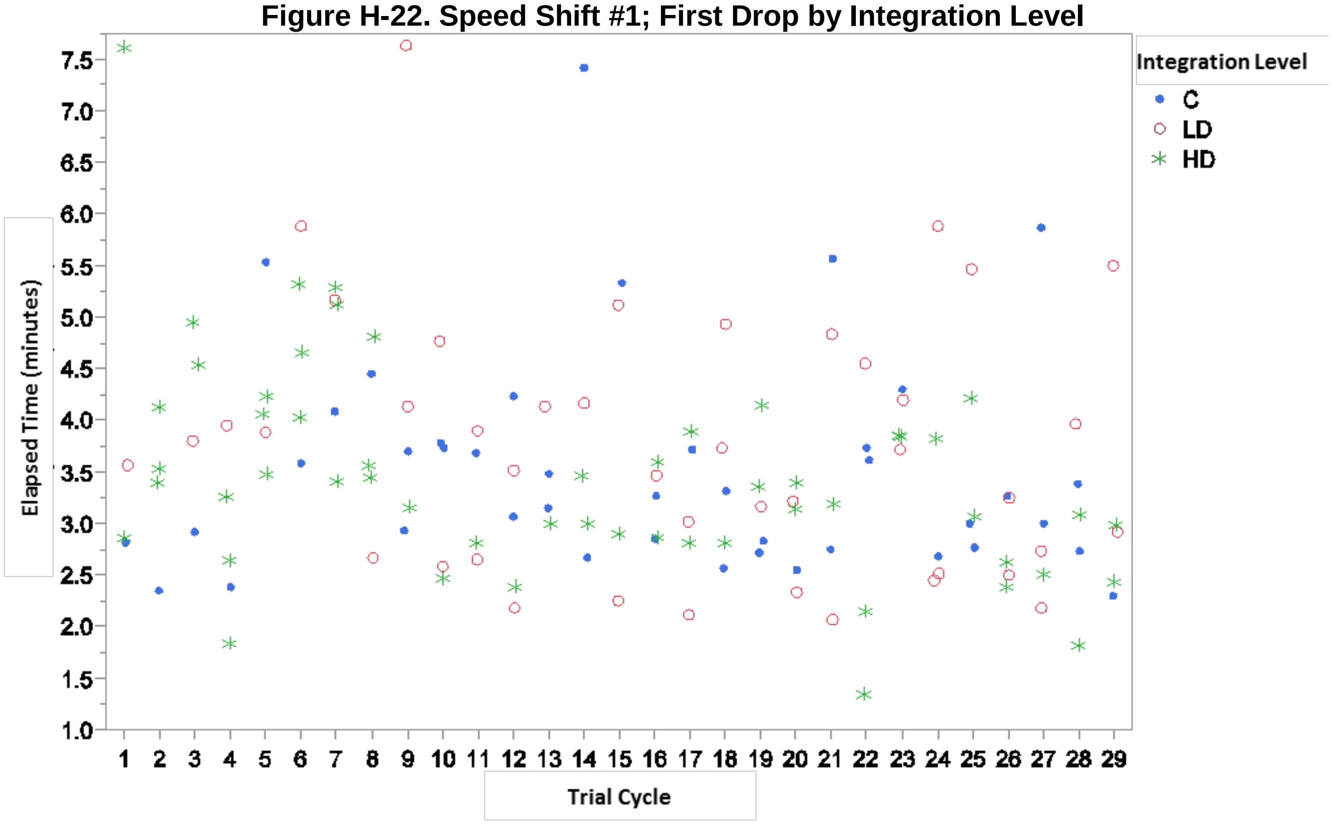
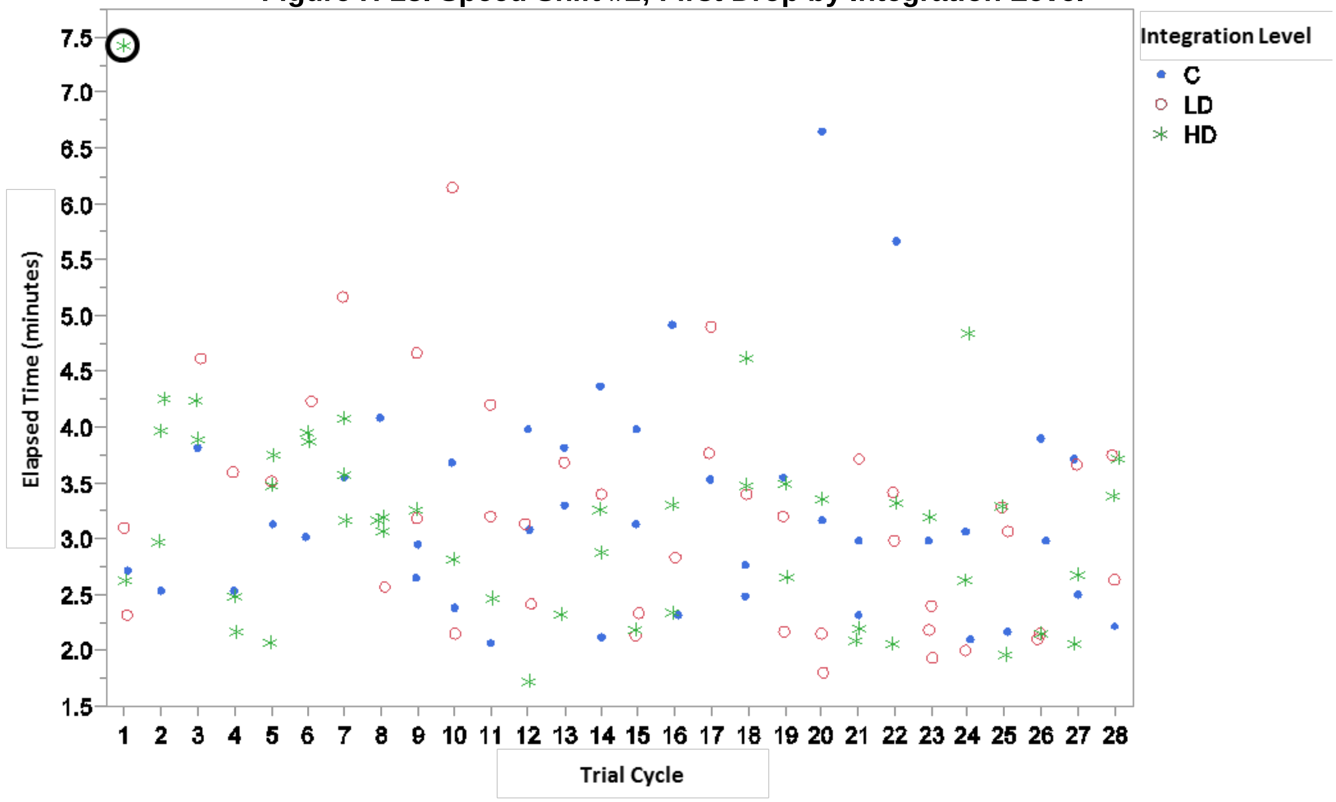


Figure H-23. Speed Shift #2; First Drop by Integration Level



H.5.4.2 Out of Traverse by Integration Level Data Table and Analysis

The tables below summarize the results of the task Out-of-Traverse analysis by Integration Level. Table H-18 compares means across metrics and integration levels. Table H-19 presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-18. Speed Shift by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Speed Shift #1; First Drop (minutes)	**C**	**42**	**3.53**	**1.08**
	LD	**43**	**3.74**	**1.25**
	HD	**54**	**3.46**	**1.05**
Speed Shift #2; First Drop (minutes)	C	41	3.24	0.96
	LD	42	3.17	0.99
	HD	49	3.16	0.98

Speed Shift #2; First Drop [Excluding Influential Points] **(minutes)**	**C**	**41**	**3.24**	**0.96**
	LD	**42**	**3.17**	**0.99**
	HD	**48**	**3.07**	**0.76**

Table H-19. Speed Shift by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Speed Shift #1; First Drop (minutes)	**0.78 (2, 136)**	**0.46**	**LD-C**	**0.21**	**5.97%**	**0.66**	**-0.21**	**0.63**	**-0.29**	**0.72**
			HD-C	**-0.07**	**-1.97%**	**0.95**	**-0.47**	**0.33**	**-0.55**	**0.41**
			HD-LD	**-0.28**	**-7.49%**	**0.44**	**-0.68**	**0.12**	**-0.76**	**0.20**
Speed Shift #2; First Drop (minutes)	0.08 (2, 129)	0.92	LD-C	-0.07	-2.10%	0.95	-0.44	0.30	-0.51	0.38
			HD-C	-0.08	-2.41%	0.92	-0.43	0.28	-0.51	0.35
			HD-LD	-0.01	-0.32%	1.00	-0.36	0.34	-0.43	0.41
Speed Shift #2; First Drop [Excluding Influential Points] (minutes)	0.39 (2, 128)	0.68	LD-C	-0.07	-2.14%	0.94	-0.41	0.27	-0.48	0.27
			HD-C	-0.17	-5.14%	0.66	-0.50	0.16	-0.56	0.16
			HD-LD	-0.10	-3.11%	0.86	-0.43	0.23	-0.49	0.23

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

H.5.4.2.1 Out of Traverse by Integration Level

For Speed Shift #1, the C and HD groups' data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01, but the LD group's data is normally distributed with a p-value of 0.02. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA

For Speed Shift #1, the C group had a mean time of 3.53 minutes; the LD group, 3.74 minutes, and the HD group, 3.46 minutes. No comparison is statistically significant.

For Speed Shift #2, the C and HD groups' data are not normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of < 0.01, but the LD group's data is normally distributed with a p-value of 0.01. We proceed with presenting ANOVA results because sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA

For Speed Shift #2, the C group had a mean time of 3.24 minutes; the LD group, 3.17 minutes; and the HD group, 3.16 minutes. No comparison is statistically significant.

For Speed Shift #2 [Excluding Influential Points], the C group had a mean time of 3.24 minutes; the LD group, 3.17 minutes; and the HD group 3.07 minutes. No comparison is statistically significant.

H.5.4.3 Speed Shift by Critical Billet—#1 and #2 Cannoneers

This second method of comparison for the Out-of-Traverse task examines performance differences between gun crews grouped only by the gender of the Marines serving in two critical billets. In this analysis, the first critical billet is the Cannoneer #1, and it is denoted by either a (M) for male or a (F) for female. The second critical billet is the Cannoneer #2 and, again, is also depicted by a (M) or (F) to denote the gender of the Marine. The tables below display the Cannoneer #1 and #2 billets as a two-letter representation, where the first letter represents the gender of the Marine occupying the Cannoneer #1 billet and the second letter represents the gender of the Marine occupying the Cannoneer #2 billet.

H.5.4.4 Speed Shift by Critical Billet—#1 and #2 Cannoneer Scatterplots

The scatterplots in figures H-24 and H-25 display the data used in the analysis of the results. All data points shown in the scatterplots were determined to be valid and used in the analysis and modeling.

Figure H-24. Speed Shift #1; First Drop by Critical Billet

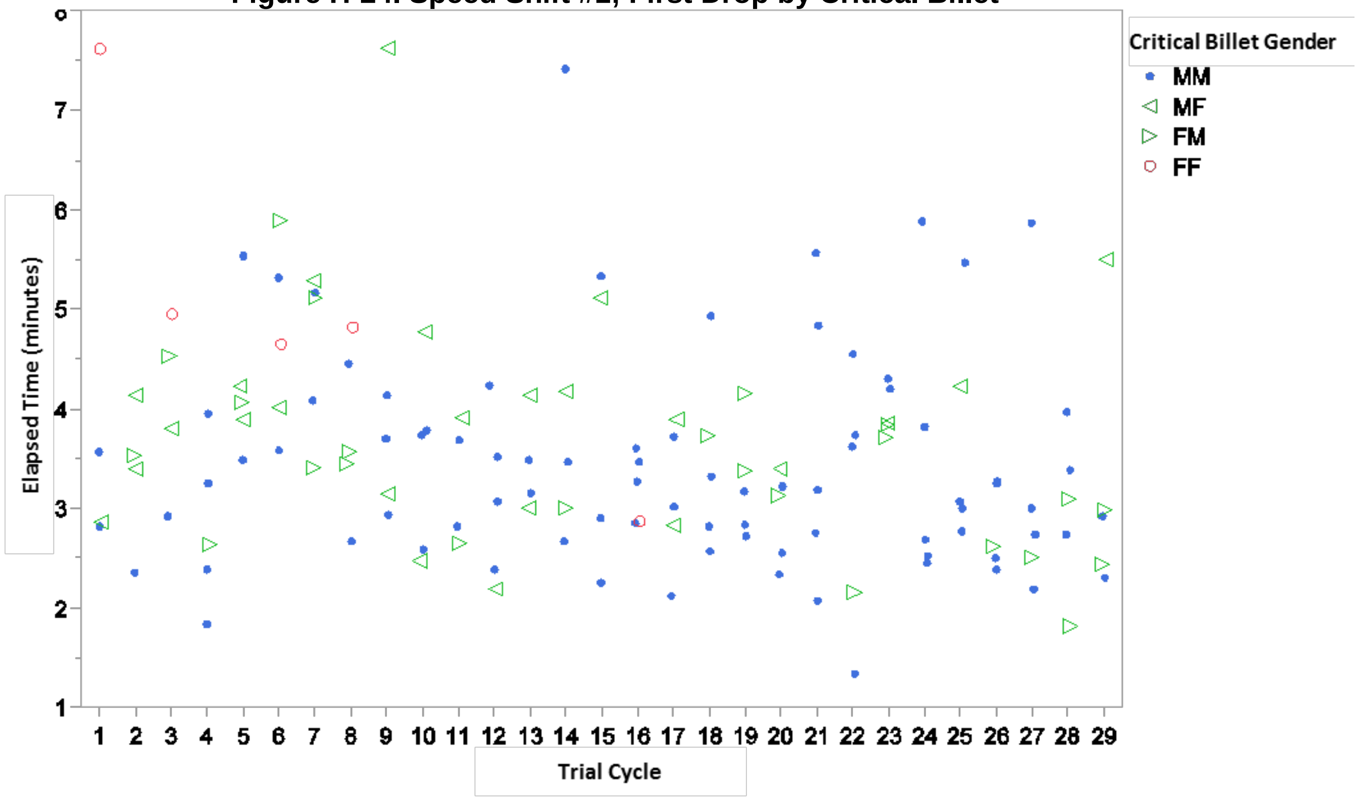
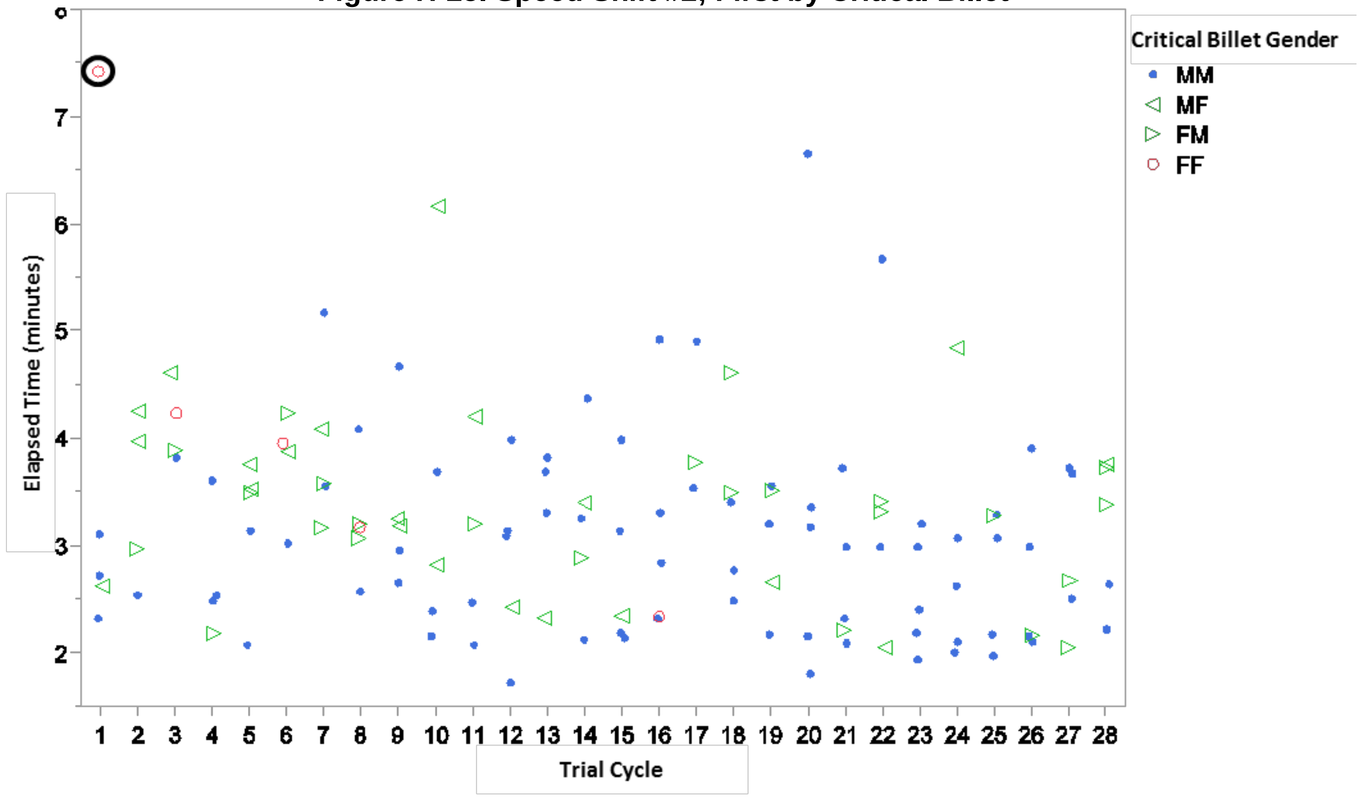


Figure H-25. Speed Shift #2; First by Critical Billet



H.5.4.5 Speed Shift by Critical Billet—#1 and #2 Cannoneers Data Table Analysis

The tables below summarize the results of the task Out-of-Traverse analysis by Critical Billets. Table H-20 compares means across metrics and integration levels. Table H-21 presents ANOVA and Tukey test results bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-20. Speed Shift Results by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Speed Shift #1; First Drop (minutes)	MM	86	3.41	1.06
	MF	24	3.99	1.14
	FM	24	3.39	0.93
	FF	5	4.98	1.70
Speed Shift #2; First Drop (minutes)	MM	82	3.03	0.92
	MF	21	3.52	1.00

	FM	**24**	**3.22**	**0.64**
	FF	**5**	**4.22**	**1.93**
Speed Shift #2; First Drop [Excluding Influential Points] **(minutes)**	**MM**	**82**	**3.03**	**0.92**
	MF	**21**	**3.52**	**1.00**
	FM	**24**	**3.22**	**0.64**
	FF	**4**	**3.42**	**0.85**

Table H-21. Speed Shift by Critical Billet ANOVA and Tukey Test

ANOVA and Tukey Test Results**Metric**	**F Statistic (df)**	**F Test P-Value**	**Comparison**	**Difference**	**% Difference**	**P-Value**	**80 % LCB**	**80% UCB**	**90% LCB**	**90% UCB**
Speed Shift #1; First Drop **(minutes)**	4.92 (3, 135)	< 0.01*	MF-MM	0.58	17.03%	0.10*	0.09	1.07	0.01	1.16
			FM-MM	-0.02	-0.60%	1.00	-0.51	0.47	-0.60	0.56
			FF-MM	1.57	46.00%	0.01*	0.59	2.55	0.42	2.72
			MF-FM	0.60	17.74%	0.22	-0.02	1.22	-0.12	1.32
			FF-MF	0.99	24.76%	0.25	-0.06	2.04	-0.24	2.21
			FF-FM	1.9	46.89%	0.02*	0.54	2.64	0.36	2.82
Speed Shift #2; First Drop **(minutes)**††	7.45	0.06*	MF-MM	0.49	16.23%	0.05	0.18	0.81	0.08	0.90
			FM-MM	0.19	6.31%	0.25	-0.02	0.41	-0.09	0.47
			FF-MM	1.19	39.17%	0.24	-0.14	2.52	-0.65	3.03
			MF-FM	0.30	9.34%	0.25	-0.03	0.63	-0.13	0.73
			FF-MF	0.70	19.73%	0.47	-0.64	2.03	-1.14	2.54

			FF-FM	1.00	30.91%	0.32	-0.33	2.33	-0.84	2.84
Speed Shift #2; First Drop [Excluding Influential Points] **(minutes)**	1.90 (3, 127)	0.13	MF-MM	0.49	16.23%	0.11	0.06	0.92	-0.01	1.00
			FM-MM	0.19	6.31%	0.79	-0.22	0.60	-0.29	0.67
			FF-MM	0.39	12.81%	0.83	-0.51	1.29	-0.67	1.44
			MF-FM	0.30	9.34%	0.67	-0.23	0.83	-0.31	0.92
			FF-MF	-0.10	-2.94%	1.00	-1.07	0.86	-1.23	1.02
			FF-FM	0.20	6.12%	0.98	-0.75	1.15	-0.91	1.31

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.
††Indicates results presented are from Robust ANOVA and Welch's t-tests with p-values compared to 0.0167 for Bonferroni adjustment. F-statistics is a Chi-square statistic from Robust ANOVA, and the F-test Pp-value is from a Robust ANOVA. The P-values are two-sided p-values from Welch's t-tests and the CIs are from Welch's t-tests

H.5.4.5.1 Out of Traverse by Critical Billet—#1 and #2 Cannoneers

For Speed Shift #1, the MM group data is not normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of < 0.01, but the FF, FM, and MF groups' data are normally distributed with p-values of 0.42, 0.38, and 0.03. We proceed with presenting ANOVA results because they were confirmed by a Kruskal-Wallis Test with a p-value of < 0.01. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA

For Speed Shift #1, the tables above show that, on average, the MM group had a mean time of 3.41 minutes. The MF group had a mean time of 3.99 minutes (17.03% slower than the MM group), and the difference is statistically significant. The FM group had a mean time of 3.39 minutes (0.60% faster than the MM group), and the difference is not statistically significant. The FF group had a mean time of 4.98 minutes (46.00% slower than the MM group), and the difference is statistically significant.

For Speed Shift #2, the MM group data is not normally distributed as evidenced by the Shapiro-Wilk Test that results in a p-value of < 0.01, but the FF, FM, and MF groups' data are normally distributed with p-values of 0.33, 0.31, and 0. 83. In addition, the FF group standard deviation is more than twice that of the FM group. As such, we recommend using the robust ANOVA results shown above. We proceed with presenting robust ANOVA results because they were confirmed by a Kruskal-Wallis Test with a p-value of 0.02.

For Speed Shift #2, the tables above show that, on average, the MM group had a mean time of 3.03 minutes. The MF group had a mean time of 3.52 minutes (16.23% slower than the MM group), and the difference between is not statistically significant. The FM group had a mean time of 3.22 minutes (6.31% slower than the MM group), and the difference is not statistically significant. The FF group had a mean time of 4.22 minutes (39.17% slower than the MM group), and the difference is statistically significant.

Although the data seem to indicate that it matters who is in the #1 position as compared to #2 (MF vs FM), it is important to note that these billets are essentially mirror images of each other. There is no advantage to dominant hand or bodily positioning in relation to the lever. In both cases, the Marine is operating a suspension lever (a second-class lever) with the fulcrum near the bottom of the howitzer and the effort approximately 4.5 feet off the deck.

For Speed Shift #2, if the Influential Points are excluded, the FF group's mean time drops significantly, averaging 3.42 minutes. The MM and FF group comparison is no longer statistically significant and drops to a 12.81% difference.

H.5.4.5.2 Out-of-Traverse Limits Contextual Comments

It is important to note that, when the Out-of-Traverse task was analyzed by integration level, there was no statistical significance between the three groups. However, when the task is analyzed by critical billet, there was a statistical difference between the MF and the FF groups.

This task is particularly vulnerable to the impacts of two individual Marines. The task itself does not lend itself to compensation or adaptation, as all Marines tend to be completely occupied with individual and group responsibilities. This is highlighted by the disparities in results between integration level and critical billet makeup.

H.5.5 Displacement by Integration Level

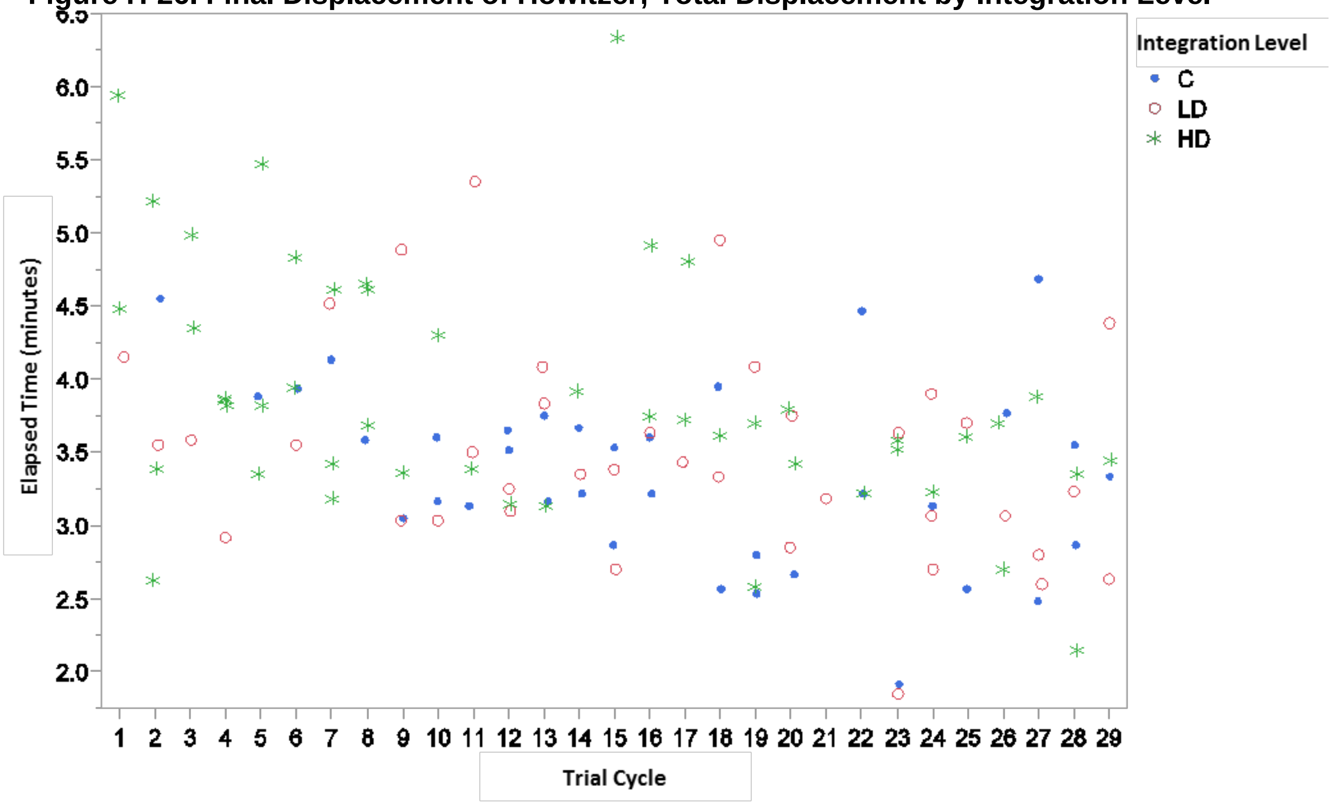
Two displacements were captured during each trial. The first displacement began at fire position one and required the volunteers to upload ammunition and retrieve the crew serve weapon in addition to displacing the howitzer. The second displacement was located at fire position two and required the volunteers to displace only the howitzer. The first displacement time hacks were complicated by the additional tasks. Therefore, the second displacement, also known as the final displacement, was chosen for primary analysis because it was purely the displacement task without additional subtasks embedded.

The actions completed by the howitzer section included the #1 Cannoneer engaging the traverse lock; #4 Cannoneer inserting the trident bar into the lunette assembly; and #1 and #2 Cannoneers pumping the suspension to ride height, attaching the howitzer to the prime mover, stowing the trail arms, and finally climbing into the back of the prime mover. The time for this task started when the Section Chief gave the order to begin and finished when the last volunteer was seated in the back of the truck. The results for total displacement time for the final displacement are shown below. See the Artillery Appendix for the first displacement, trident bar, and trail arm calculations.

H.5.6 Displacement by Integration Level Scatterplot

The scatterplot in Figure H-26 displays the data used in the analysis of the results. All data points shown were determined to be valid and used in the analysis and modeling.

Figure H-26. Final Displacement of Howitzer; Total Displacement by Integration Level



H.5.7 Displacement by Integration Level Data Table and Analysis

The tables below summarize the results of the task Displacement. Table H-22 compares means across metrics and integration levels. Table H-23 presents ANOVA and Tukey test results, bringing into focus those metrics that resulted in statistical significance along with their percentage differences.

Table H-22. Displacement by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Final Displacement of Howitzer; Total Displacement (minutes)	C	35	3.36	0.62
	LD	38	3.49	0.71
	HD	48	3.88	0.84

Table H-23. Displacement by Integration Level ANOVA and Tukey Test

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Final Displacement of Howitzer; Total Displacement (minutes)	5.63 (2, 118)	< 0.01*	LD-C	0.12	3.71%	0.76	-0.18	0.43	-0.24	0.49
			HD-C	0.52	15.43%	< 0.01*	0.23	0.80	0.18	0.86
			HD-LD	0.39	11.30%	0.04*	0.11	0.67	0.06	0.73

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a non-parametric equivalent test.

The C, LD, and HD groups' data are normally distributed as evidenced by the Shapiro-Wilk Test that results in p-values of 0.87, 0.02, and 0.33. In addition, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA

The tables above show that, on average, the C group had a mean time of 3.36 minutes; the LD group, 3.49 minutes; and the HD group, 3.88 minutes. On average, the LD group was 3.71% slower than the C group, and the difference between the LD and C group is not statistically significant in a Tukey test. On average, the HD group was 15.43% slower than the C group, and the difference between the HD and C group is statistically significant.

H.5.7.1 Displacement by Integration Level Contextual Comments

The displacement of the howitzer is critical to the survivability of the battery and maintaining responsive fires. During battle against an opponent capable of counterbattery fire, the ability to execute a fire mission and displace in a timely manner is vital. Furthermore, in a fluid battlespace with evolving friendly and enemy situations, the requirement to move the battery occurs often. The speed at which the battery can emplace and displace is directly tied to firing capability and responsiveness.

H.6 Modeling Results

H.6.1 Statistical Modeling Results

H.6.1.1 Statistical Modeling Results Overview

The previous section discussed results only as they pertain to differences due to integration level and gender in critical billets. The goal of statistical modeling as applied here is to estimate, simultaneously, the effect of gender integration levels and other relevant variables on crew performance. This section describes all significant variables in the model and their positive or negative correlation with the result. A negative correlation indicates that an increase in a particular variable will result in a decrease in the response variable, which is a desired outcome for elapsed time for some tasks. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

The descriptive statistics section above covers a total of 16 tasks. This section presents an overview of the analysis and modeling results for 11 out of the 16 tasks. The 11 selected modeled tasks were identified as necessary to inform the time-sensitive decision by Marine Corps leadership.

Table H-24. Patch Numbers and Billet Titles for the Section

Patch Number	Billet Title
1	Cannoneer #1 / A-Gunner
2	Cannoneer #2
3	Cannoneer #3
4	Cannoneer #4/5
5	Recorder
6	Driver

H.6.1.2 0811 Selected Tasks Method of Analysis

Due to the large number of personnel variables that need to be included in the model (six covariates, per variable), a mixed-effects model with all volunteer section members and all types of personnel data could not be run. Thus, we model each personnel variable with integration level separately with a random effect for who filled each position within the section. For example, age for each volunteer member of the section (six variables) and integration level are modeled with the result (elapsed time) as the response variable. Where maximum likelihood estimation converged, AIC was used for variable selection. Otherwise, we comment on the significance of individual variables in the full model. Variables reported as significant are concluded to be significant based on at least a one-sided test.

H.6.1.3 Selected Tasks Overall Modeling Results

There are no personnel data variables that are both statistically significant and have a practical impact to the model. Each time personnel data variables are statistically significant in a model, their effects are practically negligible, conflicting, and/or incomplete for the section.

Integration level appears in most of the tasks where a critical billet is statistically significant, and its effect is clear, causal, and practical. Therefore, integration level is the best variable to describe performance for most of these tasks. CFT Maneuver Under Fire also appears to be common for certain billets in certain tasks. Refer to the descriptive statistics section for the ANOVA summary for each of the 0811 tasks mentioned below .

H.6.1.3.1 First Emplacement Ammo Truck Offload

We model elapsed time for the First Emplacement's Ammo Truck Offload as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

Gender integration levels are significant and positively correlated with the First Emplacement Truck Offload time for the models that include the following personnel variables:

- None

Gender integration levels are significant and negatively correlated with the First Emplacement Truck Offload time for the models that include the following personnel variables:

- None

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- Height of the Recorder
- CFT Maneuver under Fire time for Cannoneer #1

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- AFQT Score of Cannoneer #4, Driver, and Recorder
- Weight of Cannoneer #3 and Cannoneer #4

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only the intercept, meaning that we could not determine a mixed model that fits the data well. We recommend referring back to ANOVA results in Table X for differences between integration groups.

**Cannoneers #3 and 4 have no involvement in this task and have no impact on its completion.

H.6.1.3.2 Ammo Resupply: Ammo Movement

We model elapsed time for the Ammo Resupply as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

Gender integration levels are significant and positively correlated with the Ammo Resupply time for the models that include the following personnel variables:

- None

Gender integration levels are significant and negatively correlated with the Ammo Resupply time for the models that include the following personnel variables:

- None

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- CFT Maneuver Under Fire Time of Cannoneer #1, Cannoneer #4, and Recorder
- PFT: Run of Cannoneer #3
- Rifle Score of the Driver.

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- AFQT Score of the Driver

- CFT Maneuver Under Fire Time of Cannoneer #2
- PFT: Crunches of Cannoneer #4.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only the intercept, meaning that we could not determine a mixed model that fits the data well. We recommend referring back to ANOVA results in Table X for differences between integration groups.

**This event involved the Marines moving rounds independently, and the time is only representative of the slowest Marine.

H.6.1.3.3 Ammo Truck Upload

We model elapsed time for the Ammo Truck Upload as a function of each personnel variable and integration level in one single mixed model. The covariates in the model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For the model, we report statistically significant positive and negative correlations and whether we observe any patterns.

In the overall model, gender integration levels were significant and positively correlated with Ammo Truck Upload time.

The following personnel variables are significant in the overall model and are positively correlated with the Ammo Truck Upload time:

- GCT Score of Cannoneer #1 and Cannoneer #2
- CFT Movement to Contact of Cannoneer #1
- CFT Maneuver under Fire Times of Cannoneer #2, Cannoneer #4, and Recorder
- PFT: Crunches of Cannoneer #2 and Driver
- Physical Fitness Run Time of the Driver
- AFQT Score of Cannoneer #4
- Height of Cannoneer #4
- Age of the Recorder.

The following personnel variables are significant in the overall model and are negatively correlated with the Ammo Truck Upload time:

- AFQT Score of Cannoneer #1 and Cannoneer #2
- Weight of Cannoneer #1 and Cannoneer #2
- Physical Fitness Run Time of Cannoneer #1 and Cannoneer #4
- Rifle Score of Cannoneer #2, Cannoneer #3, Recorder, and Cannoneer #4
- Height of Cannoneer #3
- Age of Cannoneer #4
- GCT Score of Cannoneer #4
- PFT: Crunches of the Recorder.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only gender integration levels. Holding all other variables constant, HD groups are expected to perform this task 0.26 minute slower than C groups, while LD groups are expected to perform this task 0.35 minute slower than C groups.

**Cannoneer #3 and 4 had no involvement in this task and had no impact on its completion.

H.6.1.3.4 Final Displacement of Howitzer; Total Displacement

We model elapsed time for the Final Displacement of Howitzer as a function of each personnel variable and integration level in one single mixed model. The covariates in the model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For the model, we report statistically significant positive and negative correlations and whether we observe any patterns.

Gender integration levels were selected by the AIC into the final model of elapsed time to displace the howitzer.

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- None.

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- Weight of Cannoneer #2 and Cannoneer #3
- CFT Maneuver Under Fire Time of Cannoneer #2
- PFT: Crunches of Cannoneer #2 and Cannoneer #4
- Rifle Score of the Recorder.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only gender integration levels. Based on the model, although all other variables remain constant, a low-density integration is expected to result in displacement happening approximately 0.24 minute faster, on average, than would be done by the non-integrated group. The model indicates that, although all other variables remain constant, a high-density integration is expected to result in displacement happening approximately 0.24 minute slower, on average, than would be done by the non-integrated group.

H.6.1.3.5 3-Round High-Angle WP; Indiv. Subsequent Rounds

We model elapsed time for the 3-Round High-Angle WP task as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The critical billet MM integration level is significant and positively correlated for the models that include the following variables:

- None

The critical billet MM integration level is significant and negatively correlated for the models that include the following variables:

- Age
- AFQT Score
- GCT Score
- Weight
- Physical Fitness Test: Crunches.

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- Age of Cannoneer #2
- AFQT Score of Cannoneer #2 and Cannoneer #3
- GCT Score of Cannoneer #2
- CFT Maneuver under Fire Time of Cannoneer #2 and Cannoneer #4
- Physical Fitness Run Time of Cannoneer #2 and Cannoneer #4.

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- AFQT Score of the Driver
- Height of Cannoneer #2.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only Critical Billet. Based on the model, although all other variables remain constant, a MM critical billet is expected to perform this task approximately 0.19 minute faster, on average, than would be done by a FF critical billet. The FM and MF levels of critical billet are insignificantly different from the FF level.

H.6.1.3.6 3-Round Low-Angle HE; Indiv. Subsequent Rounds

We model elapsed time for the 3-Round Low-Angle HE task as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The critical billet was significant and positively correlated with males for the models that include that following variables:

- None

The critical billet was significant and negatively correlated with males for the models that include that following variables:

- Age
- Weight
- Physical Fitness Test: Crunches
- Rifle Score.

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- Age of Cannoneer #2
- Physical Fitness Run Time of Cannoneer #4.

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- AFQT Score of Cannoneer #3 and Driver
- HEIGHT of Cannoneer #2, Cannoneer #3, and Cannoneer #4
- Weight of Cannoneer #3
- CFT Maneuver under Fire Time of the Driver
- PFT: Crunches of Cannoneer #4 and Driver.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only Critical Billet. Based on the model, although all other variables remain constant, a male is expected to perform this task approximately 0.11 minute faster, on average, than would be done by a female.

**Recorder's physical condition has no impact on this activity, as this Marine is sitting for the entire evolution.

H.6.1.3.7 9-Round Low-Angle HE; Indiv. Subsequent Rounds

We model elapsed time for the 9-Round Low-Angle HE task as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The critical billet M level is significant and positively correlated for the models that include the following variables:

- None

The critical billet M level is significant and negatively correlated for the models that include the following variables:

- Age
- AFQT Score
- GCT Score
- Physical Fitness Test: Crunches
- Physical Fitness Test: Run Time
- Rifle score.

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- Age of Cannoneer #1
- CFT Maneuver under Fire Time of Cannoneer #2 and Cannoneer #4
- Physical Fitness Run Time of Cannoneer #2.

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- AFQT Score of Cannoneer #4
- GCT Score of Cannoneer #4
- Height of Cannoneer #2 and Cannoneer #4
- Weight of Cannoneer #4
- CFT Maneuver under Fire Time of the Driver.

Because we did not identify any discernable patterns in the effects of the personnel variables, and because their effects are often negligible, our final model includes only Critical Billet. Based on the model, although all other variables remain constant, a male is expected to perform this task approximately 0.08 minute faster, on average, than would be done by a female.

H.6.1.3.8 3-Round Low-Angle WP; Indiv. Subsequent Rounds

We model elapsed time for the 3-Round Low-Angle WP task as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The M Critical Billet level is significant and positively correlated for the models that include the following variable:

- None

The M Critical Billet level is significant and negatively correlated for the models that include the following variable:

- None

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- AFQT Score of Cannoneer #2
- GCT Score of Cannoneer #1 and Cannoneer #2
- Height of the Recorder
- CFT Maneuver under Fire Time of Cannoneer #4.

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- Weight of Cannoneer #4
- Rifle Score of the Driver.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only Critical Billet. Based on the model, although all other variables remain constant, a male is expected to perform this task approximately 0.04 minute faster, on average, than would be done by a female.

H.6.1.3.9 3-Round Low-Angle HE from FP2; Indiv. Subsequent Rounds

We model elapsed time for the 3-Round Low-Angle HE from FP2 task as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For each model, we report statistically significant positive and negative correlations and whether we observe any patterns.

The critical billet M level is significant and positively correlated for the models that include the following variables:

- None

The critical billet M level is significant and negatively correlated for the models that include the following variables:

- Age
- AFQT Score
- GCT Score
- Height
- Weight
- Physical Fitness Test: Crunches
- Physical Fitness Test: Run Time
- Rifle Score.

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- Age of Cannoneer #3 and Cannoneer #4
- AFQT Score of Cannoneer #3
- GCT Score of Cannoneer #3
- Weight of the Recorder
- CFT Maneuver under Fire Time of Cannoneer #3 and Cannoneer #4
- PFT: Crunches of Cannoneer #3.

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- AFQT Score of the Recorder
- GCT Score of the Recorder
- Height of Cannoneer #3
- Weight of Cannoneer #3
- PFT: Crunches of the Driver
- Physical Fitness Run Time of Cannoneer #1.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only Critical Billet. Based on the model, although all other variables remain constant, a male is expected to perform this task approximately 0.09 minute faster, on average, than would be done by a female.

H.6.1.3.10 Speed Shift #1; First Drop

We model elapsed time for the Speed Shift #1; First Drop as a function of each personnel variable and integration level in one single mixed model. The covariates in the model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For the model, we report statistically significant positive and negative correlations and whether we observe any patterns.

Critical Billet levels were selected by the AIC into the final model for the elapsed time of Speed Shift 1.

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- Age of Cannoneer #2
- GCT Score of Cannoneer #2
- CFT Maneuver Under Fire Time of Cannoneer #3.

The following personnel variables are significant in their respective models and are negatively correlated with the First Emplacement Truck Offload time:

- Height of Cannoneer #4
- PFT: Run Time of Cannoneer #1.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only Critical Billet levels. Based on the model, although all other variables remain constant, a FM group is expected to perform this task approximately 1.62 minutes faster, on average, than would be done by the FF group. Again holding all other variables constant, a MF group is expected to perform this task approximately 1.02 minutes faster, on average, than would be done by the FF group. Again holding all other variables constant, a MM group is expected to perform this task approximately 1.58 minutes faster, on average, than would be done by the FF group.

H.6.1.3.11 Speed Shift #2; First Drop

We model elapsed time for the Speed Shift #2; First Drop as a function of each personnel variable and integration level in one single mixed model. The covariates in the model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each volunteer position on the section. For the model, we report statistically significant positive and negative correlations and whether we observe any patterns.

Critical Billet levels were selected by the AIC into the final model for the elapsed time of Speed Shift 1.

Critical Billet levels were selected by the AIC into the final model for the elapsed time of Speed Shift 2.

The following personnel variables are significant in their respective models and are positively correlated with the First Emplacement Truck Offload time:

- GCT Score of Cannoneer #3
- CFT Maneuver Under Fire Time of Cannoneer #1 and Cannoneer #3
- PFT: Run of the Recorder.

The following personnel variable is significant in its respective models and is negatively correlated with the First Emplacement Truck Offload time:

- PFT: Crunches of Cannoneer #1.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes only Critical Billet levels. Based on the model, although all other variables remain constant, a FM group is expected to perform this task approximately 1.03 minutes faster, on average, than would be done by the FF group. Again holding all other variables constant, a MF group is expected to perform this task approximately 0.74 minute faster, on average, than would be done by the FF group. Again holding all other variables constant, a MM group is expected to perform this task approximately 1.15 minutes faster, on average, than would be done by the FF group.

Appendix to Annex H
0811 Supplemental Information

This appendix provides supplemental information for the 0811 portion of the GCEITF experiment. It provides information regarding additional descriptive and basic inferential statistics not described in Annex H.

Section 1: Additional Tasks Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for additional 0811 tasks. Annex H contains the descriptive statistics for the remainder of the 0811 tasks. The words “metric” and “task” are used interchangeably throughout this Appendix; They both refer to the experimental task.

The tables in this appendix display results for the additional 0811 metrics, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels, and ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Section 2: Additional Task Results

Emplacement of Howitzer:

Emplacing Howitzer; Dismount (seconds) - The C group had a mean of 31.73 seconds with a SD of 6.99 seconds. The LD group had a mean of 40.64 seconds with a SD of 18.25 seconds. The HD group had a mean of 37.41 seconds with a SD of 15.57 seconds. The LD group was 28.06% slower than the C group. The HD group was 17.9% slower than the C group. The HD group was 7.94% faster than the LD group.

Emplacing Howitzer; Trident Bar Drop (seconds) - The C group had a mean of 6.27 seconds with a SD of 3.3 seconds. The LD group had a mean of 6.96 seconds with a SD of 4.93 seconds. The HD group had a mean of 6.71 seconds with a SD of 4.65 seconds. The LD group was 10.85% slower than the C group. The HD group was 6.96% slower than the C group. The HD group was 3.51% faster than the LD group.

Emplacing Howitzer; Trident Bar Drop Excluding Influential Points (seconds)- The C group had a mean of 6.27 seconds with a SD of 3.3 seconds. The LD group had a mean of 6.48 seconds with a SD of 3.79 seconds. The HD group had a mean of 6.33

seconds with a SD of 3.81 seconds. The LD group was 3.22% slower than the C group. The HD group was 0.93% slower than the C group. The HD group was 2.22% faster than the LD group.

Hasty Emplacement of Howitzer; Dismount (seconds) - The C group had a mean of 34.59 seconds with a SD of 17.28 seconds. The LD group had a mean of 41.26 seconds with a SD of 25.13 seconds. The HD group had a mean of 38.43 seconds with a SD of 22.79 seconds. The LD group was 19.3% slower than the C group. The HD group was 11.13% slower than the C group. The HD group was 6.85% faster than the LD group.

Hasty Emplacement of Howitzer; Trident Bar Drop (seconds) - The C group had a mean of 8.33 seconds with a SD of 7.45 seconds. The LD group had a mean of 8.37 seconds with a SD of 5.75 seconds. The HD group had a mean of 8.25 seconds with a SD of 8.52 seconds. The LD group was 0.43% slower than the C group. The HD group was 1.06% faster than the C group. The HD group was 1.48% faster than the LD group.

Hasty Emplacement of Howitzer; Trident Bar Drop Excluding Influential Points (seconds) - The C group had a mean of 7.51 seconds with a SD of 5.28 seconds. The LD group had a mean of 8.37 seconds with a SD of 5.75 seconds. The HD group had a mean of 7.35 seconds with a SD of 5.5 seconds. The LD group was 11.41% slower than the C group. The HD group was 2.21% faster than the C group. The HD group was 12.23% faster than the LD group.

Emplacing Howitzer; Trident Bar Drop (seconds) - The M group had a mean of 6.11 seconds with a SD of 3.64 seconds. The F group had a mean of 9.43 seconds with a SD of 6.4 seconds. The F group was 54.54% slower than the M group.

Emplacing Howitzer; Trident Bar Drop Excluding Influential Points (seconds) - The M group had a mean of 5.91 seconds with a SD of 3.01 seconds. The F group had a mean of 8.68 seconds with a SD of 5.41 seconds. The F group was 46.86% slower than the M group.

Hasty Emplacement of Howitzer; Trident Bar Drop (seconds) - The M group had a mean of 7.68 seconds with a SD of 6.08 seconds. The F group had a mean of 11.42 seconds with a SD of 11.36 seconds. The F group was 48.75% slower than the M group.

Hasty Emplacement of Howitzer; Trident Bar Drop Excluding Influential Points (seconds) - The M group had a mean of 7.38 seconds with a SD of 5.19 seconds. The F group had a mean of 9.52 seconds with a SD of 6.69 seconds. The F group was 29.03% slower than the M group.

Table H A - Emplacement Results (Descriptive Statistics by Integration Level)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Emplacing Howitzer; Dismount (seconds)	C	41	31.73	6.99	28.06%	17.9%	-7.94%
	LD	44	40.64	18.25			
	HD	51	37.41	15.57			
Emplacing Howitzer; Trident Bar Drop (seconds)	C	40	6.27	3.3	10.85%	6.96%	-3.51%
	LD	45	6.96	4.93			
	HD	52	6.71	4.65			
Emplacing Howitzer; Trident Bar Drop Excluding Influential Points (seconds)	C	40	6.27	3.3	3.22%	0.93%	-2.22%
	LD	44	6.48	3.79			
	HD	51	6.33	3.81			
Hasty Emplacement of Howitzer; Dismount (seconds)	C	41	34.59	17.28	19.3%	11.13%	-6.85%
	LD	42	41.26	25.13			
	HD	53	38.43	22.79			
Hasty Emplacement of Howitzer; Trident Bar Drop (seconds)	C	42	8.33	7.45	0.43%	-1.06%	-1.48%
	LD	46	8.37	5.75			
	HD	53	8.25	8.52			
Hasty Emplacement of Howitzer; Trident Bar Drop Excluding Influential Points (seconds)	C	41	7.51	5.28	11.41%	-2.21%	-12.23%
	LD	46	8.37	5.75			
	HD	52	7.35	5.5			

Table H B - Emplacement Results (Descriptive Statistics by critical billet)

Metric	Integration Level	Sample Size	Mean	SD	% Difference from control	80% LCB	80% UCB	90% LCB	90% UCB
Emplacing Howitzer; Trident Bar Drop (seconds)	M	114	6.11	3.64					
	F	23	9.43	6.40	54.54%	1.52	5.14	0.98	5.68
Emplacing Howitzer; Trident Bar Drop Excluding Influential Points (seconds)	M	113	5.91	3.01					
	F	22	8.68	5.41	46.86%	1.20	4.34	0.74	4.80

Metric	Integration Level	Sample Size	Mean	SD	% Difference from control	80% LCB	80% UCB	90% LCB	90% UCB
Hasty Emplacement of Howitzer; Trident Bar Drop (seconds)	M	117	7.68	6.08					
	F	24	11.42	11.36	48.75%	0.60	6.88	-0.33	7.81
Hasty Emplacement of Howitzer; Trident Bar Drop Excluding Influential Points (seconds)	M	116	7.38	5.19					
	F	23	9.52	6.69	29.03%	0.21	4.08	-0.37	4.65

Indirect Fire Missions:

3-Round Low-Angle HE; Total Fire Mission (minutes) - The C group had a mean of 1.89 minutes with a SD of 0.38 minutes. The LD group had a mean of 2.01 minutes with a SD of 0.46 minutes. The HD group had a mean of 2.1 minutes with a SD of 0.47 minutes. The LD group was 6.51% slower than the C group. The HD group was 11.27% slower than the C group. The HD group was 4.47% slower than the LD group.

3-Round Low-Angle HE; 1st Round Response (minutes) - The C group had a mean of 1.06 minutes with a SD of 0.28 minutes. The LD group had a mean of 1.12 minutes with a SD of 0.28 minutes. The HD group had a mean of 1.16 minutes with a SD of 0.39 minutes. The LD group was 6.13% slower than the C group. The HD group was 9.74% slower than the C group. The HD group was 3.4% slower than the LD group.

3-Round Low-Angle HE; 1st Round Response Excluding Influential Points - The C group had a mean of 1.06 seconds with a SD of 0.28 seconds. The LD group had a mean of 1.12 seconds with a SD of 0.28 seconds. The HD group had a mean of 1.13 seconds with a SD of 0.34 seconds. The LD group was 6.13% slower than the C group. The HD group was 7.02% slower than the C group. The HD group was 0.84% slower than the LD group.

3-Round High-Angle WP; Total Fire Mission (minutes) - The C group had a mean of 3.34 minutes with a SD of 0.48 minutes. The LD group had a mean of 3.5 minutes with a SD of 0.63 minutes. The HD group had a mean of 3.65 minutes with a SD of 0.84 minutes. The LD group was 4.81% slower than the C group. The HD group was 9.25% slower than the C group. The HD group was 4.24% slower than the LD group.

3-Round High-Angle WP; 1st Round Response (minutes) - The C group had a mean of 1.68 minutes with a SD of 0.37 minutes. The LD group had a mean of 1.77 minutes with a SD of 0.33 minutes. The HD group had a mean of 1.63 minutes with a SD of 0.33 minutes. The LD group was 5.45% slower than the C group. The HD group was 2.68% faster than the C group. The HD group was 7.71% faster than the LD group.

3-Round High-Angle WP; 1st Round Response Excluding Influential Points - The C group had a mean of 1.64 seconds with a SD of 0.27 seconds. The LD group had a mean of 1.75 seconds with a SD of 0.3 seconds. The HD group had a mean of 1.63 seconds with a SD of 0.33 seconds. The LD group was 6.73% slower than the C group. The HD group was 0.17% faster than the C group. The HD group was 6.46% faster than the LD group.

9-Round Low-Angle HE; Total Fire Mission (minutes) - The C group had a mean of 4.15 minutes with a SD of 0.6 minutes. The LD group had a mean of 4.28 minutes with a SD of 1.01 minutes. The HD group had a mean of 4.31 minutes with a SD of 0.63 minutes. The LD group was 3.12% slower than the C group. The HD group was 4.02% slower than the C group. The HD group was 0.88% slower than the LD group.

9-Round Low-Angle HE; 1st Round Response (minutes) - The C group had a mean of 1.12 minutes with a SD of 0.27 minutes. The LD group had a mean of 1.05 minutes with a SD of 0.31 minutes. The HD group had a mean of 1.06 minutes with a SD of 0.21 minutes. The LD group was 6.27% faster than the C group. The HD group was 5.3% faster than the C group. The HD group was 1.03% slower than the LD group.

9-Round Low-Angle HE; 1st Round Response Excluding Influential Points - The C group had a mean of 1.12 seconds with a SD of 0.27 seconds. The LD group had a mean of 1.02 seconds with a SD of 0.27 seconds. The HD group had a mean of 1.06 seconds with a SD of 0.21 seconds. The LD group was 8.6% faster than the C group. The HD group was 5.3% faster than the C group. The HD group was 3.61% slower than the LD group.

3-Round Low-Angle WP; Total Fire Mission (minutes) - The C group had a mean of 1.48 minutes with a SD of 0.29 minutes. The LD group had a mean of 1.54 minutes with a SD of 0.34 minutes. The HD group had a mean of 1.58 minutes with a SD of 0.32 minutes. The LD group was 4.17% slower than the C group. The HD group was 7.24% slower than the C group. The HD group was 2.95% slower than the LD group.

3-Round Low-Angle WP; 1st Round Response (seconds) - The C group had a mean of 43.29 seconds with a SD of 11.64 seconds. The LD group had a mean of 46.45 seconds with a SD of 18.31 seconds. The HD group had a mean of 42.36 seconds with a SD of 13.86 seconds. The LD group was 7.29% slower than the C group. The HD group was 2.16% faster than the C group. The HD group was 8.81% faster than the LD group.

3-Round Low-Angle WP; 1st Round Response Excluding Influential Points - The C group had a mean of 43.29 seconds with a SD of 11.64 seconds. The LD group had a mean of 44.73 seconds with a SD of 15.15 seconds. The HD group had a mean of 41.32 seconds with a SD of 12.12 seconds. The LD group was 3.33% slower than the C group. The HD group was 4.55% faster than the C group. The HD group was 7.63% faster than the LD group.

3-Round Low-Angle HE from FP2; Total Fire Mission (minutes) - The C group had a mean of 1.67 minutes with a SD of 0.32 minutes. The LD group had a mean of 1.69 minutes with a SD of 0.3 minutes. The HD group had a mean of 1.7 minutes with a SD of 0.33 minutes. The LD group was 1.6% slower than the C group. The HD group was 2.23% slower than the C group. The HD group was 0.62% slower than the LD group.

3-Round Low-Angle HE from FP2; 1st Round Response (seconds) - The C group had a mean of 58.75 seconds with a SD of 15.89 seconds. The LD group had a mean of 57.36 seconds with a SD of 12.65 seconds. The HD group had a mean of 57.32 seconds with a SD of 13.68 seconds. The LD group was 2.36% faster than the C group. The HD group was 2.43% faster than the C group. The HD group was 0.07% faster than the LD group.

3-Round Low-Angle HE from FP2; 1st Round Response Excluding Influential Points - The C group had a mean of 57.03 seconds with a SD of 12.78 seconds. The LD group had a mean of 57.36 seconds with a SD of 12.65 seconds. The HD group had a mean of 57.32 seconds with a SD of 13.68 seconds. The LD group was 0.58% slower than the C group. The HD group was 0.51% slower than the C group. The HD group was 0.07% faster than the LD group.

3-Round High-Angle WP; Total Fire Mission (minutes) - The MM group had a mean of 3.42 minutes with a SD of 0.59 minutes. The MF group had a mean of 3.86 minutes with a SD of 0.98 minutes. The FM group had a mean of 3.6 minutes with a SD of 0.76 minutes. The FF group had a mean of 3.59 minutes with a SD of 0.61 minutes. The MF group was 12.64% slower than the MM group. The FM group was 5.26% slower than the MM group. The FF group was 4.91% slower than the MM group. The MF group was 7.01% slower than the FM group. The FF group was 6.86% faster than the MF group. The FF group was 0.34% faster than the FM group.

3-Round High-Angle WP; 1st Round Response (minutes) - The MM group had a mean of 1.68 minutes with a SD of 0.34 minutes. The MF group had a mean of 1.79 minutes with a SD of 0.45 minutes. The FM group had a mean of 1.66 minutes with a SD of 0.27 minutes. The FF group had a mean of 1.63 minutes with a SD of 0.13 minutes. The MF group was 6.62% slower than the MM group. The FM group was 1.01% faster than the MM group. The FF group was 2.8% faster than the MM group. The MF group was 7.71% slower than the FM group. The FF group was 8.84% faster than the MF group. The FF group was 1.81% faster than the FM group.

3-Round High-Angle WP; 1st Round Response Excluding Influential Points - The MM group had a mean of 1.66 seconds with a SD of 0.31 seconds. The MF group had a mean of 1.73 seconds with a SD of 0.45 seconds. The FM group had a mean of 1.66 seconds with a SD of 0.27 seconds. The FF group had a mean of 1.63 seconds with a SD of 0.13 seconds. The MF group was 4.26% slower than the MM group. The FM group was 0.06% slower than the MM group. The FF group was 1.75% faster than the MM group. The MF group was 4.2% slower than the FM group. The FF group was 5.77% faster than the MF group. The FF group was 1.81% faster than the FM group.

Table H C - Indirect Fire Mission Results (Descriptive Statistics by integration level)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
3-Round Low-Angle HE; Total Fire Mission (minutes)	C	37	1.89	0.38	6.51%	11.27%	4.47%
	LD	38	2.01	0.46			
	HD	44	2.1	0.47			
3-Round Low-Angle HE; 1st Round Response (minutes)	C	37	1.06	0.28	6.13%	9.74%	3.4%
	LD	38	1.12	0.28			
	HD	46	1.16	0.39			
3-Round Low-Angle HE; 1st Round Response Excluding Influential Points (minutes)	C	37	1.06	0.28	6.13%	7.02%	0.84%
	LD	38	1.12	0.28			
	HD	45	1.13	0.34			
3-Round High-Angle WP; Total Fire Mission (minutes)	C	35	3.34	0.48	4.81%	9.25%	4.24%
	LD	39	3.5	0.63			
	HD	49	3.65	0.84			
3-Round High-Angle WP; 1st Round Response (minutes)	C	35	1.68	0.37	5.45%	-2.68%	-7.71%
	LD	39	1.77	0.33			
	HD	48	1.63	0.33			
3-Round High-Angle WP; 1st Round Response Excluding Influential Points (minutes)	C	34	1.64	0.27	6.73%	-0.17%	-6.46%
	LD	38	1.75	0.3			
	HD	48	1.63	0.33			
9-Round Low-Angle HE; Total Fire Mission (minutes)	C	31	4.15	0.6	3.12%	4.02%	0.88%
	LD	34	4.28	1.01			
	HD	39	4.31	0.63			
9-Round Low-Angle HE; 1st Round Response (minutes)	C	33	1.12	0.27	-6.27%	-5.3%	1.03%
	LD	35	1.05	0.31			
	HD	42	1.06	0.21			
9-Round Low-Angle HE; 1st Round Response	C	33	1.12	0.27	-8.60%	-5.3%	3.61%
	LD	34	1.02	0.27			

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Excluding Influential Points (minutes)	HD	42	1.06	0.21			
3-Round Low-Angle WP; Total Fire Mission (minutes)	C	38	1.48	0.29	4.17%	7.24%	2.95%
	LD	37	1.54	0.34			
	HD	44	1.58	0.32			
3-Round Low-Angle WP; 1st Round Response (seconds)	C	38	43.29	11.64	7.29%	-2.16%	-8.81%
	LD	38	46.45	18.31			
	HD	45	42.36	13.86			
3-Round Low-Angle WP; 1st Round Response Excluding Influential Points (seconds)	C	38	43.29	11.64	3.33%	-4.55%	-7.63%
	LD	37	44.73	15.15			
	HD	44	41.32	12.12			
3-Round Low-Angle HE from FP2; Total Fire Mission (minutes)	C	31	1.67	0.32	1.60%	2.23%	0.62%
	LD	33	1.69	0.3			
	HD	38	1.7	0.33			
3-Round Low-Angle HE from FP2; 1st Round Response (seconds)	C	32	58.75	15.89	-2.36%	-2.43%	-0.07%
	LD	33	57.36	12.65			
	HD	40	57.32	13.68			
3-Round Low-Angle HE from FP2; 1st Round Response Excluding Influential Points (seconds)	C	31	57.03	12.78	0.58%	0.51%	-0.07%
	LD	33	57.36	12.65			
	HD	40	57.32	13.68			

Table H D - Indirect Fire Mission Results (Descriptive Statistics by critical billet)

Metric	Integration Level	Sample Size	Mean	SD	% Difference from control	80% LCB	80% UCB	90% LCB	90% UCB
3-Round Low-Angle HE; Total Fire Mission (minutes)	M	101	1.97	0.40					
	F	18	2.19	0.60	11.38%	0.03	0.42	-0.03	0.48
3-Round Low-Angle HE; 1st Round Response (minutes)	M	103	1.10	0.33					
	F	18	1.20	0.30	9.34%	0.00	0.21	-0.03	0.24
3-Round Low-Angle HE; 1st Round Response Excluding Influential Point (minutes)	M	102	1.09	0.30					
	F	18	1.20	0.30	10.68%	0.01	0.22	-0.02	0.25

Metric	Integration Level	Sample Size	Mean	SD	% Difference from control	80% LCB	80% UCB	90% LCB	90% UCB
9-Round Low-Angle HE; Total Fire Mission (minutes)	M	85	4.18	0.72					
	F	19	4.55	0.90	8.85%	0.08	0.66	-0.01	0.75
9-Round Low-Angle HE; 1st Round Response (minutes)	M	90	1.07	0.24					
	F	20	1.10	0.34	3.47%	-0.07	0.14	-0.10	0.17
9-Round Low-Angle HE; 1st Round Response Excluding Influential Points (minutes)	M	90	1.07	0.24					
	F	19	1.06	0.28	-0.64%	-0.10	0.08	-0.13	0.11
3-Round Low-Angle WP; Total Fire Mission (minutes)	M	100	1.51	0.33					
	F	19	1.65	0.23	9.15%	0.06	0.22	0.03	0.24
3-Round Low-Angle WP; 1st Round Response (seconds)	M	102	43.61	15.34					
	F	19	45.68	11.59	4.76%	-1.93	6.09	-3.12	7.27
3-Round Low-Angle WP; 1st Round Response Excluding Influential Points (seconds)	M	100	42.50	13.20					
	F	19	45.68	11.59	7.49	-0.71	7.08	-1.87	8.24
3-Round Low-Angle HE from FP2; Total Fire Mission (minutes)	M	84	1.66	0.30					
	F	18	1.84	0.35	10.72%	0.06	0.29	0.03	0.33
3-Round Low-Angle HE from FP2; 1st Round Response (seconds)	M	86	56.62	13.86					
	F	19	63.00	13.63	11.28%	1.83	10.94	0.48	12.29
3-Round Low-Angle HE from FP2; 1st Round Response Excluding Influential Points (seconds)	M	85	55.96	12.54					
	F	19	60.00	13.63	12.57%	2.55	11.52	1.21	12.86

Table H E - Indirect Fire Mission Results (Descriptive Statistics by critical billet)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (MF-MM)	% Difference (FM-MM	% Difference (FF-MM)	% Difference (MF-FM)	% Difference (FF-MF)	% Difference (FF-FM)
3-Round High-Angle WP; Total Fire Mission (minutes)	MM	81	3.42	0.59	12.64%	5.26%	4.91%	7.01%	-6.86%	-0.34%
	MF	16	3.86	0.98						
	FM	22	3.6	0.76						
	FF	4	3.59	0.61						
3-Round High-Angle WP; 1st Round Response (minutes)	MM	81	1.68	0.34	6.62%	-1.01%	-2.8%	7.71%	-8.84%	-1.81%
	MF	16	1.79	0.45						
	FM	21	1.66	0.27						
	FF	4	1.63	0.13						
3-Round High-Angle WP; 1st Round Response Excluding Influential Points (minutes)	MM	80	1.66	0.31	4.26%	0.06%	-1.75%	4.2%	-5.77%	-1.81%
	MF	15	1.73	0.45						
	FM	21	1.66	0.27						
	FF	4	1.63	0.13						

Gun Nets:

Gun Nets (minutes) - The C group had a mean of 13.03 with a SD of 2.79. The LD group had a mean of 13.78 with a SD of 3.2. The HD group had a mean of 13.75 with a SD of 2.92. The LD group was 5.7% slower than the C group. The HD group was 5.54% slower than the C group. The HD group was 0.15% faster than the LD group.

Strike Nets (minutes) - The C group had a mean of 9.27 with a SD of 2.14. The LD group had a mean of 9.12 with a SD of 1.84. The HD group had a mean of 9.1 with a SD of 2.03. The LD group was 1.64% faster than the C group. The HD group was 1.82% faster than the C group. The HD group was 0.18% faster than the LD group.

Table H F - Gun Nets Results (Descriptive Statistics by integration level)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Gun Nets (minutes)	C	43	13.03	2.79	5.70%	5.54%	-0.15%
	LD	44	13.78	3.2			
	HD	52	13.75	2.92			
Strike Nets (minutes)	C	42	9.27	2.14	-1.64%	-1.82%	-0.18%
	LD	42	9.12	1.84			
	HD	52	9.1	2.03			

Displacement:

Displacement of Howitzer; Total Displacement (minutes) - The C group had a mean of 5.11 with a SD of 0.82. The LD group had a mean of 5.08 with a SD of 1.83. The HD group had a mean of 5.32 with a SD of 1.45. The LD group was 0.71% faster than the C group. The HD group was 3.94% slower than the C group. The HD group was 4.68% slower than the LD group.

Displacement of Howitzer; Trident Bar (seconds) - The C group had a mean of 19 with a SD of 8.5. The LD group had a mean of 23.71 with a SD of 15.32. The HD group had a mean of 19.67 with a SD of 8.42. The LD group was 24.81% slower than the C group. The HD group was 3.51% slower than the C group. The HD group was 17.07% faster than the LD group.

Displacement of Howitzer; Trident Bar Excluding Influential Points - The C group had a mean of 19 with a SD of 8.5. The LD group had a mean of 22.27 with a SD of 12.27. The HD group had a mean of 19.67 with a SD of 8.42. The LD group was 17.2% slower than the C group. The HD group was 3.51% slower than the C group. The HD group was 11.68% faster than the LD group.

Displacement of Howitzer; Trail Arms (seconds) - The C group had a mean of 14.02 with a SD of 5.32. The LD group had a mean of 14.72 with a SD of 5.95. The HD group had a mean of 14.4 with a SD of 5.57. The LD group was 4.97% slower than the C group. The HD group was 2.65% slower than the C group. The HD group was 2.21% faster than the LD group.

Final Displacement of Howitzer; Trident Bar (seconds) - The C group had a mean of 18.95 with a SD of 8.26. The LD group had a mean of 17.19 with a SD of 8.43. The HD group had a mean of 19.94 with a SD of 10.39. The LD group was 9.31% faster than the C group. The HD group was 5.22% slower than the C group. The HD group was 16.02% slower than the LD group.

Final Displacement of Howitzer; Trident Bar Excluding Influential Points - The C group had a mean of 18.95 with a SD of 8.26. The LD group had a mean of 17.19 with a SD of 8.43. The HD group had a mean of 19.31 with a SD of 9.47. The LD group was 9.31% faster than the C group. The HD group was 1.88% slower than the C group. The HD group was 12.34% slower than the LD group.

Final Displacement of Howitzer; Trail Arms (seconds) - The C group had a mean of 16.41 with a SD of 8.42. The LD group had a mean of 17.78 with a SD of 6.71. The HD group had a mean of 17.16 with a SD of 9.57. The LD group was 8.35% slower than the C group. The HD group was 4.55% slower than the C group. The HD group was 3.51% faster than the LD group.

Final Displacement of Howitzer; Trail Arms Excluding Influential Points - The C group had a mean of 15.53 with a SD of 6.45. The LD group had a mean of 17.78 with a SD of 6.71. The HD group had a mean of 16.3 with a SD of 7.43. The LD group was 14.52% slower than the C group. The HD group was 4.98% slower than the C group. The HD group was 8.33% faster than the LD group.

Final Displacement of Howitzer; Mount Prime Mover (minutes) - The C group had a mean of 1.05 with a SD of 0.63. The LD group had a mean of 1.08 with a SD of 0.59. The HD group had a mean of 1.02 with a SD of 0.52. The LD group was 3.27% slower than the C group. The HD group was 6.9% slower than the C group. The HD group was 3.52% slower than the LD group.

Displacement of Howitzer; Trident Bar (seconds) - The M group had a mean of 19.81 seconds with a SD of 9.52 seconds. The F group had a mean of 22.29 seconds with a SD of 10.93 seconds. The F group was 9.42% slower than the M group.

Displacement of Howitzer; Trident Bar Excluding Influential Points (seconds) - The M group had a mean of 19.81 seconds with a SD of 9.52 seconds. The F group had a mean of 22.29 seconds with a SD of 10.93 seconds. The F group was 12.51% slower than the M group.

Final Displacement of Howitzer; Trident Bar (seconds) - The M group had a mean of 18.04 seconds with a SD of 8.44 seconds. The F group had a mean of 22.8 seconds with a SD of 12.01 seconds. The F group was 26.42% slower than the M group.

Final Displacement of Howitzer; Trident Bar Excluding Influential Points (seconds) - The M group had a mean of 17.74 seconds with a SD of 7.87 seconds. The F group had a mean of 22.8 seconds with a SD of 12.01 seconds. The F group was 28.52% slower than the M group.

Displacement of Howitzer; Trail Arms (seconds) - The MM group had a mean of 14.75 seconds with a SD of 5.53 seconds. The MF group had a mean of 12.62 seconds with a SD of 4.57 seconds. The FM group had a mean of 13.67 seconds with a SD of 5.11 seconds. The FF group had a mean of 23 seconds with a SD of 15.56 seconds. The MF group was 14.43% faster than the MM group. The FM group was 7.33% faster than the MM group. The FF group was 55.96% slower than the MM group. The MF group was 7.67% faster than the FM group. The FF group was 82.26% slower than the MF group. The FF group was 68.29% slower than the FM group.

Final Displacement of Howitzer; Trail Arms (seconds) - The MM group had a mean of 17.33 seconds with a SD of 7.83 seconds. The MF group had a mean of 13.65 seconds with a SD of 4.65 seconds. The FM group had a mean of 21.62 seconds with a SD of 12.94 seconds. The FF group had a mean of 12 seconds with a SD of 2.83 seconds. The MF group was 21.24% faster than the MM group. The FM group was 24.76% slower than the MM group. The FF group was 30.77% faster than the MM group. The MF group was 36.87% faster than the FM group. The FF group was 12.1% faster than the MF group. The FF group was 44.51% faster than the FM group.

Final Displacement of Howitzer; Trail Arms Excluding Influential Points - The MM group had a mean of 16.97 seconds with a SD of 7.06 seconds. The MF group had a mean of 13.65 seconds with a SD of 4.65 seconds. The FM group had a mean of 19.07 seconds with a SD of 8.2 seconds. The FF group had a mean of 12 seconds with a SD of 2.83 seconds. The MF group was 19.53% faster than the MM group. The FM group was 12.38% slower than the MM group. The FF group was 29.27% faster than the MM group. The MF group was 28.4% faster than the FM group. The FF group was 12.1% faster than the MF group. The FF group was 37.06% faster than the FM group.

Table H G - Displacement Results by Integration Level

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Displacement of Howitzer; Total Displacement (minutes)	C	39	5.11	0.82	-0.71%	3.94%	4.68%
	LD	40	5.08	1.83			
	HD	46	5.32	1.45			

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Displacement of Howitzer; Trident Bar (seconds)	C	41	19	8.5	24.81%	3.51%	-17.07%
	LD	42	23.71	15.32			
	HD	54	19.67	8.42			
Displacement of Howitzer; Trident Bar Excluding Influential Points (seconds)	C	41	19	8.5	17.20%	3.51%	-11.68%
	LD	41	22.27	12.27			
	HD	54	19.67	8.42			
Displacement of Howitzer; Trail Arms (seconds)	C	41	14.02	5.32	4.97%	2.65%	-2.21%
	LD	43	14.72	5.95			
	HD	48	14.4	5.57			
Final Displacement of Howitzer; Trident Bar (seconds)	C	40	18.95	8.26	-9.31%	5.22%	16.02%
	LD	43	17.19	8.43			
	HD	50	19.94	10.39			
Final Displacement of Howitzer; Trident Bar Excluding Influential Points (seconds)	C	40	18.95	8.26	-9.31%	1.88%	12.34%
	LD	43	17.19	8.43			
	HD	49	19.31	9.47			
Final Displacement of Howitzer; Trail Arms (seconds)	C	39	16.41	8.42	8.35%	4.55%	-3.51%
	LD	41	17.78	6.71			
	HD	51	17.16	9.57			
Final Displacement of Howitzer; Trail Arms Excluding Influential Points (seconds)	C	38	15.53	6.45	14.52%	4.98%	-8.33%
	LD	41	17.78	6.71			
	HD	50	16.3	7.43			
Final Displacement of Howitzer; Mount Prime Mover (minutes)	C	31	1.05	0.63	3.27%	6.9%	3.52%
	LD	33	1.08	0.59			
	HD	40	1.02	0.52			

Table H H - Displacement Results (Descriptive Statistics by critical billet)

Metric	Integration Level	Sample Size	Mean	SD	% Difference from control	80% LCB	80% UCB	90% LCB	90% UCB
Displacement of Howitzer; Trident Bar (seconds)	M	113	20.37	11.18					
	F	24	22.29	10.93	9.42%	-1.30	5.14	-2.25	6.09
Displacement of Howitzer; Trident Bar Excluding Influential Points (seconds)	M	112	19.81	9.52					
	F	24	22.29	10.93	12.51%	-0.67	5.63	-1.60	6.56

Metric	Integration Level	Sample Size	Mean	SD	% Difference from control	80% LCB	80% UCB	90% LCB	90% UCB
Final Displacement of Howitzer; Trident Bar (seconds)	M	113	18.04	8.44					
	F	20	22.80	12.01	26.42%	1.07	8.46	-0.04	9.57
Final Displacement of Howitzer; Trident Bar Excluding Influential Points (seconds)	M	112	17.74	7.87					
	F	20	22.80	12.01	28.52%	1.38	8.74	0.27	9.84

Table H I - Displacement Results (Descriptive Statistics by critical billet)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (MF-MM)	% Difference (FM-MM	% Difference (FF-MM)	% Difference (MF-FM)	% Difference (FF-MF)	% Difference (FF-FM)
Displacement of Howitzer; Trail Arms (seconds)	MM	91	14.75	5.53	-14.43%	-7.33%	55.96%	-7.67%	82.26%	68.29%
	MF	21	12.62	4.57						
	FM	18	13.67	5.11						
	FF	2	23	15.56						
Final Displacement of Howitzer; Trail Arms (seconds)	MM	90	17.33	7.83	-21.24%	24.76%	-30.77%	-36.87%	-12.1%	-44.51%
	MF	23	13.65	4.65						
	FM	16	21.62	12.94						
	FF	2	12	2.83						
Final Displacement of Howitzer; Trail Arms Excluding Influential Points (seconds)	MM	89	16.97	7.06	-19.53%	12.38%	-29.27%	-28.4%	-12.1%	-37.06%
	MF	23	13.65	4.65						
	FM	15	19.07	8.2						
	FF	2	12	2.83						

Annex I.
M1A1 TANK CREWMAN (MOS 1812)

This annex details the M1A1 Tank Crewman (MOS 1812) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 3 March to 16 April 2015 at Range 500, aboard the Marine Corps Air-Ground Combat Center (MCAGCC), Twentynine Palms, CA. The sections below outline the Tank Crewman Scheme of Maneuver, Limitations, Deviations, Dataset Description, Descriptive and Basic Inferential Statistics, and Modeling Results.

I.1 Scheme of Maneuver

I.1.1 Experimental Cycle Overview

The Tank Crewmen (MOS 1812) were assessed in a field environment aboard MCAGCC, Twentynine Palms, CA. The tank experiment consisted of 3-day cycles divided into 2 pilot trial cycles and 12 record trial cycles, conducted over the course of a 45-day assessment. The experimental unit received 1 rest day after completing 4 cycles, resulting in an operational tempo of 12 days on and 1 day off. The day off was used as a range maintenance day to allow contractor support to conduct range and targetry upkeep. Sixty-four record trial cycles were planned, but 72 trials were conducted. All trials took place at the Range 500 training area aboard the MCAGCC, Twentynine Palms, CA.

Volunteers filled the Driver, Loader, and Gunner billets on each tank crew. The Tank Commander billets were filled by experienced noncommissioned officers, staff noncommissioned officers, and a lieutenant assigned through the normal orders process. The three volunteer billets were further divided into rotating and nonrotating billets. Due to safety considerations, the Gunner billets were designated as nonrotating billets and filled by the male volunteers with the greatest level of MOS experience. The Tank Commanders and Gunners who completed Tank Table VI gunnery qualifications together during the workup phase remained paired throughout the experiment. The remaining volunteers were randomly assigned to a different Driver or Loader billet each trial cycle.

I.1.2 Experimental Details

The Tank experiment trials' 3-day cycles consisted of a maintenance day, a non-live-fire day, and a live-fire day. Day 1 of the cycle (Maintenance) consisted of a standard series of preventive maintenance procedures, along with vehicle repairs conducted as required. Although not evaluated, these events ensured the unit's vehicles and equipment remained fully operational. Day 2 (Non-Live-Fire) consisted of seven individual and crew-level non-live-fire tasks. The day began with a crew-level maintenance task, followed by three crew-station-specific individual tasks, a crew evacuation, a CASEVAC, and a disabled-vehicle recovery task. Day 3 (Live Fire) consisted of six crew-level live-fire tasks conducted in two phases. Phase 1 consisted of uploading and transferring ammunition at the field ammunition-supply point. Phase 2 consisted of four live-fire engagements conducted on the Range 500 firing line. The live-fire evolution progressed from defense to offense, and included main gun and machinegun engagements. At the conclusion of the Day 3, the unit reorganized into new tank crews in preparation for the next cycle. Initial Fatigue surveys were given to each volunteer at the beginning of each experimental day, followed by a final Fatigue survey at the completion of that particular day's tasks. Workload surveys were administered immediately following certain specified tasks. Finally, volunteers completed cohesion surveys at the end of each trial cycle. For survey instruments, see the GCEITF EAP Annexes D and M.

Objective and subjective measures of performance were captured for each task. The objective measures of performance included the time required to perform the task and the heart rate of volunteers while performing the task (reported separately). Subjective measures of performance and ability included Fatigue, Workload, and Cohesion surveys conducted at various times during each trial cycle.

I.1.3 Additional Context

The assessment was conducted in a simulated tactical environment. During each cycle, the unit operated from a tactical assembly area; bivouacked on their vehicles at night, maintaining a 15% security posture; and subsisted on field rations. All trials were conducted wearing body armor and other personal protective equipment (PPE) appropriate for tank crewmen. These conditions were established to introduce the minimal level of cumulative fatigue that Marines would realistically experience in an operating environment with an enemy threat. It is important to note that sustained combat operations in an expeditionary environment would likely involve much longer workdays, less time for rest and recovery, and a corresponding increase in cumulative fatigue.

On average, the volunteers spent between 4 and 6 hours each day performing the experiment tasks. In addition, the volunteers performed various required maintenance procedures on their vehicles, including approximately 2 hours of preventive maintenance, checks, and services performed on each tank after the day's trials were complete. Equipment failures often required additional work to troubleshoot and repair disabled vehicles. Performing daily preventive maintenance and repair work maintained the tanks at the high level of operational readiness required for the experiment. Performing several hours of preventive maintenance or repairs each day is operationally realistic and an indispensable part of every tank crewman's job in the operating forces.

Fatigue surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. It allows for outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) to be accounted for in the analysis of the performance data. Workload surveys collected the volunteer's perceived average and maximal level of exertion during the performance-specified tasks. Cohesion surveys provided a method of collecting subjective data relating to each tank crew's ability to work as a team and its members' overall perspective on the cohesiveness of the crew.

I.1.4 Experimental Tasks

All of the tasks selected for this experiment were based on NAVMC 3500.121 Tank Training and Readiness (T&R) Manual; FM 3-20.21: *Heavy Brigade Combat Team Gunnery Manual;* MCWP 3-12: *Marine Corps Tank Employment;* M1A1 Technical Manual 08953A-10/1-3; and guidance from MOS senior leadership at Plans, Policies, and Operations (PP&O), Training and Education Command (TECOM), and Fleet Marine Force units. In addition, the tasks selected for the experiment were deemed physically demanding, repeatable, and frequently performed under time-constrained conditions in both training and combat environments.

A description of each task is given below. The tasks are listed in the order in which they were performed during each trial cycle.

I.1.4.1 Perform Maintenance Actions

Tank crewmen in the operating forces typically spend a significant percentage of available man-hours performing routine maintenance. The work is often very physically demanding, because the parts, tools, and equipment organic to a tank unit are large and heavy. A proficient and effective tank crewman will be capable of performing multiple hours of preventative maintenance or repairs each day. This is operationally realistic and an indispensable part of every tank crewman's job. A specific experiment task was developed to evaluate a tank crew's ability to perform common maintenance procedures. The task involved separating the track, removing and replacing a track section, and reinstalling the track. This task was chosen because it is physically demanding, repeatable, and commonly performed under time-constrained conditions in a combat environment. An initial Fatigue survey was given to each volunteer prior to this task. A Workload survey was administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

I.1.4.2 Reload Main Gun

In this task, the loader removed a round weighing 52.8 lb from a stowage compartment approximately 3 to 5 feet off the ground, flipped the round over, and loaded it into the breach while balancing inside a moving vehicle. All main gun engagements in the assessment used the 52.8-lb M831A, a training replica of the M830 High-Explosive Anti-Tank (HEAT) multipurpose round. Although this task is the loader's responsibility, all crewmembers must be able to perform the task. Doctrinal standards prescribed by FM 3-20.21 require each crewmember to demonstrate the ability to reload the main gun in 7 seconds or less. The rate of fire is limited by the speed at which the loader can manually reload the main gun, so reloading speed has a significant impact on the crew's effectiveness and survivability in combat operations. A Workload survey was administered immediately following this task to measure the average and maximum workload experienced by the volunteer.

I.1.4.3 Manually Manipulate Turret and Main Gun

In this task, each volunteer manually traversed the turret 180 degrees to the left, lowered and raised the main gun to its minimum and maximum elevation, then traversed the turret 180 degrees back to the right. The turret and main gun of the M1A1 operate by electric and hydraulic systems. As a backup, the turret and main gun can be operated using crank handles located in the gunner's station. The manual manipulation of the turret and main gun is a real-world action performed in the event of a hydraulic or electrical malfunction. The speed at which the gunner can manually scan for targets and place the main gun on target has a direct impact on combat effectiveness and survivability. A Workload survey was administered immediately following this task to capture the average and maximum workload experienced by the volunteer.

I.1.4.4 Prepare Commander's Weapons Station

In this task, volunteers lifted the M48 .50-caliber machinegun from the ground to the top of the tank, mounted the weapon in the cradle, set the headspace and timing, and performed a function check. The tank commander's station is the most physically challenging station to set up. All crewmembers must be capable of performing this procedure to be assigned as a tank commander. FM 3-20.21 requires each crewmember to demonstrate proficiency in assembling, setting headspace and timing, and performing a function check on the M48. The ability to assemble and install the weapon is necessary to prepare the tank for operations and to troubleshoot weapons malfunctions during operations. A Workload survey was administered immediately following this task to capture the average and maximum workload experienced by the volunteer.

I.1.4.5 Conduct Crew Evacuation

The crew evacuation task was performed with all crewmembers starting at crew stations with safety guards installed and hatches closed and locked. Upon receiving the command to evacuate, the crew exited their stations and moved to a rally point

50 meters behind the vehicle. Opening the crew hatches required lifting a 40- to 65-lb hinged door in an upward motion, away from the crewman. A combat-loaded M1A1 contains several hundred gallons of flammable fuel and hydraulic fluid, and several hundred pounds of HE ordinance. In the event of a fire or other emergency, the survivability of the crew depends on a quick evacuation of the tank. All crewmen (Tank Commander, Gunner, Driver, and Loader) exited the vehicle safely and moved as a crew to the rally point 25 meters away. Marines wore fighting loads and carried individual weapons, as would be expected when evacuating a vehicle in a tactical environment. Although the crew evacuation subtask is not as physically demanding as extracting a casualty from the tank, a quick vehicle evacuation is vital in the event of a vehicle fire or other life-threatening issue. A Workload survey was administered immediately following this task to capture the average and maximum workload experienced by the volunteer.

I.1.4.6 Evacuate Wounded Crewmember

This task measured a tank crew's response time and its ability to evacuate an incapacitated crewman to a rally point 50 meters away. To set up the task, a 205-lb casualty dummy (wearing full PPE) was staged in the gunner's station. The gunner's station within a tank is the most difficult location from which to evacuate a Marine because of its inaccessibility and limited workspace. To perform the CASEVAC, the crew lifted the dummy out the loader's hatch on the roof of the turret, and carried the casualty to a rally point 50 meters behind the vehicle. In the event that a crewmember is injured or otherwise incapacitated, the remaining three crewmembers are responsible for evacuating the wounded Marine. Performing this task quickly enables the casualty to receive medical attention as soon as possible. A Workload survey was administered immediately following this task to capture the average and maximum workload experienced by the volunteer.

I.1.4.7 Conduct Vehicle Recovery

To perform the recovery task, crewmembers from each tank dismounted from the vehicles, removed the tow bars from the stowage position on the operational tank, mounted them on the front of the disabled tank, and then held the tow bars in place (approximately 3 feet off the ground) while the operational tank maneuvered into position. In a combat environment, a quick recovery of a disabled vehicle denies the enemy the opportunity to maneuver or employ fires against a stationary tank. In the event that a tank becomes disabled, tank sections are equipped and trained to perform self-recovery. Recovering a disabled tank is a physically demanding task because the tow bars weighs 300 lb and must be moved by hand. A Workload survey was administered immediately following this task to capture the average and maximum workload experienced by the volunteer.

I.1.4.8 Conduct Ammunition Resupply

This task was designed to evaluate the tank crew's performance uploading a full complement of ammunition onto the vehicle from a field ammunition-supply point. A crewman removed the main gun rounds from the stowage containers, transported them approximately 20 meters to the tank, and lifted the rounds up to the crewmen standing on the tank. After each round was delivered to the tank the remaining crewmen took the round and stowed it inside the turret. A Workload survey was administered immediately following this task to capture the average and maximum workload experienced by the volunteer.

I.1.4.9 Transfer Ammunition

The M1A1 has two useable stowage compartments for main gun ammunition: the ready ammunition stowage compartment from which the rounds are withdrawn to be fired, and the semi-ready ammunition stowage compartment, which is used to carry additional rounds but is less accessible. When the ready ammunition stowage compartment is low on ammunition, the crew transfers ammunition into it from the semi-ready ammunition stowage compartment. Performance of this task was evaluated by measuring the total time required for the crew to transfer 16 main gun rounds between stowage compartments. The measure of performance for this task is time. In a combat situation, the speed with which the crew can perform this task determines how quickly the crew can resume engaging targets. A Workload survey was administered immediately following this task to capture the average and maximum workload experienced by the volunteer.

I.1.4.10 Engage Main Gun Targets – Defense

This task was the first of four live-fire engagements. The tank crew engaged a series of five main gun targets from a defensive fighting position. To calculate the elapsed time for each reload, time was recorded each time a round was fired and each time the loader completed the procedure to reload the gun. To minimize the rest time for the loader between reloads, targets were chosen at close ranges and azimuths that facilitated quick acquisition and engagement of subsequent targets.

Employing the main gun is a physically strenuous task for the loader (see section 2.2 Reload Main Gun). In contrast, the tank commander and gunner use fine motor skills and hand-eye coordination to identify targets, aim, and fire the weapons. Quick loading and reloading of the main gun has a direct effect on the crew's rate of fire and combat effectiveness.

I.1.4.11 Employ Loaders Weapons System

The loader's weapons system is a M240 medium machinegun pintle-mounted, adjacent to the loader's hatch. This task required the loader to engage a group of four man-sized targets approximately 250 m from a defensive fighting position. Three metrics were evaluated during this engagement: the percentage of engagements resulting in effects on target, the time to first effects on target, and the time required to reload the weapon. The time to first effects on target was calculated by collecting time hacks when the targets first appeared and when the target recorded the first hit. The reload time was calculated by collecting time hacks when the loader reported having expended all ammunition and when the loader resumed engaging the target. The targets were set to continuously reappear so that the loader would run out of ammunition and be forced to conduct the reloading portion of the task.

I.1.4.12 Reload Commander's Weapon Station

The third engagement took place in a defensive fighting position. The loader reloaded the commander's M48 heavy machinegun overhead with minimal exposure outside the protective armor of the tank. The task began with the tank commander engaging a target with his weapons station. A starting time hack was recorded when the weapon ran out of ammunition, and an ending time hack was recorded when the reload was complete and the tank commander resumed engaging the target. Performance evaluation was based on the total time required to conduct the reload.

I.1.4.13 Engage Main Gun Targets – Offense

In the final live-fire task, the tank crew engaged a series of four main-gun targets while in the offense, similar to the Engage Defensive Targets task, except that the tank was maneuvering downrange while engaging the targets rather than engaging targets from a stationary defensive fighting position. To calculate the elapsed time for each reload, time hacks were captured each time a round was fired and each time the loader completed the procedure to reload the gun. To minimize the rest time for the loader between reloads, targets were chosen at close ranges and azimuths that facilitated quick acquisition and engagement of subsequent targets.

I.1.5 Loading Events

Eighteen of the 19 volunteers participated in experiment trials each cycle. To provide equivalent loading for the one extra volunteer not participating in trials during that cycle, the extra Marine was assigned similar duties to perform. On maintenance and non-live-fire days, the extra Marine assisted tank crews and the unit's mechanics perform non-trial-specific preventive maintenance and repairs. On the live-fire day, the extra Marine worked at the FASP. Tasks at the FASP included downloading ammunition from the tanks between and after trials, staging ammunition for trials, and sorting dunnage. These loading tasks were physically similar to the experimental trials, involving similar levels of physical exertion, used the same basic muscle groups, and were performed for a similar duration as the experimental trials.

I.1.6 Scheme of Maneuver Summary

The 1812 experiment's 3-day cycles consisted of a maintenance day, a non-live-fire day, and a live-fire day. The maintenance day consisted of standard non-evaluated preventive maintenance and any necessary repair work.

The non-live-fire day consisted of seven individual and crew-level tasks common to mission preparation and rehearsal. The day began with a crew-level maintenance task, followed by three crew-station-specific individual tasks, a crew evacuation, a CASEVAC, and a disabled-vehicle recovery task.

The live-fire day consisted of six crew tasks conducted in two phases. Phase 1 involved uploading and transferring ammunition at the FASP. Phase 2 included a series of four live-fire engagements progressing from the defense to the offense and incorporating main gun and machinegun engagements. At the conclusion of the live-fire day, the unit reorganized into new tank crews in preparation for the next cycle.

I.2 Limitations

I.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a similar manner to those in an operational environment. However, under certain situations, artificial limitations or interruptions were introduced that changed or altered the way a task would normally be performed. While these limitations represent a degree of artificiality, they do not detract significantly from our abilities to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for the 1812 assessment.

I.2.1.1 Number of Volunteer Participants

For the Tank experiment, 17 male and 3 female volunteers began the experiment, and 16 male and 3 female volunteers completed the experiment. The results presented in this annex are based on the performance of 20 Marines. For population analysis (by gender and MOS), which examines the volunteer population and the comparative overall Marine Corps population, see Population Analysis, Annex Q.

I.2.1.2 Limitations on Cumulative Loading Effect

The members of the M1A1 Tank Crewman assessment executed a variety of subtasks over the course of a 3-day trial cycle meant to encompass, as accurately as possible, the demands placed on a Tank Crewman MOS during an administrative field training environment. The evaluated subtasks within the cycle were considered the most physically demanding and operationally relevant tasks that a junior tank crewman would perform on a recurring basis. These subtasks were executed as they would normally be performed; however, due to time constraints, data collection requirements, and other factors, limitations and artificialities existed in the assessment. These limitations had little to no effect on data collection, but impacted the cumulative loading effects and fatigue of the volunteers. The cumulative workload experienced by the volunteers during this assessment is most comparable to an administrative field training exercise such as semiannual gunnery qualifications, or the initial phases of an Integrated Tactical Exercise. The experiment was not designed to replicate the 20-hour work days, multi-day missions, or physical and mental hardships encountered during prolonged combat operations in an expeditionary environment.

I.2.1.3 Low MOS School Graduation Rate

Eight female Marine volunteers were recruited for assignment to the 1812 MOS-producing Marine Armor Crewman Course and follow-on assignment to the GCEITF tank platoon. Three of the 8 volunteers met all MOS school graduation requirements, while two more met the graduation requirements after participating in a remediation program not normally offered. Three of the recruited volunteers failed to meet the course requirements and were dis-enrolled. Of the five volunteers who graduated and received follow-on assignments to the GCEITF, two withdrew from the experiment voluntarily before the unit began workup training for the experiment. The remaining three Marines completed the entire research program. This limitation highlights that the inferences of results and conclusions to be applied to future physical, physiological, and performance standards for the 1812 MOS community are based on a smaller-than-planned sample size of females conducting the GCEITF experiment.

I.2.1.4 Omitted High-Concentration Group

The 1812 experiment included two comparison groups—a control group consisting of an all-male tank crew, and a low–female density group consisting of one female Marine integrated into an otherwise male tank crew. This assessment intended to include an additional high–female density group consisting of tank crews with two female Marines, but the low number of qualified female volunteers led the research team to eliminate this group.

I.2.1.5 Experience Disparities

Volunteers enrolled in the research program came from a variety of Marine Corps backgrounds and had varying levels of experience in the 1812 MOS. All male volunteers were recruited from active duty or reserve tank battalions, and possessed varying levels of operational experience. In contrast, female volunteers had no experience in the 1812 MOS before participating in the program. Prior to arriving at the test unit, female volunteers were sent to the 1812 MOS-producing Marine Armor Crewman Course.

The unit's pre-experiment workup at Camp Lejeune was designed to bring all volunteers to the minimum level of proficiency necessary to execute the experiment, and to provide an equivalent level of experience in performing the experiment-specific tasks. Unit workup training prior to the assessment consisted of approximately 20 weeks of crew-level training at Marine Corps Base Camp Lejeune, NC. The unit completed multiple non-live-fire field training exercises, Gunnery Skills Test qualifications for each crewman, and Tank Table VI crew gunnery qualifications. Each volunteer designated to rotate crew positions received extensive training in the Driver and Loader crew positions, including live-fire gunnery training in each position.

I.2.1.6 Change in SOPs Influenced Accurate Measurement of Ammunition Resupply Task

The task to conduct an ammunition resupply was designed to evaluate a tank crew's performance uploading a full complement of ammunition onto the vehicle from a FASP. During the experiment, the unit developed an SOP that required the gunner (a nonrotating, male volunteer) to perform the majority of the physical work. The gunner was responsible for removing the main gun rounds from the stowage containers, transporting them approximately 20 meters to the tank, and lifting the rounds up to the crewmen standing on the tank. After the gunner delivered each round to the tank, the driver and loader were responsible for taking the round and stowing it inside the turret. In addition to the gunner performing the most challenging components of the task, this method prevented the loader and driver from working continuously. Because the gunner's role in the retrieval and stowage of each round took the greatest amount of time, the elapsed time measures the gunner's performance and was not influenced by the performance of the other crewmen. For these reasons, this report cannot provide a meaningful performance comparison of all-male and integrated crews in completing the task of conducting an Ammunition Resupply.

I.2.1.7 Limitations Summary

The Tank Crewman assessment sought to replicate the realities of field conditions and loading Marines experience during field training and combat operations. The assessment team balanced these demands with the necessity of collecting equitable and uniform data throughout the assessment's duration and across participants. This led to artificialities that departed from normal operations but did not affect the validity of data collection or the assessment as a whole. Limitations were presented due to the lack of volunteers caused by low graduation rates at the MOS-producing school.

I.3 Deviations

I.3.1 Experimental Groups

The M1A1 Tank Crewman experiment was designed to include three experimental groups for comparison: a control group (all-male tank crew), a low-concentration group (one female crewmember per tank), and a high-concentration group (two female crewmembers per tank). The limited number of female participants caused the high-concentration group to be eliminated. This deviation from the EAP changed the number of total trials required to conduct the analysis.

I.3.2 Conduct Ammunition Resupply

This task was originally designed to evaluate the tank crew's performance in uploading a full complement of ammunition onto the vehicle from a field ammunition supply point. During the experiment, the unit developed a standard operating procedure that required the Gunner (a non-rotating, male volunteer) to perform the vast majority of the physical work. The Gunner was responsible for removing the main gun rounds from their storage containers, transporting them approximately 20 meters to the tank, and lifting the rounds up to the other crewmen standing on the tank. After the Gunner delivered each round to the tank, the driver and loader were responsible for taking the round and storing it inside the turret. In addition to the gunner performing the most challenging components of the task, this method also prevented the loader and driver from working continuously. Because the gunner's role in the retrieval and storage of each round took the greatest amount of time, the elapsed time primarily measures the gunner's performance and was not influenced by the performance of the other crewmen. For these reasons, this report cannot provide a meaningful comparison of the performance of all-male and integrated crews in completing this task.

I.4 Data Set Description

I.4.1 Data Set Overview

The 1812 portion of the experiment consisted of 2 pilot trial cycles and 12 record trial cycles. The pilot trial cycles were conducted from 3 March through 8 March 2015. Pilot data were not used in analysis due to variations in the conduct of the experimental trials. We base all analysis on the 12 record trial cycles executed from 9 March to 16 April 2015.

I.4.1 Record Test Volunteer Participants

At the beginning of the first record cycle, there were 17 male and 3 female volunteers. There was one Marine who voluntarily withdrew during the execution of the experiment. The final number of volunteers was 16 males and 3 females.

I.4.2 Planned, Executed, and Analyzed Trials

Table I-1 displays the number of trials planned, executed, and analyzed by task. Sixty-four record trials were planned, with the option of conducting an additional 12 trials. A total of 72 record trials were ultimately conducted. All 72 trials were analyzed with respect to the non-live-fire tasks. During two record cycles, abnormalities occurred preventing the live-fire tasks from being conducted or recorded. On one particular day, data collection error resulted in the loss of all live-fire trials for all six tank crews. On another occasion, one tank crew was unable to conduct the live-fire trials due to a drop on request (DOR) mid-trial cycle. A total of 65 trials were analyzed with respect to the live-fire tasks.

Table I-1. 1812 Planned, Conducted, and Analyzed Trials

Metric	Integration Level	Number Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Not Analyzed
Perform Maintenance Actions (minutes)	C	33	36	31	5[a]
	LD	33	36	31	5[a]
Load and Arm Main Gun (seconds)	C	66	72	64	8[a]
	LD	66	72	66	6[a]
Manipulate Turret/Main Gun (minutes)	C	66	72	66	6[a]
	LD	66	72	66	6[a]
Prep Commander's Weapon Station (minutes)	C	66	72	63	9[a]
	LD	66	72	66	6[a]
Conduct Crew Evacuation (seconds)	C	33	36	32	4[a]
	LD	33	36	33	3[a]
Evacuate Wounded Crewmen (minutes)	C	33	36	30	5[a] 1[b]
	LD	33	36	33	3[a]
Conduct Vehicle Recovery; Operational Tank (minutes)	C	33	36	32	4[a]
	LD	33	36	33	3[a]
Transfer Ammunition (minutes)	C	33	36	34	1[a], 1[b]
	LD	33	36	36	N/A
Engage Main Gun Targets – Defense (seconds)	C	165	175	160	15[a]
	LD	165	180	164	15[a], 1[b]
Employ Loader's M240; First Hit (proportion)	C	33	35	35	N/A
	LD	33	36	36	N/A
Employ Loader's M240: First Hit (seconds)	C	33	35	22	13[a]
	LD	33	36	26	10[a]
Employ Loader's M240: Reload (seconds)	C	33	35	32	3[a]
	LD	33	36	33	3[a]
Reload Commander's Weapon Station (M48) (minutes)	C	33	35	32	3[a]
	LD	33	36	32	3[a], 1[b]

Metric	Integration Level	Number Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Not Analyzed
Engage Main Gun Targets – Offense (seconds)	C	132	140	127	12[a], 1[b]
	LD	132	144	132	12[a]

[a.] Data Collection Error: data was not captured or captured incorrectly due to Data Collector (human) error or data processing (equipment) error.
[b.] Trial was identified as an outlier using box plot analysis, as described in the Methodology Annex.

I.5 Descriptive and Basic Inferential Statistics

I.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of 13 common Tank Crewman tasks and were indicative of crew-level proficiency in core MOS skills. Performance in these critical MOS tasks provides a good indication of the efficiency, survivability, and lethality of a tank crew. To further understand the descriptive statistics analysis, a brief explanation of tank crew billets and how the crews were organized for comparison of integration levels and individual performance is provided below.

This report describes the performance of various tank crew combinations when integrating MOS-qualified female Marines with MOS-qualified male Marines. It is important to quantify the billets within a tank crew, and to understand which billets the volunteers filled during the experiment. A tank crew consists of the four Marines who operate an M1A1 main battle tank. The crew positions include the Driver, the Loader, the Gunner, and the Tank Commander; these billets will be mentioned throughout this Annex. In a typical crew, the driver is the most junior and least-experienced Marine, followed by the loader. Both of these crew positions are normally filled by Marines in grades E-1 to E-3. The gunner is typically an E-4 or E-5, with a minimum of several years of experience in the MOS. The crew is led by the Tank Commander, a senior noncommissioned officer or staff noncommissioned officer who is the most experienced Marine on the crew.

The majority of the volunteers rotated between driver and loader billets by random selection without regard to MOS knowledge or experience. The gunner billets were filled by six nonrotating male volunteers with extensive MOS experience. The tank commanders were all direct-assignment male Marines, also with extensive MOS experience. The Marines assigned as gunners and tank commanders did not rotate because of the requirement for gunners and tank commanders to perform workup training and qualify together prior to conducting live fire (FM 3-20.21). For the safety of the volunteers, it was necessary to avoid assigning inexperienced crewmen to perform the gunner or tank commander’s responsibilities, and to avoid conducting trials with gunner–tank commander combinations that had not previously qualified together.

This report refers to various concentrations of vehicle crews in terms of integration levels. A control (C) group refers to an all-male crew, and is representative of the tank community as it currently exists. The low-density (LD) group refers to a gender-integrated crew consisting of three male Marines and one female Marine. If a task is performed by the entire crew, the comparison is between the performance of the C and LD groups.

Due to the individual nature of some tasks, the gender of the Marine in a critical billet matters more than the total integration level of the crew. If a task is individual in nature, the comparison is between the performance of male (M) volunteers and female (F) volunteers. If a task is performed at the crew level, but disproportionally influenced by the performance of one individual, both methods of comparison are used.

Special caution should be taken when comparing similar tasks executed by different MOSs across the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks, such as distances, techniques, leadership, load carried, group size, and group composition.

I.5.2 Selected Tasks Descriptive Statistics Results

Table I-2 and Table I-3 display descriptive statistics results for six selected 1812 tasks that were also modeled using personnel variables (see Modeling Results Section I.6). The remaining tasks are included in Appendix I M1A1 Tank Crewman (MOS 1812).

Table I-2 displays the metrics and integration levels with respect to sample sizes, means, and standard deviations. Table I-3 displays analysis of variance (ANOVA) and Welch's test results. The ANOVA table includes metrics and integration levels, p-values indicating statistical significance, integration-level elapsed-time and percent differences, and the upper and lower confidence bounds. In addition to conducting an ANOVA, a Welch's t-test was performed to examine the variances. P-values less than the significance level of 0.10 identify statistical significance. The words "metric" and "task" are used interchangeably throughout this Annex; both refer to the experimental task.

Table I-2. 1812 Selected Tasks Results

Metric	Integration Level	Sample Size	Mean	SD
Perform Maintenance Actions (minutes) *	C	31	77.92	22.07
	LD	31	86.64	22.52
Manipulate Turret/Main Gun (minutes) *	M	100	8.14	2.48
	F	32	8.97	3.08
Evacuate Wounded Crewmen (minutes)	C	30	3.97	1.33
	LD	33	4.12	1.07
Conduct Vehicle Recovery; Operational Tank (minutes)	C	32	8.26	2.14
	LD	33	8.50	2.59
Engage Main Gun Targets – Defense*	C	160	11.57	6.23

Metric	Integration Level	Sample Size	Mean	SD
(seconds)	LD	164	14.07	8.01
	M	234	12.42	7.97
	F	90	13.92	4.97
Engage Main Gun Targets – Offense * (seconds)	C	127	8.84	3.50
	LD	132	9.53	3.20
	M	187	8.99	3.46
	F	72	9.72	3.05

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

Table I-3. Analysis of Variance (ANOVA) and Welch's t-test

Metric	F statistic (df)	F Test P-value	Comparison	Diff	% Diff.	1-Sided p-Value	2-Sided p-Value	80% LCB	80% UCB	90% LCB	90% UCB
Perform Maintenance Actions (minutes)*	2.37 (1, 60)	0.13	LD - C	8.72	11.19%	0.06*	0.13	1.38	16.06	-0.74	18.18
Manipulate Turret/Main Gun*	2.41 (1, 130)	0.12	F - M	0.83	10.21%	0.09*	0.17	0.05	1.61	-0.18	1.84
Evacuate Wounded Crewmen	0.22 (1, 61)	0.64	LD - C	0.15	3.57%	0.32	0.64	-0.61	0.39	-0.75	0.53
Conduct Vehicle Recovery	0.16 (1, 63)	0.69	LD - C	0.24	2.86%	0.34	0.69	-0.53	1.00	-0.75	1.22
Engage Main Gun Targets – Defense*	9.74 (1, 322)	< 0.01*	LD - C	2.5	21.53%	< 0.01*	<0.01*	1.47	3.51	1.18	3.81
	2.79 (1, 322)	0.10*	F - M	1.5	12.11%	0.02*	0.04*	0.55	2.45	0.28	2.72
Engage Main Gun Targets – Offense*	2.73 (1, 257)	0.10*	LD - C	0.69	7.78%	0.05*	0.10*	0.15	1.22	0.00	1.38
	2.48 (1, 257)	0.12	F - M	0.73	8.15%	0.05*	0.10*	0.17	1.30	0.01	1.46

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

The conditions and performance steps for each selected task will be described, along with scatterplots depicting the results of each trial conducted for that task. In each scatterplot, the solid blue dots represent the control (C) or male group, and the red circles represent the low-density (LD) integration or female group. Outliers excluded from analysis are identified by a solid black circle placed around the data point. Contextual comments and additional insights may be presented with each experimental task to help familiarize readers with the nature and importance of each task in a combat environment.

I.5.3 Perform Maintenance Actions Overview

Tank crewmen in the operating forces typically spend a significant percentage of available man-hours performing routine maintenance. This work is often very physically demanding because the parts, tools, and equipment organic to a tank unit are large and heavy. A proficient and effective tank crewman will be capable of performing multiple hours of preventative maintenance or repairs each day. This is operationally realistic and an indispensable part of every tank crewman's job. There were two integration levels assessed for this task—a control (C) group of all male Marines, and a low-density (LD) group with one female in the tank crew.

A specific experiment task was developed to evaluate a tank crew's ability to perform common maintenance procedures. The task involved separating the track, removing and replacing a track section, and reinstalling the track. This task was chosen because it is physically demanding, repeatable, and commonly performed under time-constrained conditions in a combat environment. The time for this task started when the tank commander gave the order to begin, and finished when the task was complete with all crewmen back in their stations.

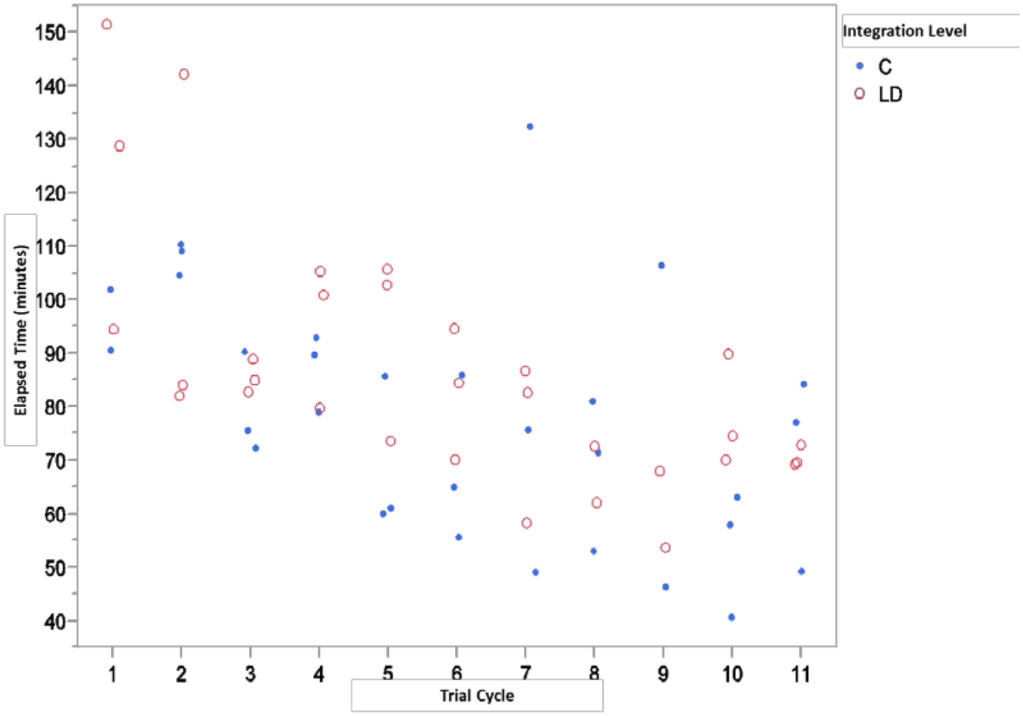
I.5.3.1 Performance Maintenance Action

For this task, sample sizes are sufficiently large ($n > 30$) to satisfy the normality assumption for ANOVA.

On average, the C crew performed 11.19% faster than a LD crew. The C group performed the task in a mean time of 77.92 minutes while the LD group had a mean time of 86.64 minutes, which results in a statistically significant p-value of 0.06.

Figure I-1 (scatterplot) displays data used in the analysis of the results.

Figure I-1. Perform Maintenance



I.5.3.2 Performance Maintenance Action Contextual Comments

Maintenance tasks similar to this one are commonly required to repair damage sustained during operations. When a tank sustains damage to its track or suspension during a mission, the crew will be required to perform repairs encompassing some or all of the performance steps required for this task. In a combat environment, it is critical to perform repairs and regain mobility quickly, denying the enemy the opportunity to attack a stationary target.

I.5.3.3 Performance Maintenance Action Additional Insights

In relation to performing preventative maintenance or repairs in a garrison environment, if we extend the result of this task out to an average maintenance workday, we can expect that an integrated crew would require approximately 10 hours to perform the work that a non-integrated crew would perform in 9 hours. In real-world operations, this difference translates into increased times required to prepare for a mission or perform maintenance after an operation.

I.5.4 Manually Manipulate Turret and Main Gun Overview

The turret and main gun of the M1A1 are operated primarily by electric and hydraulic systems. As a backup, they can be manipulated manually using crank handles located in the gunner's station. These backups provide a way for the crew to continue the fight in the event of a hydraulic or electrical malfunction. This task required each volunteer to manually traverse the turret 180 degrees to the left, raise the main gun to its maximum elevation, lower the main gun to its minimum elevation, then traverse the turret 180 degrees back to the right. This task is an individual event; other crewmembers were unable to influence the performance of the task. The data collected from this task were analyzed according to the gender of the Marine performing the task rather than the integration level of the crew. Time started when the tank commander gave the command to begin, and ended when the Marine completed the course of movement.

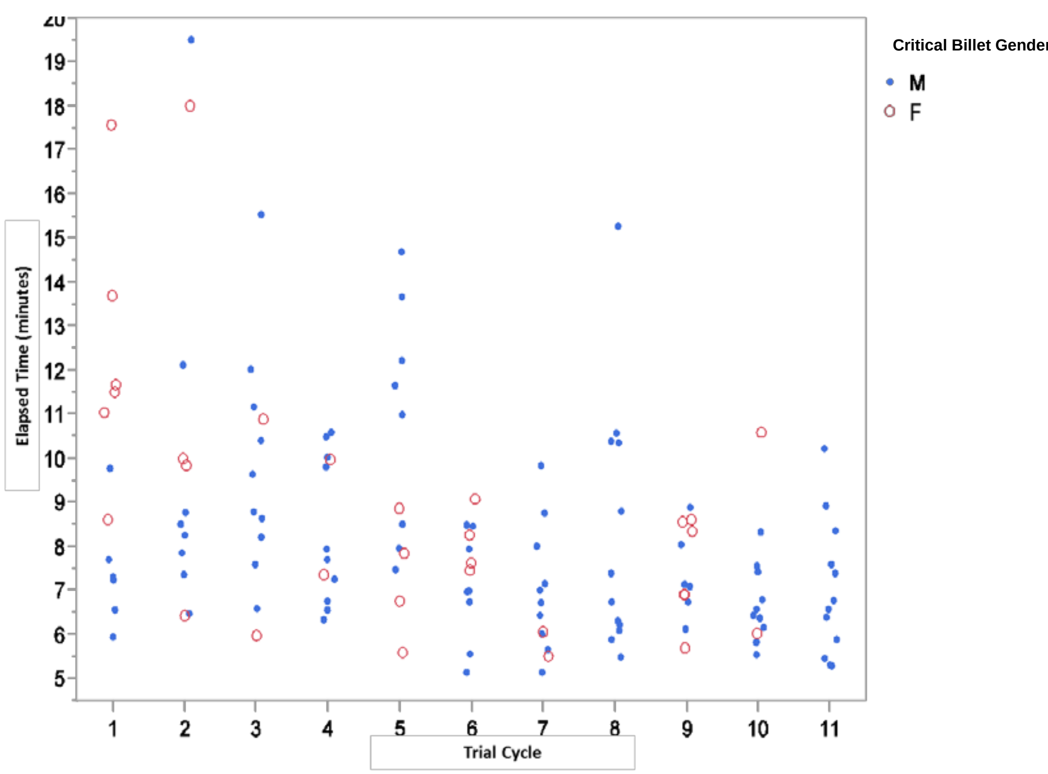
I.5.4.1 Manually Manipulate Turret and Main Gun

For this task, sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

Males (M) performed the task in a mean time of 8.14 minutes, while females (F) performed the task in a mean time of 8.97 minutes. The results of the data analysis show that the male crewman performed this task an average of 10.21% faster than the female crewman. These results are statistically significant, as identified by a p-value of 0.09. The male group performed the task an average of 49.8 seconds faster than the female group.

Figure I-2 (scatterplot) displays data used in the analysis of the results.

Figure I-2. Manually Manipulate Turret and Main Gun



I.5.4.2 Manually Manipulate Turret and Main Gun Contextual Comments

Similar to the reloading speed task, the ability to acquire and engage targets faster than an adversary provides a competitive advantage in a combat environment. The manual traverse and elevation mechanisms are backup systems that allow the M1A1 to remain operational when degraded. Mechanical failures in the primary system can be expected to occur in time. Tank crewmen are also required to manually traverse the turret as a component of several routine maintenance procedures.

I.5.5 Evacuate Wounded Crewmen Overview

This task measured each tank crew's ability to evacuate an incapacitated crewman to a simulated landing zone (LZ) 50 meters away. A 205-lb dummy served as a simulated casualty, staged in the gunner's station and equipped with full personal protective equipment. The gunner's station is the most difficult location from which to evacuate a Marine because of its inaccessibility and limited workspace. The task involved lifting the dummy out the loader's hatch and onto the roof of the turret, followed by carrying the casualty to a simulated LZ 50 meters behind the vehicle. The time for this task started when the crew was given the command to evacuate the casualty, and finished when all crewmen (including the dummy) reached the LZ.

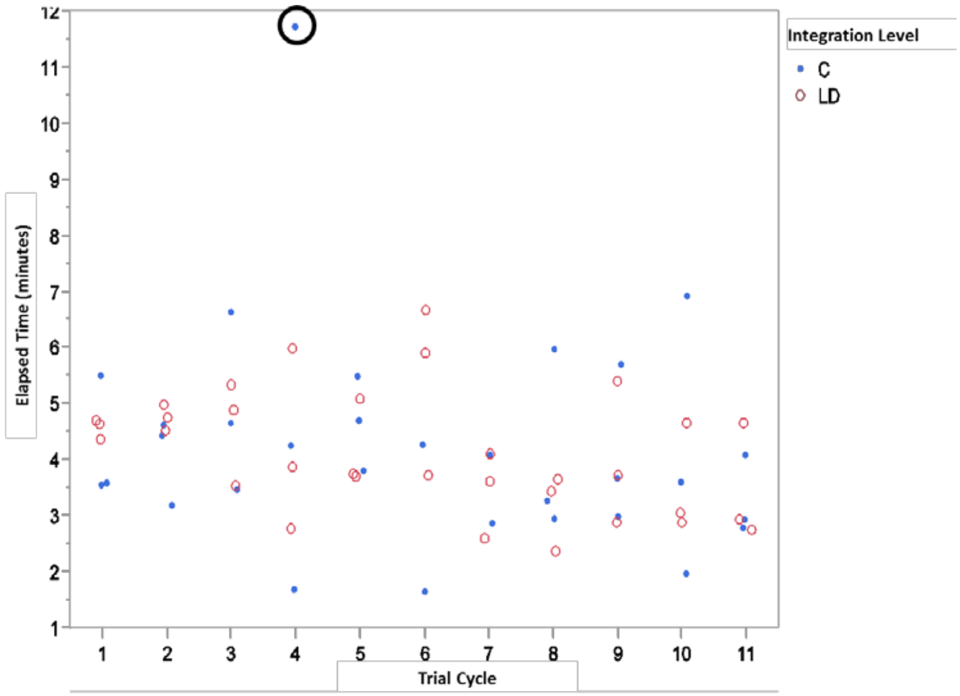
I.5.5.1 Evacuate Wounded Crewmen

For this task, sample sizes are sufficiently large ($n > 30$) to satisfy the normality assumption for ANOVA.

On average, the C group performed the task 3.57% faster than the LD group. The C group performed the task in a mean time of 3.97 minutes while the LD group had a mean time of 4.12 minutes. Based on a p-value of 0.32 in a single-sided t-test, there is no statistically significant difference between the C and LD groups.

Figure I-3 (scatterplot) displays data used in the analysis of the results.

Figure I-3. Evacuate Wounded Crewman



I.5.5.2 Evacuate Wounded Crewmen Contextual Comments

Performing a CASEVAC quickly enables the casualty to receive immediate medical attention. In situations such as a fire, it may be necessary to evacuate a casualty from a damaged tank before the sympathetic detonation of onboard fuel or high-explosive rounds create additional casualties. In this type of situation, evacuation speed directly impacts the crew's chances of survival.

I.5.6 Conduct Vehicle Recovery Overview

Tank sections are equipped and trained to perform self-recovery in the event that a tank becomes disabled. Performing tank recovery involves crewmembers from each tank dismounting from their vehicles, removing the 300-lb. tow bars from the stowage position on the operational tank, mounting the tow bars on the front of the disabled tank, and then holding the tow bars in place approximately 3 feet off the ground while the operational tank maneuvers into position to connect the tow bar's lunette eye to a pintle hitch. The task begins and ends with all crewmen in their assigned stations. The time started when the section leader gave the command to "conduct recovery" and stopped when both vehicles were rigged and ready to tow.

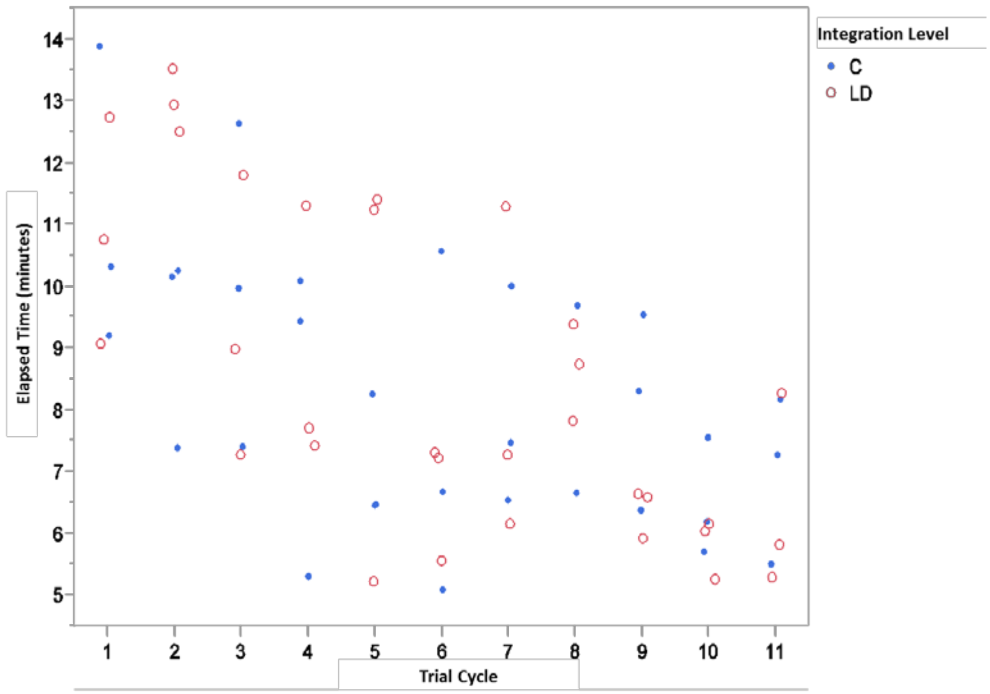
I.5.6.1 Conduct Vehicle Recovery

For this task, sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

On average, the all-male control group was 2.86% faster than the low-density integrated group. The C group performed the task in a mean time of 8.26 minutes. The LD group performed the task in a mean time of 8.50 minutes. Based on a p-value of 0.34 in a single-sided t-test, there is no statistically significant difference between the C and LD groups.

Figure I-4 (scatterplot) displays data used in the analysis of the results.

Figure I-4. Conduct Vehicle Recovery



I.5.6.2 Conduct Vehicle Recovery Contextual Comments

In a combat environment, it is desirable to recover a disabled vehicle quickly, denying the enemy the opportunity to maneuver or employ fires against a stationary tank. A time difference of approximately 30 seconds may or may not have a significant operational impact, depending on the specific circumstances of the situation.

I.5.7 Engage Main Gun Targets – Defense Overview

This task was the first of four live-fire engagements. The tank crew engaged a series of five main-gun targets from a defensive fighting position. The engagements were similar to those conducted during semiannual crew and section gunnery qualifications. Employing the main gun is a physically strenuous task for the loader (see 1.2 Reload Main Gun). Although other crewmembers were not directly involved in reloading the main gun, the performance of the driver can have a significant effect on the loader's performance. Erratic acceleration and braking by the driver while moving up and down a defensive fighting position made the loader's job more difficult to perform.

To calculate the elapsed time for each reload, time hacks were captured each time a round was fired and each time the loader completed the procedure to reload and arm the gun. To minimize the rest time for the loader between reloads, targets were chosen at close ranges and azimuths that facilitated quick acquisition and engagement of targets by the gunner.

I.5.7.1 Engage Main Gun Targets – Defense

For this task, sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

Data collected from this task were analyzed in two separate ways. The first comparison looked for performance differences between the control (C) group and the low-density (LD) group, a crew with a female in the driver or loader position. Loaders in the control group crews reloaded the main gun in an average time of 11.57 seconds, while the low-density crew had an average reload time of 14.07 seconds. The 21.53% difference between the means is statistically significant. An all-male crew can be expected to outperform a low-density integrated crew by an average 2.5 seconds on each of the five reloads.

The second comparison looked for performance differences between crews grouped only by the gender of the loader (regardless of the crew composition). Any crew with a male loader would be considered part of the control group using this method. The male loaders (M) reloaded the main gun in an average time of 12.42 seconds, while the female loaders (F) had an average reload time of 13.92 seconds. The 12.11% difference between the means is statistically significant. A male loader can be expected to outperform a female loader by an average of 1.5 seconds.

Figure I-5. Engage Main Gun Targets – Defensive

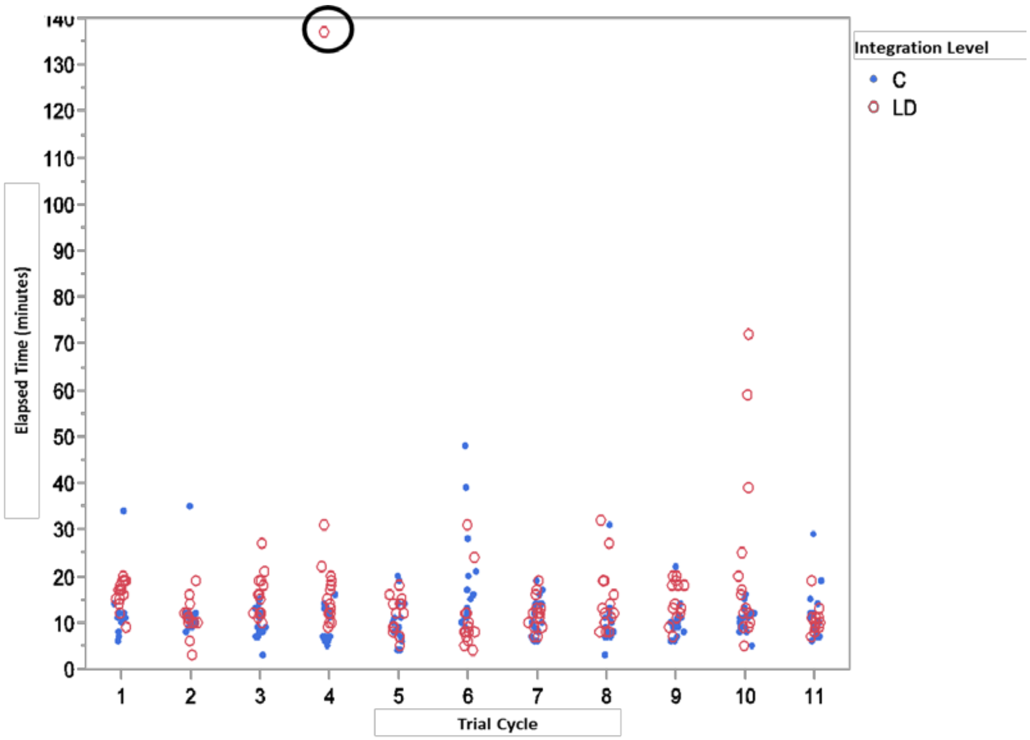
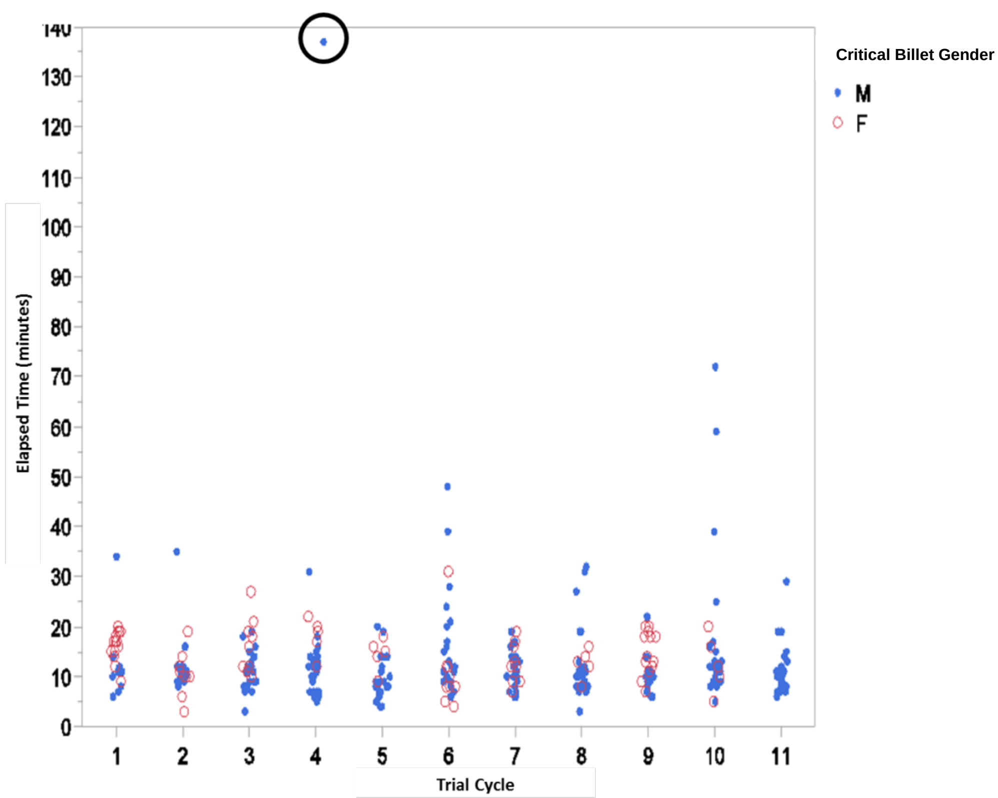


Figure I-6. Engage Main Gun Targets – Defensive



I.5.7.2 Engage Main Gun Targets – Defense Contextual Comments

In combat operations, faster reload times are highly desirable. Because the rate of fire is primarily limited by the speed at which the loader can manually reload the main gun, reloading speed has a significant impact on the crew's combat effectiveness and survivability. The ability to reload the main gun rapidly is absolutely critical when re-engaging a target or engaging multiple targets. Although this task is primarily the loader's responsibility, the potential for combat losses requires all crewmen to be capable of performing the task.

I.5.8 Engage Main Gun Targets – Offense Overview

In the fourth and final live-fire task, the tank crews engaged a series of four main-gun targets while in the offense. This task is similar to the Engage Main Gun Targets – Defense task described in section 1.10, except that the crews engaged targets while maneuvering downrange, instead of shooting from a defensive fighting position. To calculate the elapsed time for each reload, time hacks were captured each time a round was fired and each time the loader completed the procedure to reload the gun. To minimize the rest time for the loader between reloads, targets were chosen at close ranges and azimuths that facilitated quick acquisition and engagement of subsequent targets. As with the Engage Defensive Targets task, data collected from this task were analyzed by crew composition and by gender of the loader. Because the performance of the driver had a much smaller effect on the loader in the offense than in the defense, the second gender-based method (as identified in the following sections) should be considered the primary measure for analyzing the results of this task.

I.5.8.1 Engage Main Gun Targets – Offense

For this task, sample sizes are sufficiently large ($n > 30$) to satisfy the normality assumption for ANOVA.

The first comparison looked for performance differences between the control (C) group and the low-density (LD) group. Loaders in the control group crews reloaded the main gun in an average time of 8.84 seconds, while the low-density crew had an average reload time of 9.53 seconds. The 7.78% difference between the means is statistically significant, yielding a p-value of 0.05.

The second comparison looked for performance differences between crews grouped only by the gender of the loader (regardless of crew composition). Male (M) loaders reloaded the main gun in an average time of 8.99 seconds, while female (F) loaders had an average reload time of 9.72 seconds. The 8.15% difference between the means is statistically significant. A double-sided t-test produces a p-value of 0.10, and a single-sided t-test produces a p-value of 0.05.

Figure I-7 and Figure I-8 (scatterplots) depict the results of each trial conducted.

Figure I-7. Engage Offensive Targets

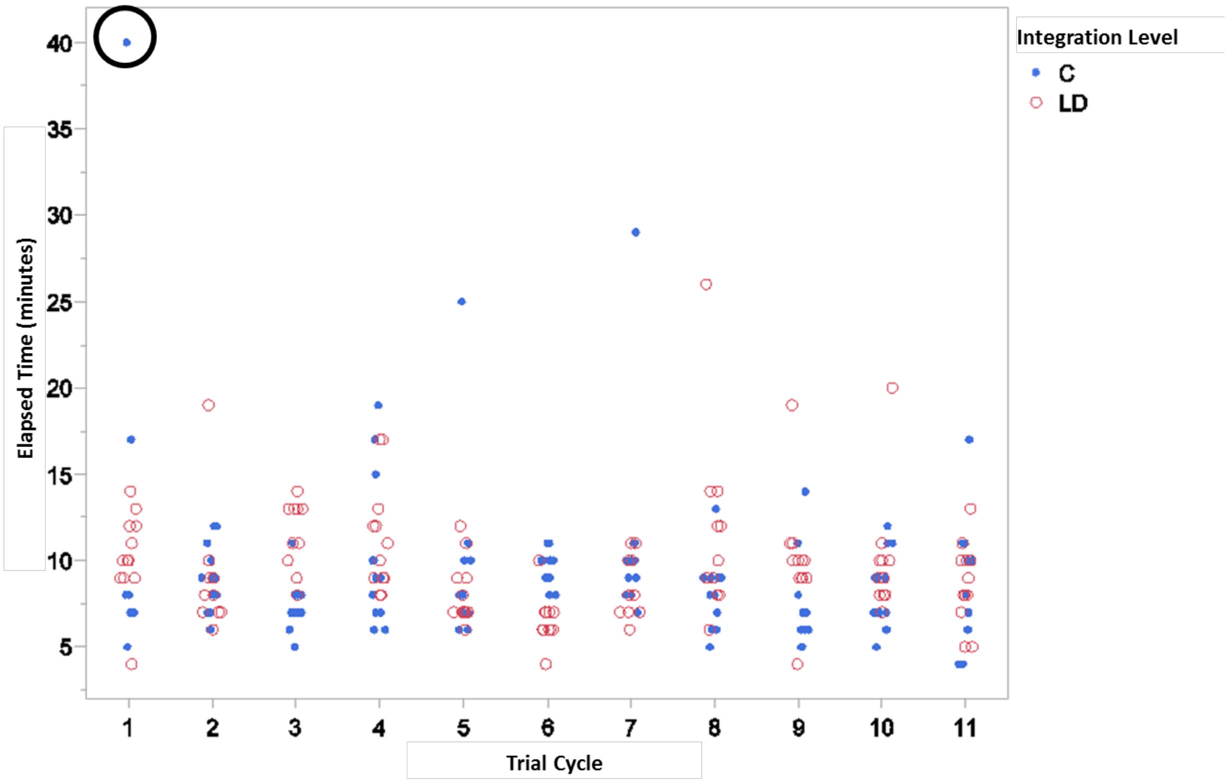
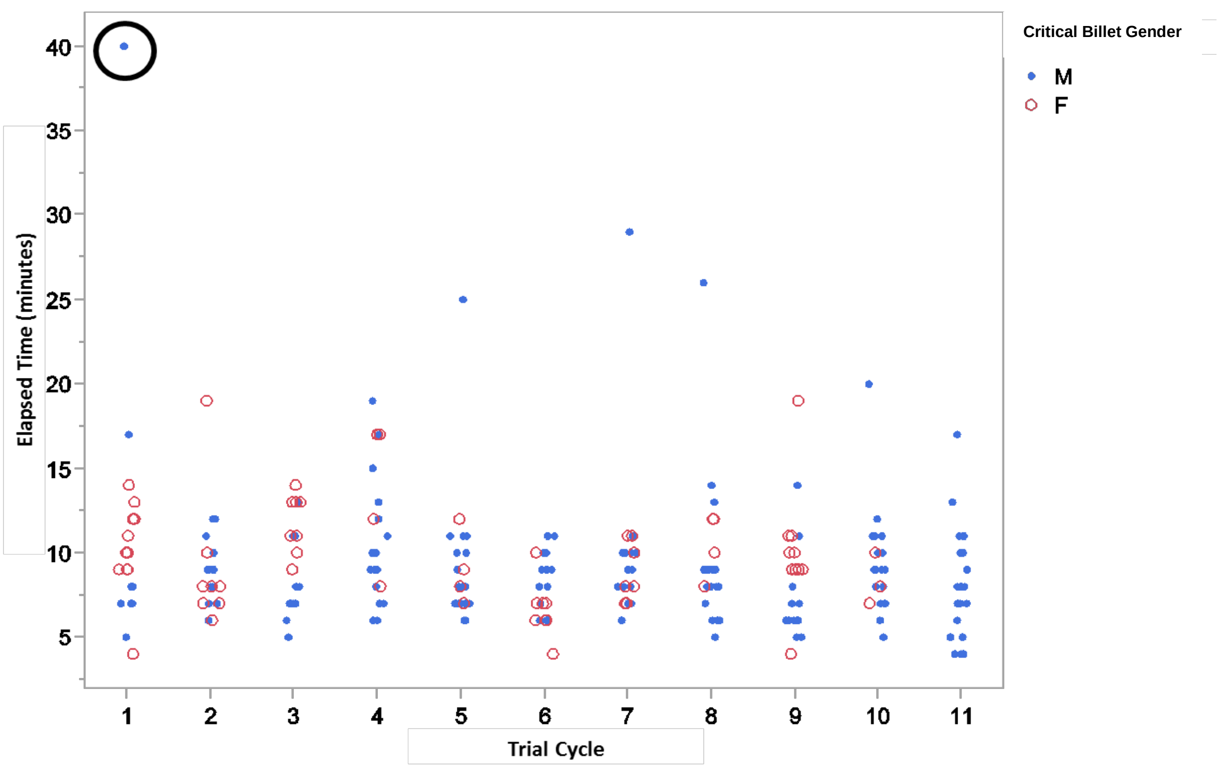


Figure I-8. Engage Offensive Targets



I.5.8.2 Engage Main Gun Targets – Offense Contextual Comments

In combat operations, faster reload times are highly desirable. Because the rate of fire is primarily limited by the speed at which the loader can manually reload the main gun, reloading speed has a significant impact on the crew's combat effectiveness and survivability. The ability to reload the main gun rapidly is absolutely critical when re-engaging a target or engaging multiple targets. Although this task is primarily the loader's responsibility, the potential for combat losses requires all crewmen to be capable of performing the task.

I.6 Statistical Modeling Results

I.6.1 Statistical Modeling Results Overview

The previous section discussed results only as they pertain to differences due to integration level alone. The goal of statistical modeling, as applied here, is to estimate simultaneously the effect of gender-integration levels and other relevant variables on tank crew performance. (Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.)

For the selected tasks described in the previous section, this section presents an overview of the analysis and results, and then presents the modeling results for each task.

For each task, we describe the significant variables in the model, and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates that the increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time.

I.6.2 Selected Tasks Method of Analysis

A mixed-effects model with all crew members and all types of personnel data does not work for the Tank dataset. Thus we model each personnel variable with integration level separately with a random effect for who filled each position in the Tank crew. For example, age for each member of the Tank crew (three variables), a random-effect for who filled each billet, and integration level are modeled with the result (response time) as the response variable. For tasks that had critical billets (i.e., only one or two Marines were performing the task without help from others), only their personnel data and gender were included in the model. Where maximum likelihood estimation converged, AIC was used for variable selection. Otherwise, we comment on the significance of individual variables in the full model. Variables reported as significant are concluded to be significant based on at least a one-sided test.

I.6.3 Selected Tasks Overall Modeling Results

There are no personnel data variables that are both statistically significant and have a practical impact in the model. Each time personnel data variables are statistically significant in a model, their effects are practically negligible, conflicting, and/or incomplete for the tank crew (i.e., there are no tasks for which a variable is significant for all or even most members of the tank crew).

Integration level consistently appears as statistically significant in each task, and its effect is clear, causal, and practical; integration level is the best variable to describe performance for each task. (Refer to section I.5.2 Descriptive Statistics Results for the ANOVA summary for each of the below-mentioned tasks.)

I.6.3.1 Perform Maintenance Actions

We modeled elapsed time for maintenance actions as a function of each personnel variable for each crewmember and integration level in a mixed-effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the tank crew. We report statistically significant positive and negative correlations, and whether we observe any patterns.

The crew composition is significant and positively correlated with performing maintenance actions, indicating that integrated crews perform this task slower than all-male crews.

The models for the following variables do not run due to missing values:

- None.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Age
- AFQT
- GCT
- Height
- Weight
- CFT MTC
- CFT MANUF
- PFT Run Score
- Rifle Score.

The LD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with performing maintenance actions:

- PFT Crunch score of the Driver.

The following personnel variables are significant in their respective models and are negatively correlated with performing maintenance actions:

- GCT Score of the Loader
- CFT MANUF of the Gunner.

The following personnel variables have no significant variables in their respective models:

- None.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes integration level only, where LD has a difference of 10.33 minutes when compared to a C group. The comparison is statistically significant. This difference is an increase from the 8.72-minute difference identified in the descriptive statistics, which is an 18.46% change.

I.6.3.2 Manually Manipulate Turret and Main Gun; Driver

We modeled elapsed time for manually manipulating the turret and main gun by the critical billet (Driver) as a function of the personnel variables and gender of the Marine performing the task in a mixed-effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the tank crew. We report statistically significant positive and negative correlations.

The gender of the Driver was not significant and ultimately not selected into the final model for elapsed time for manually manipulating the turret, indicating that Driver gender is not a good predictor of time to execute this task.

The models for the following variables do not run due to missing values:

- None.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The LD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following Driver variables are significant in the model and are positively correlated with manually manipulating the turret:

- PFT Crunch Score.

The following Driver variables are significant in the model and are negatively correlated with manually manipulating the turret:

- Height.

There are no patterns for any Driver variables for manually manipulating the turret and main gun.

Because integration level is not significant in the final model, there is no final mixed effects model for this task. (Refer to the section I.5.2 for this task to see the ANOVA results.)

I.6.3.3 Manually Manipulate Turret and Main Gun; Loader

We modeled elapsed time for manually manipulating the turret and main gun by the critical billet (Loader) as a function of the personnel variables and gender of the Marine performing the task in a mixed-effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the tank crew. We report statistically significant positive and negative correlations.

The gender (male) Loader is significant and negatively correlated with manually manipulating the turret, indicating that males perform this task faster than females.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The LD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following Loader variables are significant in the model and are positively correlated with manually manipulating the turret:

- None.

The following Loader variables are significant in the model and are negatively correlated with manually manipulating the turret:

- Gender
- Age.

Because the effects of the personnel variables do not have any patterns and their effects are often negligible, our final model includes critical billet gender only, where females have a difference of 1.98 minutes when compared to males. The comparison is statistically significant.

I.6.3.4 Evacuate Wounded Crewman

We modeled elapsed time for evacuating a wounded crewman as a function of each personnel variable for each crew member and integration level in a mixed-effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the tank crew. We report statistically significant positive and negative correlations, and whether we observe any patterns.

Gender integration levels were not selected by the AIC into the final model of elapsed time to evacuate a wounded crewman, indicating that integration level is not a good predictor of time to evacuate a wounded crewman.

The models for the following variables do not run due to missing values:

- None.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The LD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with evacuating a wounded crewman:

- PFT Crunch Score of the Driver.

The following personnel variables are significant in their respective models and are negatively correlated with evacuating a wounded crewman:

- Rifle Score of the Gunner.

The following personnel variables have no significant variables in their respective models:

- Age
- AFQT
- GCT
- Height
- Weight
- CFT MTC
- CFT MANUF
- PFT Run.

There are no patterns for any personnel variables for evacuating a wounded crewman.

Because integration level is not significant in the final model and there are no variables that are significant for the crew, there is no final mixed-effects model for this task. (Refer to the section I.5.2 for this task to see the ANOVA results.)

I.6.3.5 Conduct Vehicle Recovery

We modeled elapsed time for conducting a vehicle recovery as a function of each personnel variable for each crewmember and integration level in a mixed-effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the tank crew. We report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not run due to missing values:

- None.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- None.

The LD integration level is significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with conducting a vehicle recovery:

- None.

The following personnel variables are significant in their respective models and are negatively correlated with conducting a vehicle recovery:

- CFT MTC of the Loader
- CFT MANUF of the Loader
- PFT Run score of the Loader.

The following personnel variables have no significant variables in their respective models:

- Age
- AFQT
- GCT
- Height
- Weight
- PFT Crunch
- Rifle Score.

There are no patterns for any personnel variables for conducting a vehicle recovery.

Because integration level is not significant in the final model and there are no variables that are significant for the crew, there is no final mixed-effects model for this task. (Refer to the section I.5.2 for this task to see the ANOVA results.)

I.6.3.6 Engage Main Gun Targets – Defense

We modeled elapsed time for a defensive engagement with the main gun as a function of each personnel variable for each crew member and integration level in a mixed-effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the tank crew. We report statistically significant positive and negative correlations, and whether we observe any patterns.

The gender of the Loader was not significant and ultimately not selected into the final model for elapsed time for a defensive engagement with the main gun, indicating that the Loader gender is not a good predictor of time to execute this task.

The models for the following variables do not run due to missing values:

- None.

The following Loader variables are significant in the model and are positively correlated with defensive engagements with the main gun:

- GCT Score
- CFT MTC.

The following Loader variables are significant in the model and are negatively correlated with defensive engagements with the main gun:

- Height
- CFT MANUF
- PFT Crunch Score
- PFT Run Score.

There are no patterns for any personnel variables for defensively engaging main gun targets. (Refer to the section I.5.2 for this task to see the ANOVA results.)

I.6.3.7 Engage Main Gun Targets – Offense

We modeled elapsed time for an offensive engagement with the main gun as a function of each personnel variable for each crew member and integration level in a mixed-effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the tank crew. We report statistically significant positive and negative correlations, and whether we observe any patterns.

The gender of the Loader was not significant and ultimately not selected into the final model for elapsed time for an offensive engagement with the main gun, indicating that the Loader gender is not a good predictor of time to execute this task.

The following Loader variables are significant in the model and are positively correlated with offensive engagements with the main gun:

- GCT Score
- Weight.

The following Loader variables are significant in the model and are negatively correlated with offensive engagements with the main gun:

- AFQT Score
- Height
- CFT MANUF
- PFT Crunch Score

- PFT Run Score.

There are no patterns for any personnel variables for offensively engaging main gun targets. (Refer to the section I.5.2 for this task to see the ANOVA results.)

Appendix to Annex I
1812 Supplemental Information

This appendix provides supplemental information for the 1812 portion of the GCEITF experiment. It provides additional descriptive and basic inferential statistics not described in Annex I.

Section 1: Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences.

This section presents results for additional 1812 tasks. Annex I contains the descriptive statistics for the remainder of the 1812 tasks. The words “metric” and “task” are used interchangeably throughout this Appendix. They both refer to the experimental task. Table I A displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels. Table I B displays ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare the groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the LD group is different from that in the C group.

Table I A - 1812 Additional Tasks Results

Task and Metric	Integration Level	Sample Size	Mean	SD	% Difference	
Load and Arm Main Gun (seconds)	M	97	6.45	1.39	F – M	15.51%
	F	33	7.45	1.97		
Prep Commander’s Weapon Station (minutes)	M	97	4.59	1.48	F – M	11.56%
	F	32	5.12	2.07		
Conduct Crew Evacuation (seconds)	C	32	37.03	5.49	LD - C	-2.38%
	LD	33	36.15	4.51		
Transfer Ammunition (minutes)	M	53	8.04	2.60	F – M	14.80%
	F	18	9.23	2.74		
Employ Loader’s M240; First Hit (proportion)	M	35	0.78	N/A	F – M	-8.33%
	F	13	0.72	N/A		
Employ Loader’s M240: First Hit (seconds)	M	35	20.74	19.75	F – M	67.99%
	F	13	34.85	28.42		
Employ Loader’s M240: Reload (seconds)	M	47	45.43	16.40	F – M	-13.17%
	F	18	39.44	15.11		
Reload Commander’s Weapon Station	M	46	1.30	0.57	F – M	21.60%

Task and Metric	Integration Level	Sample Size	Mean	SD	% Difference	
(M48) (minutes)	F	18	1.58	0.92		

Table I B - Analysis of Variance and Welch's T-test

Task and Metric	F statistic (df)	F Test P-value	Comparison	Difference	% Diff.	T-Test P-Value one-sided (two-sided)	80% LCB	80% UCB	90% LCB	90% UCB
Load and Arm Main Gun	10.18 (1, 128)	< 0.01*	F - M	1.00	15.51%	< 0.01* (0.01*)	0.52	1.48	0.38	1.62
Prep Commander's Weapon Station	0.00 (1, 127)	0.98	F - M	0.53	11.56%	0.49 (0.98)	0.02	1.04	-0.13	1.20
Conduct Crew Evacuation	0.50 (1, 63)	0.48	LD - C	-0.88	-2.38%	0.24 (0.48)	-2.50	0.74	-2.97	1.21
Transfer Ammunition	2.73 (1, 69)	0.10	F - M	1.19	14.80	0.06* (0.12)	0.22	2.16	-0.07	2.45
Employ Loader's M240: First Hit†	3.78 (1, 46)	0.06*	F - M	14.11	67.99%	0.09* (0.17) †	0.00†	26.00†	-1.00†	33.00†
Employ Loader's M240: Reload	1.80 (1, 63)	0.18	F - M	-5.99	-13.17	0.09* (0.17)	-11.59	-0.37	-13.24	1.28
Reload Commander's Weapon Station (M2)	2.17 (1, 62)	0.15	F - M	0.28	21.60%	0.12 (0.24)	-0.03	0.59	-0.12	0.68

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level according to ANOVA or an equivalent nonparametric test.

†Due to lack of normality, p-values and confidence intervals have been replaced with Mann-Whitney Test results.

Section 2: Additional Task Results:

Load and Arm Main Gun. An essential skill for all tank crewmen is the ability to rapidly reload the tank's main gun. During this task, the Tank Loader was required to retrieve a 53 pound dummy round from a storage compartment approximately three to five feet off the ground, flip the round over, and loading it into the breach on a stationary vehicle. The dummy rounds used in this task were inert rubber and steel training replicas of the 120mm M830 High-Explosive Anti-Tank (HEAT) multipurpose round. The dummy rounds are identical in size, shape, weight, and weight distribution to service ammunition rounds. Due to the individual nature of this task the method of comparison looked at performance differences between crews grouped by the gender of the Marine performing the task. Data collectors began recording time when the loader received the command to reload and stopped when the round was loaded and the main gun armed.

For this task, group sample sizes are sufficiently large ($n \geq 30$) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. Male loaders performed the task in a mean time of 6.45 seconds compared to female loaders with a mean time of 7.45 seconds. On average, the male loaders executed the task 15.51% faster than the female loaders. A two-sided t-test yields a p-value of 0.01, meaning the performance difference between genders is statistically significant. Males performed the task an average of 1 second faster than females.

- **Contextual Comments**. Combat operations faster reload times (even measured in fractions of a second) are highly desirable. Because the rate of fire is primarily limited by the speed at which the loader can manually reload the main gun, reloading speed has a significant impact on the crew's combat effectiveness and survivability. Although this task is primarily the loader's responsibility, the potential for combat losses requires all crewmen to be capable of performing the task. Doctrinal standards prescribed by FM 3-20.21, the Heavy Brigade Combat Team Gunnery Manual, require each crewmember to demonstrate the ability to reload the main gun in 7.00 seconds or less on a stationary vehicle. On average, the female loaders did not meet the standard established by the reference. Numerous near-peer competitor platforms utilize autoloading systems that offer ever increasing reload speeds. With the service life of the M1A1 projected to extend to 2050, selecting tank crewmen who are able to quickly reload the main gun plays a significant role in maintaining the competitiveness of the platform.

Prepare Commander's Weapon Station. The most physically challenging crew station to set up is the tank commander's weapon station. This task requires the crewman to lift the M48 .50 caliber machine gun from the ground up to the top of the tank, mount the weapon in the cradle, set the headspace and timing, and perform a functions check on the weapon. All crewmembers must be capable of performing this procedure to be assigned as a tank commander. Additionally, FM 3-20.21 requires each crewmember to demonstrate proficiency in assembling, setting headspace and timing, and performing a functions check on the M48. The time starts when the crewman performing the task is given the command to begin and stops when the installation is complete. This task is an individual event; the other crewmembers are unable to influence the performance of the task. For this reason, the data collected from this task was analyzed according to the gender of the Marine performing the task rather than according the to the integration level of the crew.

For this task, group sample sizes are sufficiently large ($n \geq 30$) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. The Males (M) performed the task in a mean time of 4.59 minutes, while females (F) had a mean time of 5.12 minutes. On average males executed the task 11.56% faster than females, with a statistically significant p-value of 0.09. The male group performed the task an average of 31.8 seconds faster than the female group.

- **Contextual Comments**. This task, as scripted for the experiment, is unlikely to be performed while in contact with the enemy because it is by nature a preparatory task. However, various components of this task could be required to troubleshoot and correct a malfunction that occurs while employing the weapon in a combat situation. Additionally, removing/installing and disassembling/assembling the weapon for cleaning and inspection are required components of before and after operations maintenance procedures.

Crew Evacuation. The crew evacuation task required the entire crew to rapidly exit their stations and move to a rally point 50 meters behind the vehicle. It was performed with all crewmembers starting in their crew stations with safety guards installed and hatches closed and locked. The physically demanding aspects of this task include lifting a 40 to 65 pound crew compartment hatch up and away from the crewmen on a hinge, climbing down off the tank, and sprinting 50 meters. The time starts when the Tank Commander gives the command to evacuate and stops when the last crewman reaches the rally point.

For this task, group sample sizes are sufficiently large ($n \geq 30$) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. On average, the C group performed the task 2.38% slower than an integrated crew. The C group performed the task in a mean time of 37.03 seconds, while the LD group had a mean time of 36.15 seconds. No statistical significant difference is present between the C and LD groups.

- **Contextual Comments**. A combat loaded M1A1 contains over five hundred gallons of highly flammable fuel and hydraulic fluid, and over two thousand pounds of explosive ordinance. In the event of a fire or similar emergency, the survivability of the crew is dependent on their ability to evacuate the tank quickly.

Conduct Ammunition Resupply Overview. This task was originally designed to evaluate the tank crew's performance in uploading a full complement of ammunition onto the vehicle from a field ammunition supply point. Unfortunately, the way that the task was performed did not present a meaningful opportunity to determine whether a performance difference exists between the C and LD groups.

The standard operating procedure developed by the test unit required the Gunner (a non-rotating, male volunteer) to perform the vast majority of the physical work. The Gunner was responsible for removing the main gun rounds from their storage containers, carrying them approximately 20 meters to the tank, and lifting the rounds up to the other crewmen onboard the tank. After the Gunner delivered each round to the tank, the driver and loader were responsible for storing it inside the turret. In addition to the gunner performing the most challenging components of the task, this method also prevented the loader and driver from working continuously. Because the gunner's role in the retrieval and storage of each round took the greatest amount of time, the elapsed time primarily measures the gunner's performance and was not significantly influenced by the performance of the other crewmen. For these reasons, this report cannot provide a meaningful comparison of the performance of all-male and integrated crews in completing this task.

Transfer Ammunition. The M1A1 has two useable storage compartments for main gun ammunition: the ready ammunition storage compartment from which the rounds are withdrawn to be fired, and the semi-ready ammunition storage compartment which stores additional rounds but is much less accessible. When the ready ammunition storage compartment is low on ammunition the crew transfers more ammunition into it from the semi-ready ammunition storage compartment. This task evaluated the crew's ability to perform this procedure by measuring the total time required for the crew to transfer 16 main gun rounds between storage compartments. The time started when the tank commander gave the order to begin and ended when all 16 rounds had been transferred and both storage compartment doors were secured. All of the main gun engagements in the assessment utilized the 52.8 pound M831A, a training replica of the M830 High-Explosive Anti-Tank (HEAT) multipurpose round. This task was performed entirely by the loader and tank commander without assistance from the other two crewmen. All the tank commanders in the experiment were direct assignment male Marines. For this reason, the data collected from this task was analyzed according to the gender of the Marine serving as the loader.

For this task, the M group data are not normally distributed as evidenced by a Shapiro-Wilk Test p-value of less than 0.01, while the F group data are normally distributed with Shapiro-Wilk p-value of 0.04. We recommending using the ANOVA results presented above which are confirmed by a Mann-Whitney Test (p-value = 0.10). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

Male loaders (M) performed the task in a mean time of 8.04 minutes while female loaders (F) performed it in a mean time of 9.23 minutes. On average, a male loader is 14.80% faster than a female loader in performing this task. Male loaders performed this task an average of one minute and eleven seconds faster than female loaders.

- **Contextual Comments**. In a combat situation, a tank crew is required to perform this task after expending approximately half of their stored ammunition. The amount of time required to transfer ammunition determines how quickly the crew can resume engaging targets. Additionally, this task provides insight into similar tasks ammunition handling tasks that are common to the 1812 MOS. Uploading ammunition onto the tank and redistributing ammunition between tanks are common tasks which place similar physical demands on the tank crewmen. These tasks may, or may not, be performed under time critical conditions. Whether the performance difference between groups is operationally significant is entirely dependent on the specific tactical situation.

Employ Loader's M240. This was the second live-fire engagement. The loader's weapons system was a M240 medium machine-gun pintle-mounted adjacent to the loader's hatch. The Loader's M240 is primarily used for close defense of the tank. This task required the loader to engage a group of four man-sized targets at a range of 250 meters, pausing to reload the weapon during the engagement. The targets were set to continuously reappear so that the loader would run out of ammunition and be forced to conduct the reloading portion of the task. Three metrics were evaluated during this engagement: the number of trials in which the loader was able to achieve effects on target, the time to first effects on target (if effects occurred), and the time required to reload the weapon. The time to first effects on target was calculated by collecting time hacks when the targets first appeared and when a target recorded the first hit. The reload time was based on time hacks recorded when the loader ran out of ammunition and again when the loader resumed engaging the target.

The task was performed as an individual event; the loader's performance on this task was in no way dependent on other crewmembers. For this reason, performance comparisons are made based on the gender of the loader rather than by crew integration level. In assessing effects on target, males (M) achieved a hit 78% of the time, while females (F) hit the target 72% of the time.

For time to first hit, the M group data are not normally distributed as evidenced by a Shapiro-Wilk Test p-value of less than 0.01, while the F group data are normally distributed with Shapiro-Wilk p-value of 0.02. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA, however, due to lack of normality, we recommend using the results of the Mann-Whitney Test presented in Table I-B.

Regarding the length of time required hitting the target (with the sample size only including those trials where a hit did occur), male loaders (M) hit the target in a mean time of 20.74 seconds, while female loaders (F) hit the target in a mean time of 34.85 seconds. On average, males were 67.99% faster than females. A one-sided t-test for this task yields a p-value of 0.06. A male loader can be expected to achieve effects on target an average of 14.11 seconds faster than a female loader.

For time to reload, the M group data are not normally distributed as evidenced by a Shapiro-Wilk Test p-value of less than 0.01, while the F group data are normally distributed with Shapiro-Wilk p-value of 0.26. We recommending using the ANOVA results presented above which are confirmed by a Mann-Whitney Test (p-value = 0.30). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

In assessing the third component of this task, reloading the M240, male loaders (M) performed the task in a mean time of 45.43 seconds while female loaders (F) had a mean time of 39.44 seconds. On average, females were 13.17%, or 5.99 seconds, faster than males in reloading the M240. A one-sided t-test for this task yields a p-value of 0.09, and a two-sided t-test yields a p-value of 0.17. A female loader can be expected to reload the M240 an average of 5.99 seconds faster than a male loader.

- **Contextual Comments.** The loader's M240 is primarily used to provide close range defense of the tank against lightly armored or dismounted threats such as an anti-tank missile team or a vehicle-borne improvised explosive device. In this context, it is extremely important for the loader to quickly achieve effects on target.

  Additionally, it is important to note that the loader's M240 on a M1A1 takes a great deal of physical strength to employ accurately. Unlike the common M240B variant, the loader's M240 has duel spade grips and a butterfly trigger instead of a buttstock and pistol grip. It is also pintle-mounted on a skate ring with no bipods for support. While these modifications greatly increase the quick maneuverability of the weapon and enable the loader to cover a 180 degree sector of fire, they also necessitate the application of a significant amount of force to keep the weapon on target while firing. A loader unable to apply sufficient force to the weapon during employment will have difficulty hitting the target.

Reload Commander's Weapon Station. The third engagement took place from a stationary position and required the loader to reload the commander's M48 heavy machine gun overhead with minimal exposure outside the protective armor of the tank. Performance evaluation was based on the total time required to conduct the reload. The task began when the tank commander ran out of ammunition while engaging a target. The loader removed the empty ammunition can, locked a full ammunition can in place, and fed the rounds into the weapon. A starting time hack was recorded when the tank commander gave the order to "reload" and an ending time hack was recorded when the tank commander resumed engaging the target. The task was performed as an individual event; the loader's performance on this task was in no way dependent on other crewmember's performance. For this reason, performance comparison was made by gender rather than by crew integration level.

For this task, both the M and F group data are not normally distributed as evidenced by Shapiro-Wilk Test p-values of less than 0.01. We recommending using the ANOVA results presented above which are confirmed by a Mann-Whitney Test (p-value = 0.62). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

Male loaders (M) performed the task in a mean time of 1.30 minutes while female loaders (F) performed the task in a mean time of 1.58 minutes. On average, male loaders were 21.60% faster than female loaders. Due to the high degree of variance in the data produced by each group, the results cannot be considered statistically significant.

- **Contextual Comments**. Although the commander weapon station is fired from inside the crew compartment, the weapon itself is mounted on top of the tank without any armor protection. In order to reload the weapon, the Loader must lean out the loaders hatch exposing part of the body (normally the head, shoulders, and arms) outside the protective armor of the tank. Because this is inherently dangerous in a combat environment, the challenge is to perform the task while keeping as much of the body inside the tank as possible. This means lifting the 30-lb ammunition can overhead to feed the rounds and lock it in place.

Annex J.
AAV Crewman (MOS 1833)

This annex details the Assault Amphibious Vehicle (AAV) Crewman (MOS 1833) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 3 March – 26 April 2015 at Range 500, aboard the Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, CA. The sections outline the AAV Crewman Scheme of Maneuver, Limitations, Deviations, Data Set Description, Descriptive and Basic Inferential Statistics, and Modeling Results.

J.1 Scheme of Maneuver

J.1.1 Experimental Cycle Overview

The AAV portion of the GCEITF Experiment was executed from 3 March to 26 April 2015 at Range 500, aboard the MCAGCC, Twentynine Palms, CA. The second portion of the experiment for AAVs took place from 9 May to 18 May 2015 at Assault Craft Unit FIVE (ACU-5) and Training Area Red Beach aboard CAMPEN, CA. The AAV experiment was originally scheduled to commence record trials on 9 March 2015 and continue until 1 May 2015, with makeup days through 11 May 2015 at Range 500. The amphibious record trials were scheduled to commence on 18 May 2015 and continue until 1 June 2015, with makeup days inclusive in those dates. This timeline included 2 pilot trial cycles, 16 record trial cycles, and 3 makeup trial cycles. The pilot trials at 29 Palms were conducted as planned from 3-6 March 2015, while 16 record trial cycles were conducted from 9 March through 26 April 2015, vice 1 May 2015 as originally planned. A range maintenance day was conducted every four cycles (12 days) to allow contractor support to conduct range upkeep. This also allowed volunteers to receive one recovery day at Camp Wilson. The CAMPEN experimental trials were conducted from 9-18 May 2015. Each evaluated subtask in both phases was considered the most physically demanding and operationally relevant tasks that an 1833 AAV crewman is expected to perform. An 1833 AAV crew is made up of three Marines: a Rear Crewman, Driver, and Turret Gunner.

J.1.2 Experimental Details

The GCEITF volunteers of the AAV experiment executed a variety of subtasks over the course of a 3-day trial cycle at the MCAGCC in Twentynine Palms, CA and a 1-day cycle at Camp Pendleton (CAMPEN), CA. The first day of the 3-day cycle at the MCAGCC consisted of live-fire evaluated subtasks divided into three categories: a modified gun table V, combat-simulated interior reload, and a combat-simulated exterior reload. The second day of the trial cycle consisted of non-evaluated maintenance tasks meant to keep the vehicles and equipment combat ready executing subsequent trial cycles. The third day consisted of non-live-fire evaluated subtasks divided into three categories: conducting maintenance actions, loading weapons and ammunition, and conducting CASEVACs. The 1-day cycle subtasks at CAMPEN were divided into two categories: CASEVACs conducted in an amphibious environment, and water and land towing operations. There were no breaks in the record trials at CAMPEN. Initial Fatigue surveys were given to each volunteer at the beginning of each experimental day, followed by a final Fatigue survey at the completion of that particular day's tasks. Workload surveys were administered at the completion of a particular day's tasks. Lastly, volunteers completed cohesion surveys at the end of each trial cycle.

J.1.3 Additional Context

Marines wore a standard vehicle fighting load while conducti.ng all subtasks. This included a CVC protective suit, an IPC with SAPI plates, a Kevlar or CVC helmet, and an M4 Carbine. This gear weighed approximately 35 lb. The Marines bivouacked at a platoon position in the vicinity of the AAVs. All volunteers not randomly selected to participate on a particular day's trials conducted tasks outside the assessed events to provide realistic loading. These tasks were equivalency tasks to ensure equity between individuals participating in a trial cycle and those not chosen via random selection. These tasks are discussed in detail in the loading section below.

Baseline and Post-Trial Fatigue surveys were administered to the volunteers before the beginning of the first trial and after the end of the last trial. These surveys were designed to capture the volunteers' cumulative fatigue level at the beginning and end of each day's trials. The data collected provide additional insight into apparent aberrations in the performance level of a given volunteer on a specific day. It allows for outside fatigue-related factors (minor illness, lack of sleep the night before, etc.) to be accounted for in the analysis of the performance data. Additionally, all volunteers completed Workload surveys at the completion of each day's trials. These surveys collected the volunteers' perceived average and maximal level of exertion during the performance of the day's trials. Lastly, all volunteers completed Crew Cohesion Surveys at the conclusion of each cycle. These surveys provided a method of collecting subjective data relating to each AAV crew's ability to work as a team.

J.1.4 Scheme of Maneuver Experimental Tasks

J.1.4.1 Conduct Modified Gun Table V and Simulated Reloads

The AAV live-fire trial day consisted of three separate subtasks: conduct a modified gun table V, conduct a combat-simulated interior reload, and conduct a combat-simulated exterior reload. Prior to the start of the subtasks, each vehicle crewman completed a baseline fatigue survey. All subtasks were conducted as a crew event, with an AAVP7A1 crew of a Driver, Rear Crewman, and Turret Gunner.

J.1.4.1.1 Conduct a modified Gun Table V

The AAV crews were required to engage multiple targets ranging from 400 to 1500 m from their positions, distances and exposure times dictated by the US Army Heavy Brigade Combat Team Gunnery Manual. The Marines wore fighting loads and conducted four separate engagements: one offensive and three defensive, and engaged 12 vehicle or troop targets.

J.1.4.1.2 Conduct Interior Reload

This event required the AAV Turret Gunner and Rear Crewman to conduct a reload of 200 rounds of .50-caliber ammunition and 64 rounds of 40-mm grenade rounds. The Rear Crewman assisted the Turret Gunner by moving boxes of ammunition from the appropriate troop-compartment stowage spaces and handing them to the Turret Gunner, who sat in the vehicle turret. This task was executed twice during the course of fire and simulated the requirement for the Rear Crewman to assist the Turret Gunner with the reloading of both weapons systems within the turret of the AAV.

J.1.4.1.3 Conduct External Reload

This event required the AAV Turret Gunner and Rear Crewman to conduct a reload of 200 rounds of .50-caliber ammunition and 96 rounds of 40-mm grenade rounds. A loaded can of .50-caliber ammunition weighs approximately 35 lb. and a loaded can of 40-mm grenade rounds weighs approximately 45 lb. The Rear Crewman assisted the Turret Gunner by moving boxes of ammunition from the appropriate troop-compartment stowage spaces to the top of the vehicle and into the turret from the exterior. This task was executed once during the course of fire and simulated the requirement for the Rear Crewman to assist the Turret Gunner with the reloading of both weapons systems from the exterior of the AAV.

J.1.4.2 Conduct Maintenance Actions

The AAV non-live-fire trial day consisted of three separate subtasks: conduct maintenance actions, load weapons and ammunition, and conduct simulated CASEVACs. Prior to the start of the subtasks, each vehicle crewman completed a baseline fatigue survey. All subtasks were conducted as a crew event, with an AAVP7A1 crew of a Driver, Rear Crewman, and Turret Gunner.

J.1.4.2.1 Break and Reassemble Track

This subtask required the AAV crew to use the vehicle SL-3 and to break track at the rear of the vehicle. The crews then moved the vehicle forward until the broken track was at the mid-point of the AAV at the dual support roller. The vehicle was then placed in reverse and moved back until the crew was able put the track back together. Upon completion of the setup, an AAV Platoon Staff member inspected the vehicle to ensure proper reassembly and the crew had followed all necessary safety requirements. Upon verification, the crew stowed all materials used and moved the vehicle approximately 20 meters to a marked point, signaling completion of the subtask. The crew had 90 minutes to accomplish this task. This task simulated multiple possible suspension and vehicle track issues that could arise during the course of operations. An initial Fatigue survey was given to each volunteer prior to the beginning of this task.

J.1.4.2.2 Conduct a Manual Ramp Raise

This subtask required the AAV crew to raise the AAV ramp, manually, with a ramp jack. The crews did not have responsibilities dictated for this task and were allowed to conduct the task as they saw fit until completion. Upon completion, an AAV Platoon Staff member inspected the ramp to ensure it had been properly secured. There was no maximum time limit for this task.

J.1.4.3 Load Weapons and Ammunition

The vehicle crew loaded the M2 .50-caliber machine gun, the MK-19 40-mm grenade launcher, and a full complement of ammunition (two cans of .50-caliber ammunition and three cans of 40-mm grenades) from the lowered ramp of the AAV. The crew was required to execute this task from the exterior of the vehicle. An interior loading of the fully assembled M2 was authorized. A fully assembled M2 weighs approximately 85 lbs. and the MK-19 weighs approximately 79 lb. Upon completion, an AAV Platoon Staff member inspected to ensure correct installation of the weapons. There was no maximum time limit for this task. This task simulated the AAV crew loading weapons and ammunition from various positions on the vehicle.

J.1.4.3.1 Conduct Manual Turret Manipulation

This subtask required the AAV Turret Gunner, without assistance, to manually traverse, elevate, and depress the vehicle turret. The Turret Gunner used the manual traverse and elevation controls within the turret. The Turret Gunner manually maximized elevate, maximized depress, and then traversed the turret 360 degrees. A data collector captured the total elapsed time; there was no maximum time limit. This task simulated the requirement of a Turret Gunner to manipulate the turret, manually, scanning and engaging targets.

J.1.4.3.2 Conduct CASEVAC of an Incapacitated Turret Gunner

This subtask required the Rear Crewman and Driver to evacuate a simulated casualty from the turret of the AAV, externally and internally. The simulated casualty was a dummy wearing the appropriate vehicle crewman fighting load weighing approximately 205 lb. The Rear Crewman performed each evacuation alone for the first 60 seconds while the Driver radioed a simulated CASEVAC 9-line report. If the dummy was not evacuated by the 60-second mark, the Driver assisted. The Driver and Rear Crewman then attempted to execute the evacuation until the 6-minute mark. The Turret Gunner then assisted between the 6-minute mark and the time limit of 9 minutes. Once the dummy was placed on top of the vehicle for the exterior evacuation or on the floor of the troop compartment for the interior evacuation, the data collector recorded a stop time. The data collector was responsible for verifying the dummy had not sustained additional head injuries during the evacuation using an accelerometer check light fastened to the head of the dummy. The three choices signaled by the check light were green, yellow, and red. A red check light meant the casualty sustained significant head injury during the casualty evacuation, a yellow light signaled additional head injury but not life threatening, and a green light signaled no injury. This task simulated a crew conducting a CASEVAC from the turret and is a training and readiness requirement for 1833s. A final fatigue survey was conducted at the end of this non-live fire day task. A workload survey was also administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

J.1.5 Conduct Amphibious Operations

The AAV Amphibious trials conducted aboard CAMPEN, CA consisted of 1- day cycle consisting of three separate subtasks: conduct a secure and unsecure of an AAVP7A1, conduct an evacuation of an incapacitated Rear Crewman at sea, and perform recovery operations. All subtasks were conducted as a crew event, with an AAVP7A1 crew of a Driver, Rear Crewman, and Turret Gunner.

J.1.5.1 Conduct a Secure and Unsecure of an AAVP7A1

This subtask required the AAV Crew to secure and unsecure an AAV to a Landing Craft, Air Cushioned (LCAC). The crew secured the AAV with eight sets of chains and dogging brackets, then unsecured the AAV and returned the chains and dogging brackets to their appropriate stowage areas. A data collector captured separate start and stop times for secure and unsecure. There was no maximum time associated with this task. This task simulated an AAV crew securing and unsecuring a vehicle in preparation to debark from Navy Amphibious shipping. An initial Fatigue survey was given to each volunteer prior to the beginning of this task.

J.1.5.2 Conduct Water Evacuation of an Incapacitated Rear Crewman

This subtask required the AAV Turret Gunner and Driver to lift a simulated casualty from the floor of the troop compartment to the top of the vehicle through the starboard right cargo hatch. The casualty was simulated by a 205-lb dummy. The Turret Gunner and

Driver had 25 minutes to accomplish this task. This task is a training and readiness requirement for 1833s and simulated conducting a CASEVAC from the troop compartment of the vehicle.

J.1.5.3 Perform Recovery Operations

This subtask required the AAV crew to conduct a water and land recovery of a simulated disabled AAV. The crews of the recovery and disabled AAVs worked together for each portion of the recoveries. The water recovery required the Turret Gunner and the Rear Crewman from each vehicle to throw water-tow ropes to the recovery vehicle or use a boat hook to secure the thrown ropes. The crews then connected both vehicles stern-to-stern and towed the disabled vehicle in the water. The data collector recorded a stop time upon commencement of the tow. The land tow required both crews to manipulate the two vehicles and the land tow bar weighing approximately 150 lbs. and attach it to two points on the disabled vehicle. The operational vehicle then backed up until the tow pintle seated properly. Once the vehicle was attached, both crews loaded into the recovery vehicle and the recovery vehicle began to tow. The data collector captured the stop time when the recovery vehicle had towed the disabled vehicle 125 ft. up the beach to a designated area. Upon completion of each tow, the crews swapped roles and executed the recovery again. There was no maximum time limit for this subtask. The water and land tow tasks are requirements outlined by the training and readiness manual and are somewhat likely to occur during normal operations on land and in the water. A final fatigue survey was conducted at the end of this day's task. A workload survey was also administered immediately following this task to measure the average and maximum workload experienced by each volunteer.

J.1.6 Loading Events

Loading events ensured, to the greatest extent possible, equity of physical activity amongst all volunteers and attempted to simulate the intense physical and mental workload present in combat or in a combat-focused training environment. Loading activities included nightly security and observation posts, maintenance actions, and execution of all subtasks. The first enduring task was turret watch, held by all Marines scheduled to participate in the following days' trials and consisting of 2 hours of watch. This watch only occurred on nights prior to assessing events. Maintenance day, also known as PMCS, included maintaining winch cables and resetting any gear stowed incorrectly during evaluations. Major maintenance actions were performed during the maintenance day. At the end of the established 12-day cycle, crews conducted lube orders on their vehicle prior to movement back to Camp Wilson. The day spent at Camp Wilson was also a range maintenance day where contractors performed repairs or maintenance on all targets. Weekly maintenance checks and services were conducted every other PMCS day.

J.1.7 Loading Challenges

There were multiple challenges enacting a loading plan that effectively simulated the physical and mental difficulties of combat or a combat-training environment. The foremost problem was the mindset of the volunteers. When executing subtasks during loading trials, Marines moved slower or exerted less effort knowing their trial would not be a matter of formal data collection efforts. Concerns for troop welfare and morale support led to decisions that allowed Marines to bivouac in a covered administrative position every maintenance day during a trial rather than maintain a tactical posture on their vehicles over the course of multiple trials. This administrative posture alleviated the cumulative effect that fatigue has on Marines in a tactical environment.

J.1.8 Scheme of Maneuver Summary

The AAV Experiment consisted of a 3-day trial cycle consisting of live-fire, maintenance, and non-live-fire days in Twentynine Palms and a 1-day trial cycle at CAMPEN. During the course of the experiment, the AAV platoon executed 2 pilot trial cycles and 16 record trial cycles at the MCAGCC, and 1 pilot trial cycle and 9 record trial cycles at CAMPEN. During trial execution, Marines wore a standard vehicle fighting load, which weighed approximately 35 lbs., while conducting all subtasks. When not participating in a trial, Marines executed all subtasks to ensure equivalency loading amongst volunteers. Knowing their efforts were not being directly observed and recorded did not ensure equivalency of effort by participants. Additional physical loading included watch rotations, vehicle maintenance, and a vehicle bivouac site. A rotation into an administrative posture every maintenance day alleviated the cumulative effect that fatigue has on Marines in a tactical environment.

J.2 Limitations

J.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a similar manner to those in an operational environment; however, artificial limitations or interruptions were introduced that changed or altered the normal performance of a task. While these limitations represent a degree of artificiality, they do not detract significantly from the data collection plan. The following limitations were observed for the 1833 AAV Crewman assessment.

J.2.2 Number of Volunteer Participant

For the AAV experiment, 17 male and 10 female volunteers began the trials, but by the end 16 males and 10 females were able to complete the assessment. The results presented in this annex are based on the performance of 27 Marines.

J.2.3 Experience Inequalities

Volunteers enrolled in the program came from a variety of Marine Corps backgrounds and had varying levels of experience in the 1833 MOS. The unit's pre-experiment workup was designed to bring all volunteers to the minimum level of proficiency necessary to execute the experiment, and to provide an equivalent level of experience in performing the experiment-specific tasks.

Unit workup training prior to the assessment consisted of approximately 20 weeks of crew-level training at Marine Corps Base Camp Lejeune, NC. The unit completed multiple non-live-fire field training exercises, Gunnery Skills Test qualifications for each crewman, and AAV Table I-V crew gunnery. Each volunteer designated to rotate crew positions received extensive training in all crew positions.

J.3 Deviations

Deviations to the execution of the AAV Crewman scheme of maneuver were made; however, no deviations that occurred affected the analysis methodology outlined in the EAP.

J.4 Data Set Description

J.4.1 Data Set Overview

The 1833 portion of the experiment at Range 500 consisted of 2 pilot trial cycles and 16 record trial cycles. The pilot trial cycle was conducted from 3-8 March 2015. Pilot trial data were not used in analysis due to variations in the conduct of the experiment. All analysis was based on the 16 record trial cycles executed from 9 March 2015 to 28 April 2015. The 1833 Amphibious Operations were conducted at Camp Del Mar aboard Camp Pendleton, Ca from 9-18 May 2015. One pilot trial day was conducted on 9 May 2015. Record trials were executed from 10-18 May 2015.

J.4.2 Record Test Volunteer Participants

At the beginning of the first record trial, there were 17 male volunteers and 10 female volunteers. There was one male DOR during the execution of the experiment. The final number of volunteers was 16 males and 10 females.

J.4.3 Planned, Executed and Analyzed Trial Cycles

Table J-1 below shows the number of trials planned, executed and analyzed by task for the AAV experiment for Twentynine Palms and Camp Pendleton evolutions. The notes column provides insight on the delta between the trials analyzed versus trials conducted. The data collection process allowed for identification of potential outliers and invalid data points due to data collector or Toughbook program error, or equipment malfunction. For 29 Palms, the AAV experiment began with 8 trials on any given day, but changed after the first trial cycle to include 9 trials. This increased the number of executed trials for many of the tasks. The decision to increase the daily number of trials was made based on the possible loss of data due to forecasted wind delays. Live Fire Engagements data was lost because of wind delays that eventually cancelled the trials for specific days. The increase of trials from 8 to 9 proved beneficial and maintained the goal of reaching the planned number of trials.

Table J-1. 1833 Planned, Executed and Analyzed Trial Cycles[1]

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
Conduct a modified Gun Table V	C	74	74	72	Data Collection Errors (2)
	LD	76	84	84	
	HD	90	88	87	Data Collection Errors (1)
Conduct Interior Reload (#1 and #2)	C	74	74	74	
	LD	76	84	84	
	HD	90	74	74	
Conduct External Reload	C	37	39	37	Lost data points due to Data Collection Processing
	LD	38	42	42	
	HD	45	46	45	Lost data point due to Data Collection Processing
Break and Reassemble Track	C	37	38	36	Data Collection Errors (2)
	LD	38	42	42	
	HD	45	46	46	
Conduct a Manual Ramp Raise	C	37	39	38	Lost data point due to Data Collection Processing
	LD	38	42	42	
	HD	45	46	46	
Load Weapons and Ammo	C	72	75	75	
	LD	76	84	83	Data Collection Errors (1)
	HD	90	92	91	Data Collection Errors (1)
Conduct Manual Turret Manipulation	C	37	38	38	
	LD	38	42	42	
	HD	45	46	45	Lost data point due to Data Collection Processing
Conduct	C	74	75	74	Data Collection Error (1)

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
CASEVAC of an Incapacitated Turret Gunner (Internal and External)	LD	76	82	82	
	HD	90	90	89	Data Collection Error (1)
Conduct Land Recovery	C	25	22	22	
	LD	25	23	22	Mechanical Failure (1)
	HD	30	27	27	
Conduct a Secure and Unsecure of an AAVP7A1	C	50	44	44	
	LD	50	46	46	
	HD	60	54	54	
Conduct Water Casevac	C	25	22	21	Data Collection Error (1)
	LD	25	25	23	Data Collection Processing
	HD	30	27	19	Data Collection Error (8)
Conduct Water Recovery	C	25	25	22	Lost data points due to Data Collection Processing
	LD	25	24	23	Data Collection Error (1)
	HD	30	27	27	

1. Some data was not captured or captured incorrectly due to human (Data Collector - DC) error or data processing equipment (Toughbook - TB) error. Wind delays also caused a loss of data on some occasions. Some data points were classified as outliers or potential influential points and were excluded from the analysis as described in the methodology section.

J.5 Descriptive and Basic Inferential Statistics

J.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common AAV Crewman tasks and are indicative of unit-level proficiency during either field exercises or combat operations. Seven selected tasks are presented in this section. The Appendix to this Annex contains the descriptive statistics for the remainder of the tasks.

This section accounts for various AAV Crew combinations when integrating MOS-qualified female Marines with MOS-qualified male Marines. An AAV Crew consists of a three-Marine complement per vehicle. The crew positions include Rear Crewman, Driver, and Turret Gunner. In a typical crew, the Rear Crewman is the most junior and least MOS-experienced Marine. The Driver is the next senior and more MOS-experienced Marine, while the Turret Gunner, also called the Crew Chief, is the most senior and MOS-experienced Marine.

This report will refer to various concentrations of vehicle crews in terms of integration levels. A control group (C) refers to an all-male crew filling any combination of the three billets; this type of crew was the control for the experiment. The low-density (LD) crew was made up of two males and one female filling any combination of the three billets. The high-density (HD) crew included one male and two females. The Marines rotated through the billets via random selection, without respect to MOS knowledge or experience, in an effort to examine the effects of integration at all crew levels. In addition to integration level comparisons, there were certain tasks also analyzed by critical billet. A critical billet was identified for tasks where a physically demanding duty was primarily performed by one individual.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA), Tukey Tests (or non-parametric equivalent as necessary), and scatter plots. The first table titled Descriptive Statistics displays the metric, integration levels, sample sizes, means and standard deviations. The second table shows ANOVA test results including, but not limited to, metrics, p-values suggesting statistical significance, integration level elapsed time differences and percentage differences between integration levels. If non-parametric tests were needed, the second table displays these results instead of ANOVA and Tukey test results. Subsequent subsections will cover each task in detail along with scatterplots of the data. If p-values are less than the a-priori significance level of 0.10, we conclude that there is statistical evidence that the mean elapsed time for the experimental groups, LD and HD, are different from the C group.

Special caution should be taken when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks such as distances, techniques, leadership, load carried, group size, and group composition. Lastly, the words “metric” and “task” are used interchangeably throughout this Annex. They both refer to the experimental task.

J.5.2 Conduct Interior Reload Overview

The AAV Turret Gunner and Rear Crewman reloaded 200 rounds of .50-caliber ammunition and 64 rounds of 40-mm grenade rounds. The Rear Crewman moved boxes of ammunition from the appropriate troop compartment stowage spaces and handed them to the Turret Gunner, who sat in the vehicle turret. This task was executed twice during the course of fire, simulating the requirement for the Rear Crewman to assist the Turret Gunner reload both weapons systems within the turret of the AAV.

The time began when the Rear Crewman received the command to “reload” from the Master Gunner and stopped when the Turret Gunner had completed the interior reload and was ready to re-engage targets. The Data Collector was inside the troop compartment and recorded start and stop times triggered by the commands from the Master Gunner and notifications from the AAV Turret Gunner via vehicle intercom.

The interior reload occurred twice during the modified Gun Table V course of fire derived from the HBCT Gunnery Manual. The data collected for this task was analyzed in two ways. The first comparison looked at performance differences by integration level, while the second comparison looked for performance differences by the gender of the Marines in two critical billets (Rear Crewman and Turret Gunner).

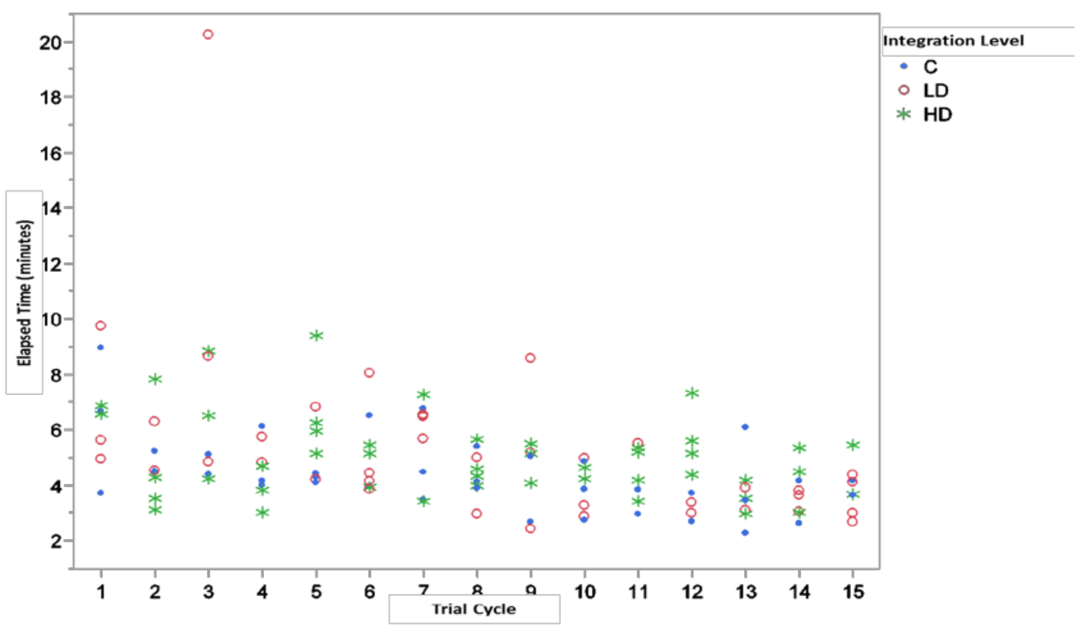
J.5.2.1 Conduct Interior Reload #1 by Integration Level

Conducting interior reload is a physically demanding task: the Rear Crewman is moving two boxes of .50-caliber ammunition weighing approximately 35 pounds each and two boxes of 40-mm ammunition weighing approximately 45 pounds each. The task was performed as a crew event, with the AAV Rear Crewman lifting the ammunition individually, with the primary measure of performance being the time required to complete the task.

J.5.2.1.1 Conduct Interior Reload #1 by Integration Level Scatterplots

The scatter plots below display results for completing the Interior Reload # 1 task by integration level. No outliers were identified for this sub-section.

Figure J-1. Interior Reload #1 by Integration Level



J.5.2.1.2 Conduct Interior Reload #1 by Integration Level Data Table and Analysis

The tables below summarize the results of the task, Interior Reload #1 by Integration Level. The first table compares sample sizes and means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-2. Interior Reload #1 by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Interior Reload #1 (minutes)	C	37	4.43	1.37
	LD	42	5.32	2.97
	HD	46	5.02	1.50

Table J-3. Interior Reload #1 by Integration Level ANOVA and Welch's Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	1- Sided P-Value	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Interior Reload #1 (minutes)	2.46 (2, 76.5)	0.09††	LD-C	0.89	20.02%	0.05	0.09	0.07	1.70	-0.09	1.86
			HD-C	0.59	13.37%	0.03*	0.06	-0.20	1.39	-0.36	1.55
			HD-LD	-0.29	-5.54%	0.29	0.57	-1.06	0.47	-1.22	0.63

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances)

For this task, samples sizes are sufficiently large (n>30) to satisfy normality assumptions for ANOVA. Since the standard deviation of the LD group is more than twice that of the C group, we recommend using the robust ANOVA results presented in the table above.

Table J-4 shows the control group averaged faster times than the LD and HD groups.

Table J-3 shows the LD group was 20.02% slower than the C group; this difference is not statistically significant. The HD group was 13.37% slower than the C group; this difference is not statistically significant. The HD group was 5.54% faster than the LD group; this difference is not statistically significant.

J.5.2.1.3 Conduct Interior Reload #1 by Integration Level Contextual Comments

In combat operations, faster reload times (even measured in a few seconds) are highly desirable. Faster reloading increases the combat effectiveness and survivability of an AAV crew. Per the MCWP 3-13, "The AA unit should continue to engage the enemy as fast as possible, because fire placed in the enemy's area will less his effectiveness and give friendly weapons time to adjust. A wasted opportunity to engage a target may never be regained." With the service life of the AAVP7 projected to extend to 2035, the selection of AAV Rear Crewmen who are able to quickly move boxes of ammunition and assist the Turret Gunner effectively plays a significant role in maintaining the competitiveness of the platform.

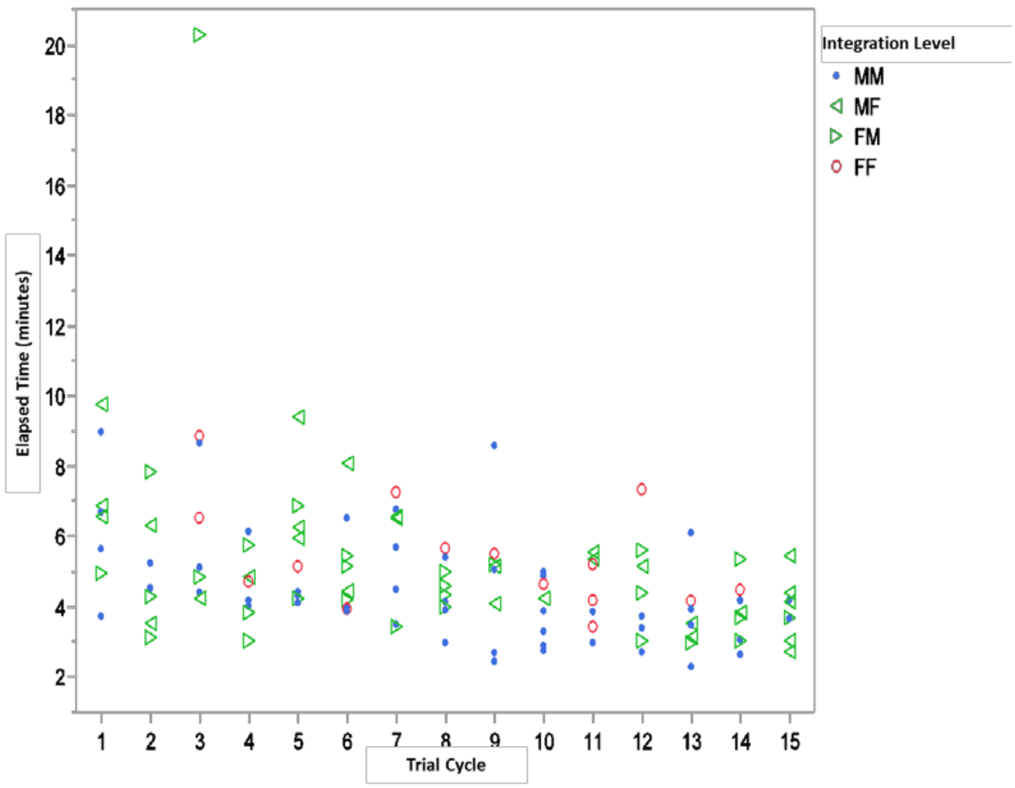
J.5.2.2 Conduct Interior Reload #1 by Critical Billet

See Section J.5.2.1 for task description.

J.5.2.2.1 Conduct Interior Reload #1 by Critical Billet Scatterplots

The scatter plots below display results for completing the Interior Reload # 1 task by critical billet. Despite the data point on trial cycle three having a significant higher elapsed time, it was determined to be a valid point and was included in the analysis. No outliers were identified for this sub-section.

Figure J-2. Interior Reload #1 by Critical Billet



J.5.2.2.2 Conduct Interior Reload #1 by Critical Billet Data Table and Analysis

The tables below summarize the results of the task, Interior Reload #1 by Critical Billet. The first table compares sample sizes and means across metrics and critical billets. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-4. Interior Reload #1 by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Interior Reload #1 (Rear Crewman/Turret Gunner) (minutes) *	MM	51	4.47	1.55
	MF	29	5.44	1.86
	FM	30	5.04	3.09
	FF	15	5.41	1.50

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.

Table J-5. Interior Reload #1 by Critical Billet ANOVA

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Interior Reload #1 (Rear Crewman/Turret Gunner) (minutes)	5.73 (3, 7.8)	0.02†	MF-MM	0.97	21.70%	0.01*	0.40	1.40	0.28	1.50
			FM-MM	0.57	12.84%	0.33	-0.08	0.65	-0.20	0.73
			FF-MM	0.94	21.01%	0.01*	0.45	1.42	0.30	1.58
			MF-FM	0.40	7.86%	0.10	0.12	1.20	0.00	1.30
			FF-MF	-0.03	-0.57%	0.87	-0.68	0.70	-0.88	0.92
			FF-FM	0.37	7.25%	0.09	0.18	1.22	0.02	1.40

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
†Results presented are from Kruskal-Wallis and Mann-Whitney tests due to lack of normality.

Because of lack of normality, we recommend using the Kruskal-Wallis and Mann-Whitney test results presented in the table above. Also of note, the standard deviation of the FM group is more than twice that of the FF group, however, we continue to recommend the Kruskal-Wallis and Mann-Whitney results above since it is more conservative than parametric alternatives. For this task, the MF and FF groups are normally distributed with Shapiro-Wilk Test p-values of 0.11 and 0.20, respectively, while the MM and FM groups are not normally distributed with p-values less than 0.01.

Table J-5 display the Rear Crewman and Turret Gunner billets as a two-letter representation, where the first letter represents the gender of the Marine occupying the Rear Crewman billet and the second letter represents the gender of the Marine occupying the Turret Gunner billet, M for male and F for female.

The mean time for interior reload #1 was faster when a male occupied the Turret Gunner position. The MF and the FF combination performed the interior reload 21.70% and 21.01% slower, respectively, than the MM combination.

The mean time for interior reload #1 was faster when a male occupied the Rear Crewman position. The FM combination performed the interior reload 12.84% slower than the MM combination.

The MF combination was 21.7% slower than the MM combination and 7.86% slower than the FM combination, both show no statistical significance. The FM combination was 12.84% slower than the MM combination: this difference is not statistically significant.

J.5.2.2.3 Conduct Interior Reload #1 by Critical Billet Contextual Comments

The Rear Crewman is responsible for the lifting of the boxes of ammunition and assisting the Turret Gunner in the reloads. This does not appear to be an experience disparity as the Rear Crewman is strictly moving boxes of ammunition up to the Turret Gunner.

In combat operations, faster reload times (even measured in a few seconds) are highly desirable. Faster reloading increases the combat effectiveness and survivability of an AAV crew. Per the MCWP 3-13, “The AA unit should continue to engage the enemy as fast as possible, because fire placed in the enemy’s area will less his effectiveness and give friendly weapons time to adjust. A wasted opportunity to engage a target may never be regained.” With the service life of the AAVP7 projected to extend to 2035, selecting AAV Rear Crewmen who are able to quickly move boxes of ammunition and assist the Turret Gunner plays a significant role in maintaining the competitiveness of the platform.

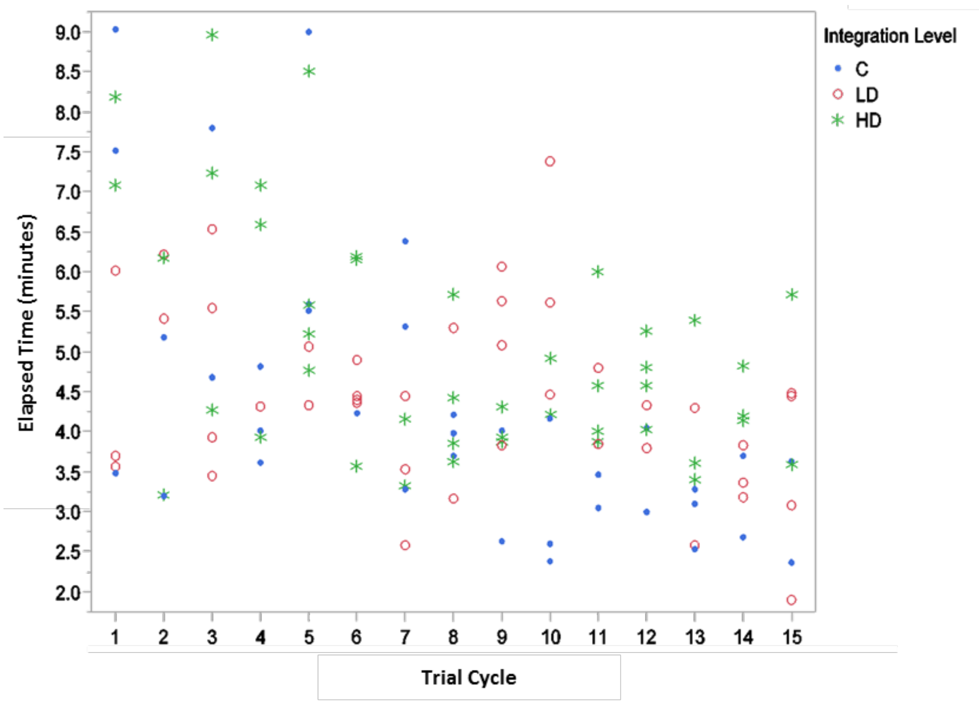
J.5.2.3 Conduct Interior Reload #2 by Integration Level

See Section J.5.2.1 for task description.

J.5.2.3.1 Conduct Interior Reload #2 by Integration Level Scatterplots

The scatterplot below displays the data used in the analysis for the Interior Reload task. No outliers were identified.

Figure J-3. Interior Reload #2 by Integration Level



J.5.2.3.2 Conduct Interior Reload #2 by Integration Level Data Table and Analysis

The tables below summarize the results of the task, Interior Reload #2 by Integration Level. The first table compares sample sizes and means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-6. Interior Reload #2 by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Interior Reload #2 (minutes) *	C	37	4.32	1.72
	LD	42	4.41	1.14
	HD	44	5.03	1.45

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

Table J-7. Interior Reload #2 by Integration Level ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Interior Reload #2 (minutes)	3.01 (2, 120)	0.05*	LD-C	0.10	2.23%	0.95	-0.47	0.66	-0.58	0.77
			HD-C	0.71	16.49%	0.07*	0.16	1.27	0.04	1.38
			HD-LD	0.62	13.95%	0.12	0.08	1.15	-0.03	1.26

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

For this task, samples sizes are sufficiently large (n>30) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

Table J-6 and Table J-77 show the control group executed the task faster than the LD and HD groups. The LD group was 2.23% slower than the C group; this difference is not statistically significant. The HD group was 16.49% slower than the C group; this difference is statistically significant, while the HD group was 13.95% slower than the LD group; this difference is not statistically significant in a Tukey test.

J.5.2.3.3 Conduct Interior Reload #2 by Integration Level Contextual Comments

In combat operations, faster reload times (even measured in a few seconds) are highly desirable. Faster reloading increases the combat effectiveness and survivability of an AAV crew. Per the MCWP 3-13, “The AA unit should continue to engage the enemy as fast as possible, because fire placed in the enemy’s area will less his effectiveness and give friendly weapons time to adjust. A wasted opportunity to engage a target may never be regained.” With the service life of the AAVP7 projected to extend to 2035, selecting AAV Rear Crewmen who are able to quickly move boxes of ammunition and assist the Turret Gunner plays a significant role in maintaining the competitiveness of the platform.

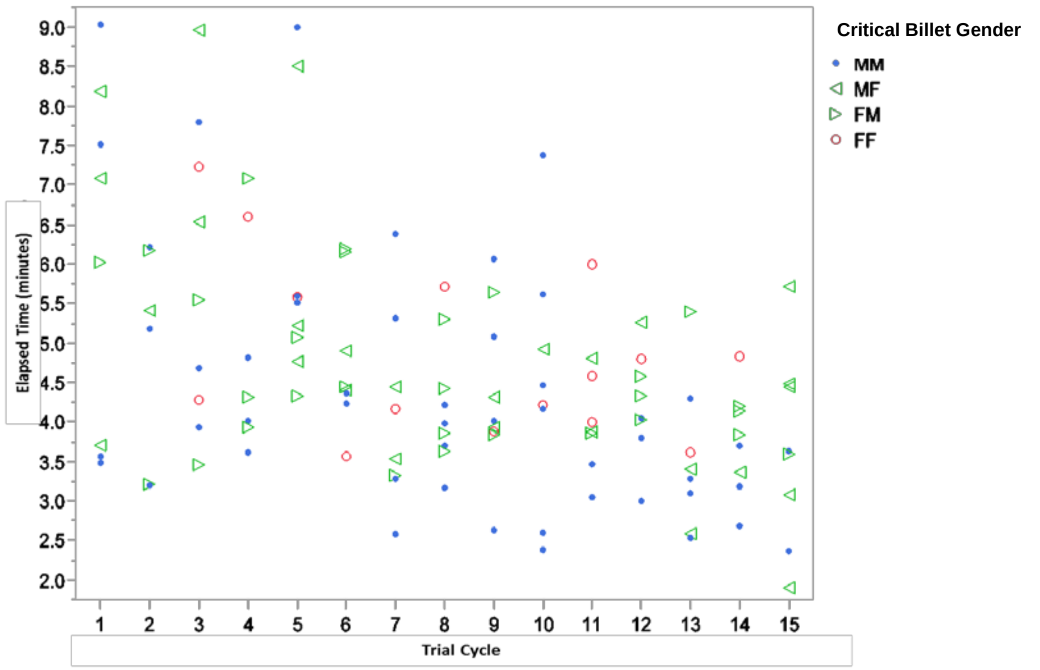
J.5.2.4 Conduct Interior Reload #2 by Critical Billet

See Section J.5.2 for task description.

J.5.2.4.1 Conduct Interior Reload #2 by Critical Billet Scatterplots

The scatterplot below displays the data used in the analysis for this task.

Figure J-4. Interior Reload #2 by Critical Billet



J.5.2.4.2 Conduct Interior Reload #2 by Critical Billet Data Table and Analysis

The tables below summarize the results of the task, Interior Reload #2 by Critical Billet. The first table compares sample sizes and means across metrics and critical billets. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-8. Interior Reload # 2 by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Interior Reload #2 (Rear Crewman/Turret Gunner) (minutes)	MM	51	4.38	1.62
	MF	28	4.85	1.70
	FM	29	4.62	1.03
	FF	15	4.87	1.12

Table J-9. Interior Reload # 2 by Critical Billet ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Interior Reload #2 (Rear Crewman/Turret Gunner (minutes)	0.82 (3, 119)	0.49	MF-MM	0.47	10.70%	0.53	-0.22	1.15	-0.33	1.27
			FM-MM	0.24	5.42%	0.90	-0.44	0.92	-0.55	1.03
			FF-MM	0.49	11.21%	0.67	-0.37	1.35	-0.51	1.49
			MF-FM	0.23	5.01%	0.93	-0.54	1.00	-0.67	1.13
			FF-MF	0.02	0.46%	1.00	-0.91	0.96	-1.07	1.11
			FF-FM	0.25	5.50%	0.95	-0.67	1.18	-0.83	1.34

Table J-8 and Table J-9 display the Rear Crewman and Turret Gunner billets as a two-letter representation, where the first letter represents the gender of the Marine occupying the Rear Crewman billet and the second letter represents the gender of the Marine occupying the Turret Gunner billet, M for Male and F for female.

The MF combination was 10.7% slower than the MM combination and 5.01% slower than the FM combination; these differences are not statistically significant. The FM combination was 5.42% slower than the MM combination and still not statistically significant. See tables above for detailed analytical results.

J.5.2.4.3 Conduct Interior Reload #2 by Critical Billet Contextual Comments

In combat operations, faster reload times (even measured in a few seconds) are highly desirable. Faster reloading increases the combat effectiveness and survivability of an AAV crew. Per the MCWP 3-13, “The AA unit should continue to engage the enemy as fast as possible, because fire placed in the enemy’s area will less his effectiveness and give friendly weapons time to adjust. A wasted opportunity to engage a target may never be regained.” With the service life of the AAVP7 projected to extend to 2035, selecting AAV Rear Crewmen who are able to quickly move boxes of ammunition and assist the Turret Gunner plays a significant role in maintaining the competitiveness of the platform

J.5.2.5 Conduct External Reload Overview

The AAV Turret Gunner and Rear Crewman conducted a reload of 200 rounds of .50-caliber ammunition and 96 rounds of 40-mm grenade rounds. A loaded can of .50-caliber ammunition weighs approximately 35 pounds and a loaded can of 40-mm grenade rounds weighs approximately 45 pounds. To assist the Turret Gunner, the Rear Crewman moved boxes of ammunition from the appropriate troop compartment stowage spaces to the top of the vehicle and into the turret from the exterior. This task was executed once during the course of fire. The task simulated the requirement for the Rear Crewman to assist the Turret Gunner with reloading weapons systems from the exterior of the AAV. The Data Collector was located in the Troop Compartment of the AAV and recorded start times based on a verbal command from either the Master Gunner via radio or from the AAV Turret Gunner via vehicle intercom. The Data Collector recorded stop times when the Turret Gunner notified the entire vehicle that the reload was complete via vehicle intercom. We analyzed performance on this task two ways: by integration level, as well as by the gender of the Marine occupying the two critical billets (Rear Crewman and Turret Gunner).

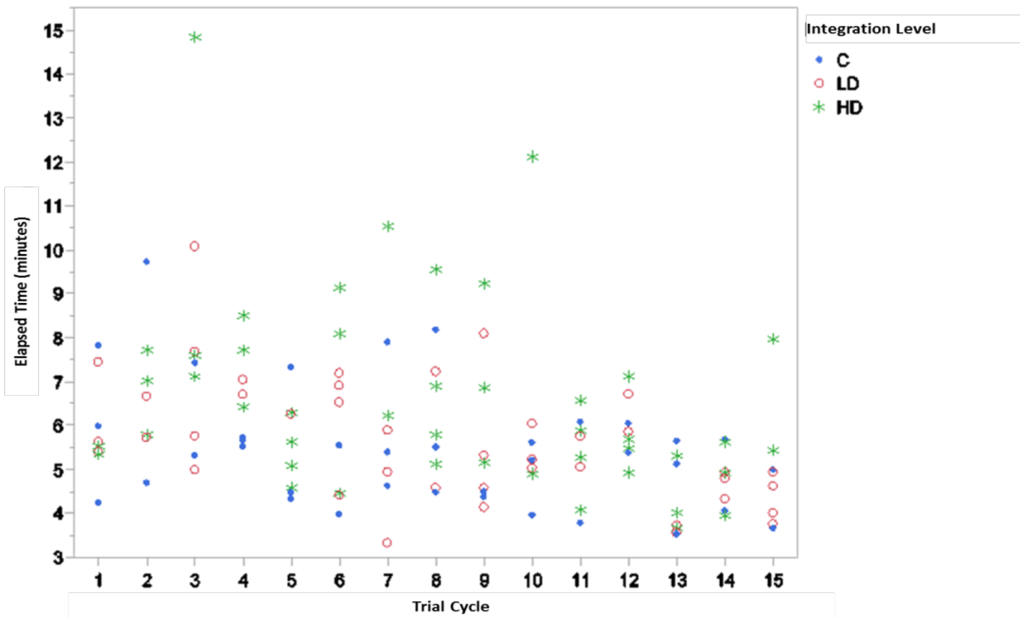
J.5.2.5.1 Conduct External Reload by Integration Level

An exterior reload is a physically demanding task: in addition to moving the boxes of ammunition, it requires the Marine to open and lock the starboard rear cargo hatch, lift the ammunition approximately 6.5 ft., and then pull themselves up to the top of the vehicle in a full combat load. The task was performed as a crew event with the AAV Rear Crewman lifting the ammunition individually, with the primary measure of performance being the time required to complete the task.

J.5.2.5.2 Conduct External Reload by Integration Level

The scatterplot below displays the data used in the analysis for the external reload task.

Figure J-5. Exterior Reload by Integration Level



J.5.2.5.3 Conduct External Reload by Integration Level Data Table and Analysis

The tables below summarize the results of the task, Exterior Reload by Integration Level. The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-20. Exterior Reload by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
External Reload (minutes)*	C	37	5.45	1.40
	LD	42	5.64	1.38
	HD	45	6.55	2.20

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level

Table J-3. Exterior Reload by Integration Level ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value (one-sided)	80 % LCB	80% UCB	90% LCB	90% UCB
External Reload (minutes)	4.94 (2, 121)	0.01*	LD-C	0.19	3.41%	0.88	-0.49	0.86	-0.62	0.99
			HD-C	1.10	20.24%	0.01*	0.44	1.77	0.31	1.90
			HD-LD	0.92	16.28%	0.04*	0.28	1.56	0.15	1.69

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

For this task, samples sizes are sufficiently large (n>30) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

Table J-20 and Table J-31 show the LD group was 3.41% slower than the C group; this is not statistically significant. The HD group was 20.24% slower than the C group; this difference is statistically significant in a Tukey test. The HD group was 16.28% slower than the LD group; this difference is statistically significant. The control group had a mean time of 5.45 minutes, faster than the LD and HD groups, with mean times of 5.64 and 6.55 minutes, respectively. The HD group was significantly slower than the C and LD groups.

J.5.2.5.4 Conduct External Reload by Integration Level Contextual Comments

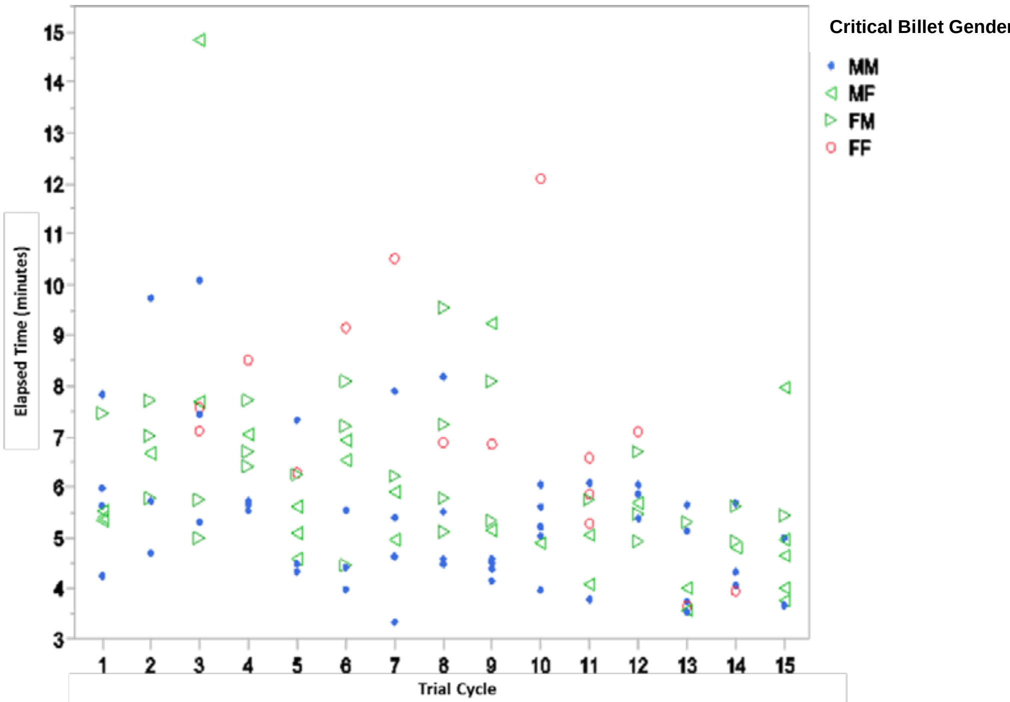
Similar to the interior reloads, in combat operations, faster reload times (even measured in fractions of a second) are highly desirable. Faster reloading increases the combat effectiveness and survivability of an AAV crew. Per the MCWP 3-13, “The AA unit should continue to engage the enemy as fast as possible, because fire placed in the enemy's area will less his effectiveness and give friendly weapons time to adjust. A wasted opportunity to engage a target may never be regained.” With the service life of the AAVP7 projected to extend to 2035, selecting AAV Rear Crewmen who are able to quickly move boxes of ammunition and assist the Turret Gunner plays a significant role in maintaining the competitiveness of the platform.

J.5.2.6 Conduct External Reload by Critical Billet

See Section J.5.2.5 and J5.2.5.1 for Task Description.

The scatterplot below displays the data used in the analysis of the external reload by critical billet task.

Figure J-6. Exterior Reload by Critical Billet



J.5.2.6.1 Conduct External Reload by Critical Billet Data Table and Analysis

The tables below summarize the results of the task, Exterior Reload (Analysis by Critical Billet). The first table compares means across metrics and integration levels. The second table presents ANOVA and Tukey test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-12. Exterior Reload by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
External Reload (Rear Crewman/Turret Gunner) (minutes)*	MM	51	5.38	1.45
	MF	28	5.85	2.21
	FM	30	6.26	1.21
	FF	15	7.16	2.25

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

Table J-13. Exterior Reload by Critical Billet ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
External Reload (Rear Crewman/Turret Gunner)	4.72 (3, 120)	< 0.01*	MF-MM	0.47	8.75%	0.65	-0.33	1.27	-0.46	1.40
			FM-MM	0.87	16.24%	0.12	0.09	1.65	-0.04	1.79
			FF-MM	1.78	33.08%	< 0.01*	0.78	2.78	0.62	2.94

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
(minutes)			MF-FM	-0.40	-6.45%	0.81	-1.29	0.49	-1.44	0.64
			FF-MF	1.31	22.38%	0.08*	0.22	2.40	0.04	2.58
			FF-FM	0.91	14.49%	0.34	-0.17	1.98	-0.35	2.16

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

We proceed with presenting ANOVA results because they were confirmed by a Kruskal-Wallis test with a p-value of less than 0.01. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For this task, the FM and FF groups are normally distributed with Shapiro-Wilk Test p-values of 0.08 and 0.30, respectively, while the MM and MF groups are not normally distributed with p-values less than 0.01.

Table J-12 and Table J-13 display the Rear Crewman and Turret Gunner billets as a two-letter representation, where the first letter represents the gender of the Marine occupying the Rear Crewman billet and the second letter represents the gender of the Marine occupying the Turret Gunner billet, M for Male and F for female.

The MF combination was 8.75% slower than the MM combination, but 6.45% faster than the FM combination. In both cases, the differences are not statistically significant. The FF combination was 33.08% slower than the MM combination and 22.38% slower than the MF combination. In both cases, the differences are statistically significant. See tables above for further detailed analysis among all comparisons.

J.5.2.6.1.1 Conduct External Reload by Critical Billet Contextual Comments

It was important to identify who was serving in the critical billets of AAV Rear Crewman and AAV Turret Gunner to remove the non-participating billet from the analysis. When a male served in the Turret Gunner billet for all combinations except one (MF), the task was executed faster when compared to other combinations. This may be attributed to the experience of the Turret Gunner in taking both weapons systems from Condition 3, once the Rear Crewman had finished with the movement of the ammunition cans to the top of the vehicle, to Condition 1.

Similar to the interior reloads, in combat operations, faster reload times (even measured in fractions of a second) are highly desirable. Faster reloading increases the combat effectiveness and survivability of an AAV crew. Per the MCWP 3-13, "The AA unit should continue to engage the enemy as fast as possible, because fire placed in the enemy's area will less his effectiveness and give friendly weapons time to adjust. A wasted opportunity to engage a target may never be regained." With the service life of the AAVP7 projected to extend to 2035, selecting AAV Rear Crewmen who are able to quickly move boxes of ammunition and assist the Turret Gunner plays a significant role in maintaining the competitiveness of the platform.

J.5.2.7 Break and Reassemble Track Overview

In this Maintenance Action subtask, the AAV crew used the vehicle SL-3 to break the vehicle track at the rear of the vehicle. The crews then moved the vehicle forward until the broken track was at the mid-point of the AAV at the dual-support roller. The vehicle was then placed in reverse and moved back until the crew was able to put the track back together. Upon completion of the setup, an AAV Platoon Staff member inspected the vehicle to ensure it had been properly reassembled and that the crew had followed all necessary safety requirements. Upon verification, the crew would then stow all materials used and move the vehicle forward 60 feet to a marked point, signaling completion of the subtask. The crew had 1 hour and 30 minutes to accomplish this task. This task simulates multiple possible suspension and vehicle track issues that could arise during the course of operations. We compared task performance by integration level of the crew as well as by the gender of the Marines in the two critical billets (Rear Crewman and Turret Gunner).

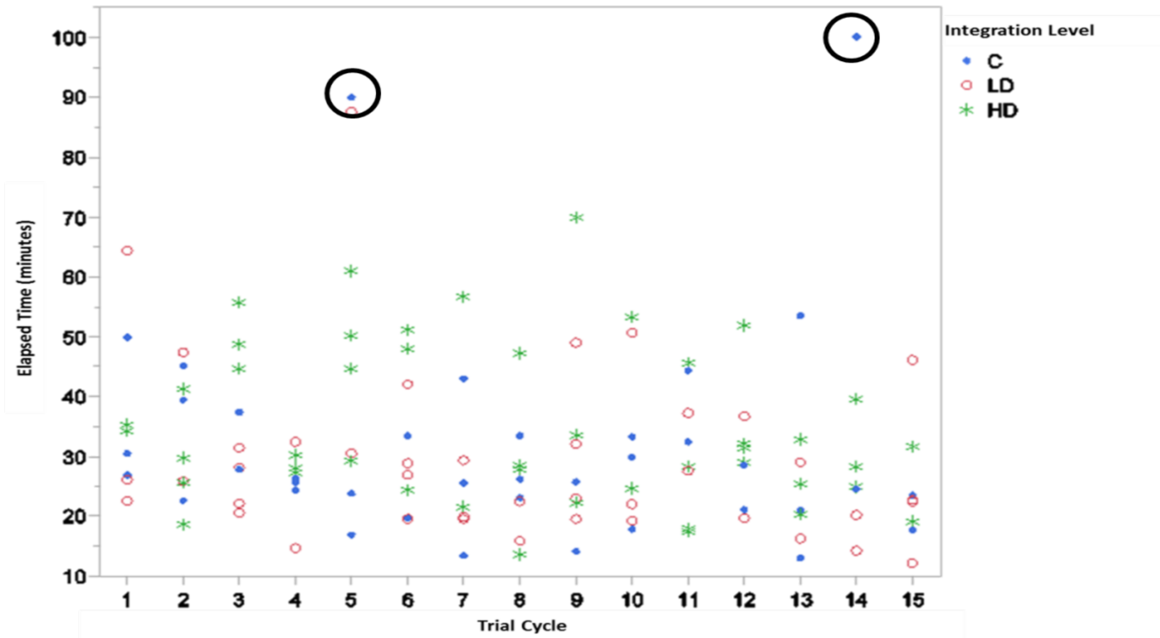
J.5.2.7.1 Break and Reassemble Track by Integration Level

Breaking and reassembling track is a physically demanding task: the AAV crew used sledgehammers, tank bars, and other assorted SL-3 to break the 35-lb individual track blocks apart before putting them back together. The task was performed as a crew event. However, the Driver did not participate in the more physically demanding part of the task.

J.5.2.7.1.1 Break and Reassemble Track by Integration Level Scatterplots

The scatterplot below displays the data used in the analysis of the results of the Break and Reassemble Track by integration level task. On trial cycles 5 and 14, two C-group data points were removed from analysis due to data entry errors.

Figure J-7. Break/Reassemble Track by Integration Level



J.5.2.7.1.2 Break and Reassemble Track by Integration Level Data Table and Analysis

The tables below summarize the results of the task, Break and Reassemble Track +by Integration Level. The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-14. Break and Reassemble Track by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Break/Reassemble Task (minutes)*	C	36	28.10	9.92
	LD	42	29.02	14.50
	HD	46	34.42	12.81

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.

Table J-15. Break and Reassemble Track by Integration Level ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Break/Reassemble Task (minutes)	3.11 (2, 121)	< 0.05*	LD-C	0.92	3.28%	0.95	-4.05	5.89	-5.05	6.89
			HD-C	6.32	22.51%	0.07*	1.46	11.19	0.48	12.17
			HD-LD	5.40	18.62%	0.12	0.73	10.07	-0.20	11.01

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.

**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.

For this task, samples sizes are sufficiently large (n>30) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The HD group was 22.51% slower than the C group; this difference is statistically significant. The HD group was 18.62% slower than the LD group and this difference is not statistically significant. On average, the control group performed the task faster than the LD and HD groups, with a mean time of 28.1 minutes, which was 55 seconds faster than the LD-group mean and 6:19 faster than the HD-group mean. The analytical results show metric mean values to be statistically significant only when we compare the HD and C groups: a 22.51% difference.

J.5.2.7.1.3 Break and Reassemble Track by Integration Level Contextual Comments

However, in a combat environment, it is desirable to fix suspension issues that would cause a vehicle to be disabled, quickly, to deny the enemy the opportunity to maneuver or employ fires against a stationary AAV.

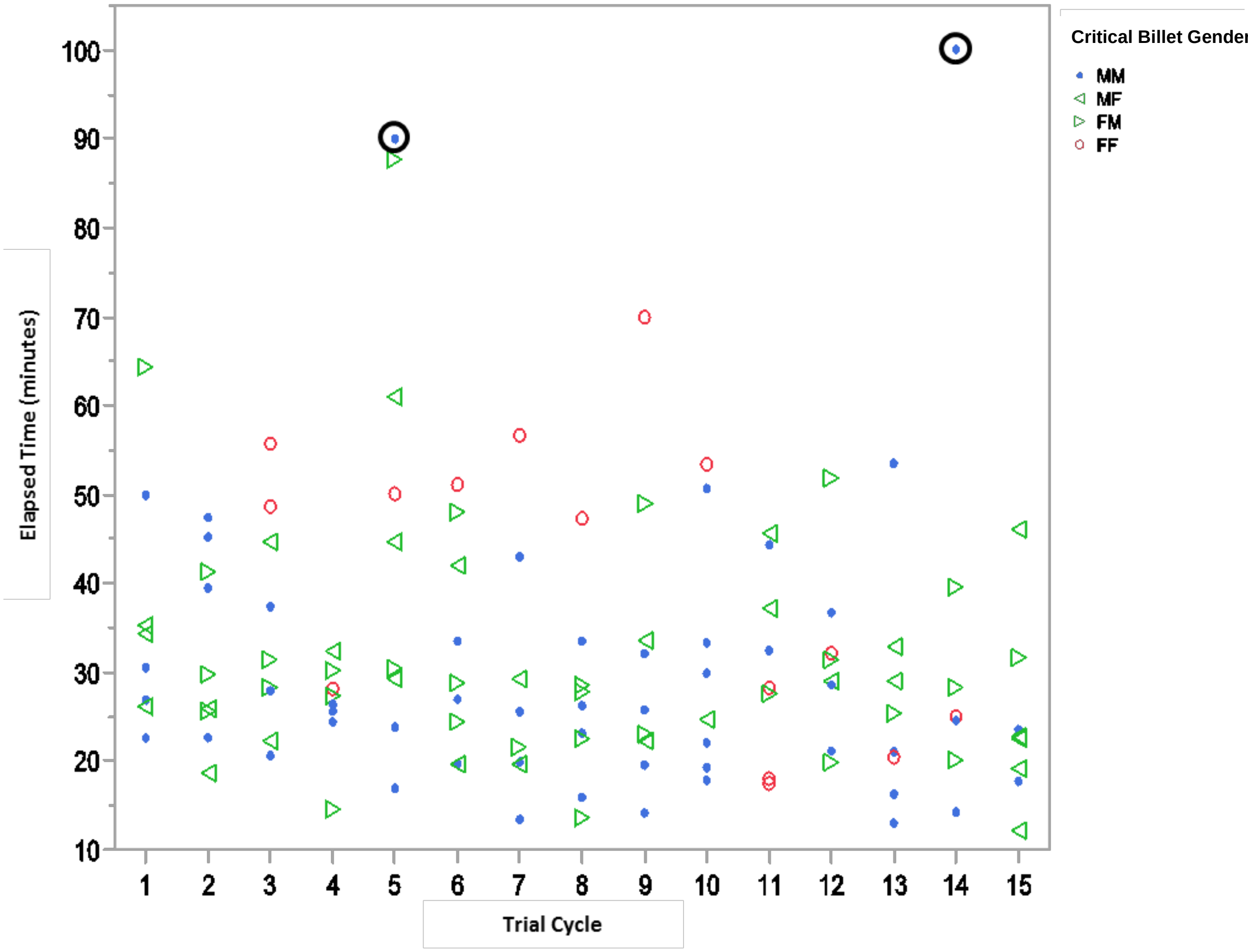
J.5.2.7.2 Break and Reassemble Track by Critical Billet

See Section J.5.2.7 and J.5.2.7.1 for task description.

J.5.2.7.3 Break and Reassemble Track by Critical Billet Scatterplots

The scatter plots below display the results to complete the Break and Reassemble Track Task by critical billet. On trial cycles 5 and 14, the data points contained inside dark black circles were determined to be data errors and were excluded from the analysis.

Figure J-8. Break/Reassemble Track by Critical Billet



J.5.2.7.4 Break and Reassemble Track by Critical Billet Data Table and Analysis

The tables below summarize the results of the task, Break and Reassemble Track by Critical Billet. The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-17. Break/Reassemble Track by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Break/Reassemble Task (Rear Crewman/Turret Gunner) (minutes)*	MM	50	27.51	10.32
	MF	29	30.39	10.89
	FM	30	32.04	15.07
	FF	15	39.71	15.91

*Indicates there is a statistically significant difference, in a two-sided hypothesis test between the metric's mean values for the Integration Level.

Table J-16. Break/Reassemble Track by Critical Billet ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Break/Reassemble Track (Rear Crewman/Turret Gunner) (minutes)	3.82 (3, 120)	0.01*	MF-MM	2.88	10.46%	0.76	-2.90	8.65	-3.87	9.62
			FM-MM	4.52	16.43%	0.40	-1.19	10.24	-2.15	11.19
			FF-MM	12.20	44.34%	< 0.01*	4.92	19.48	3.69	20.71
			MF-FM	-1.64	-5.13%	0.96	-8.09	4.80	-9.17	5.88
			FF-MF	9.32	30.67%	0.09*	1.45	17.19	0.13	18.51
			FF-FM	7.68	23.97%	0.21	-0.15	15.50	-1.46	16.82

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

We proceed with presenting ANOVA results because they were confirmed by a Kruskal-Wallis test with a p-value of 0.05. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. For this task, the MF and FF groups are normally distributed with Shapiro-Wilk Test p-values of 0.10 and 0.16, respectively, while the MM and FM groups are not normally distributed with p-values less than 0.01.

Rear Crewman and Turret Gunner billets are represented two-letters, where the first letter represents the gender of the Marine occupying the Rear Crewman billet and the second letter represents the gender of the Marine occupying the Turret Gunner billet, M for Male and F for female.

All three combinations that included a male (i.e., MM, MF, FM) in one of the critical billets had faster break-track times than the all-female combination (FF). When the (MM) and (MF) combinations in which the rear crewman was a male were compared to the (FF) combination, they showed a significant difference and a much higher statistical significance result than other comparative groups.

Table J-116 and Table J-1617 show that the MF and FM combinations were 10.46% and 16.43%, respectively, slower when compared to the MM combination; in both cases, the differences are not statistically significant. The MF combination was 5.13% faster than the FM combination; this difference is not statistically significant.

The FF combination was 44.34% slower than the MM combination and 30.67% slower than the MF combination. In both cases, these differences were statistically significant.

J.5.2.7.5 Break and Reassemble Track by Critical Billet Contextual Comments

This second comparison analyzed the performance differences between crews grouped only by the gender of the Marines serving in the critical billets. In this particular task, the critical billet's responsibilities are not scripted (meaning it is a group-effort task shared mostly between the two critical billets vice all three crewmembers). The impact of the driver on this metric is considered to be further minimalized because he/she is operating the vehicle in strict accordance with the ground guide's instructions and not operating independently nor solely responsible for the movement of the vehicle during this task.

In a combat environment, it is desirable to fix suspension issues that would cause a vehicle to be disabled, quickly, to deny the enemy the opportunity to maneuver or employ fires against a stationary AAV. The difference demonstrated in this experiment may have a significant impact during ground combat operations.

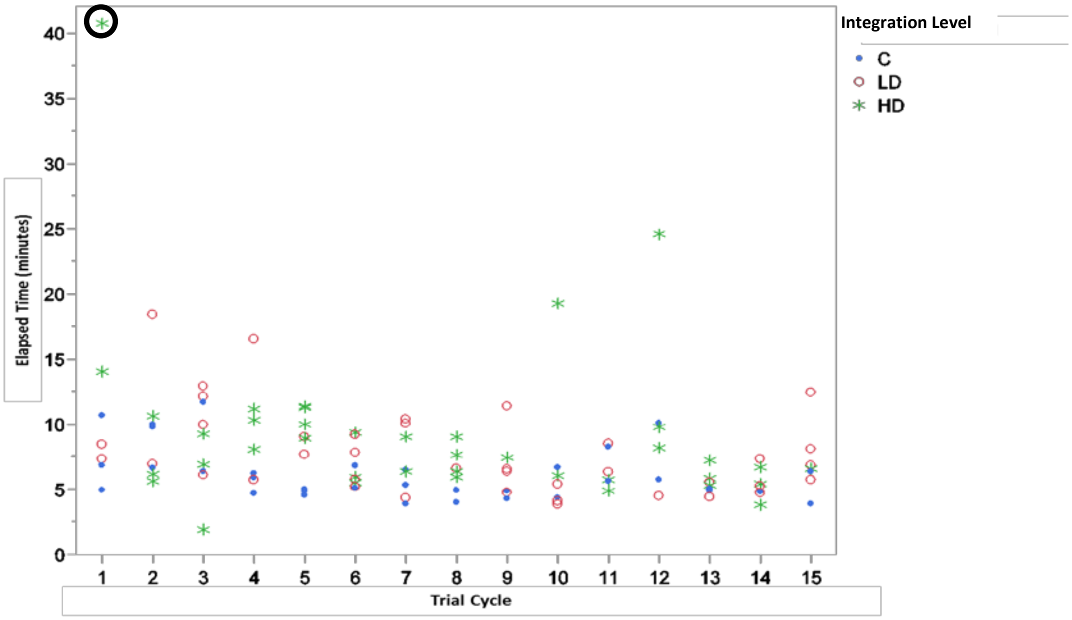
J.5.2.8 Load Weapons and Ammunition Overview

The vehicle crew loaded the M2 .50-caliber machine gun, the MK-19 40-mm grenade launcher, and a full complement of ammunition (two cans of .50-caliber ammunition and three cans of 40-mm grenades) from the lowered ramp of the AAV. The crew executed this task from the exterior of the vehicle (external mount of M2, MK-19, and Ammo). An interior loading of the fully assembled M2 was conducted, as well (see Appendix for those results). A fully assembled M2 weighs approximately 85 lb. and the MK-19 weighs approximately 79 lb. Upon completion, an AAV Platoon Staff member inspected to ensure the weapons were installed correctly. The Data Collector was located on the near the Troop Compartment of the AAV, and initiated this task with a verbal start time, recording start and stop times. There was no maximum time limit for this task. This task simulated the AAV crew loading weapons and ammunition from various positions on the vehicle. The task was performed as a crew event, with the primary measure of performance being the time required to complete the task. The data collected from this task was analyzed by crew integration only, and not by critical billet, as each Marine in the AAV Crew had an active role in execution.

J.5.2.8.1 External Mount of M2, MK-19, and Ammo Scatterplot

The scatter plots below display the results to complete the External Mount of M2, MK-19, and Ammo task. The data point contained inside the dark black circle was determined to be a data error and was excluded from the analysis.

Figure J-9. Exterior Mount of M2, MK-19, and Ammo



J.5.2.8.2 External Mount of M2, MK-19, and Ammo Data Table and Analysis

The tables below summarize the results of the task, External Mount of M2, MK-19, and Ammo by integration level. The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-18. External Mount of M2, MK-19, and Ammo

Metric	Integration Level	Sample Size	Mean	SD
Ext Mount of M2 & Mk19 and Ammo (minutes)*	C	37	6.17	2.01
	LD	42	7.62	3.25
	HD	45	7.92	3.50

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.

Table J-19. External Mount of M2, MK-19, and Ammo ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Ext Mount of M2 & Mk19 and Ammo (minutes)	3.76 (2, 121)	0.03*	LD-C	1.46	23.63%	0.09*	0.28	2.64	0.04	2.88
			HD-C	1.75	28.43%	0.03*	0.59	2.92	0.36	3.15
			HD-LD	0.30	3.88%	0.89	-0.83	1.42	-1.05	1.65

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.

**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

For this task, samples sizes are sufficiently large (n>30) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The C group performed the task faster with a mean time of 6:17 minutes when compared to the LD and HD groups which were significantly slower than the control group. The LD group was 23.63% slower than the C group, while the HD group was 28.43% slower than the C group. Both differences are statistically significant.

J.5.2.8.3 External Mount of M2 & MK-19, and Ammo Contextual Comments

Unless under an extreme situation, either in combat or in training, the difference of roughly 1.5 to 2 minutes in mounting both weapons systems externally and allocating ammunition could be considered negligible. Environments in which crews will be expected to perform this task will dictate the level of importance a 2-minute difference can make on subsequent AAV actions.

J.5.2.9 Conduct CASEVAC of an Incapacitated Turret Gunner Overview

In this task, the Rear Crewman and the Driver evacuated a simulated casualty from the turret of the AAV, externally and internally. The simulated casualty was a dummy wearing the appropriate vehicle crewman fighting load, weighing approximately 205 lb. The Rear Crewman performed each evacuation alone for the first 60 seconds, while the Driver radioed a simulated CASEVAC 9-line report. If the simulated casualty was not evacuated by the 60-second mark, the Driver then assisted. The Driver and the Rear Crewman then executed the evacuation until the 6-minute mark. Then, the Turret Gunner assisted between the 6-minute mark and the time limit of 9 minutes. This additional assistance simulated a Marine from another AAV assisting the Rear Crewman and the Driver in cases when further help is necessary. Once the simulated casualty was placed on top of the vehicle for the exterior evacuation or on the floor of the troop compartment for the interior evacuation, the Data Collector recorded the stop time. The Data Collector was located in the Troop Compartment of the AAV and was responsible for verifying that the dummy had not sustained additional head injuries during the evacuation. An accelerometer check-light (green, yellow, or red) identified any additional injury. A red check light meant the casualty sustained significant head injury during the casualty evacuation; a green light signaled no injury; while a yellow check light signaled additional head injury but not life threatening. This task simulated conducting a CASEVAC from the turret, and is a T&R requirement for 1833s. The task was performed as a crew event, with the primary measure of performance being the time required to complete the task and the AAV Turret Gunner not actively participating until after the 6-minute mark.

The data collected from this task was analyzed two ways. The first comparison looked for performance differences between the different integration-level groups. The second comparison looked for performance differences between crews grouped only by the gender of the Marines serving in the critical billets (AAV Rear Crewman and AAV Driver), and actively participating in the CASEVAC.

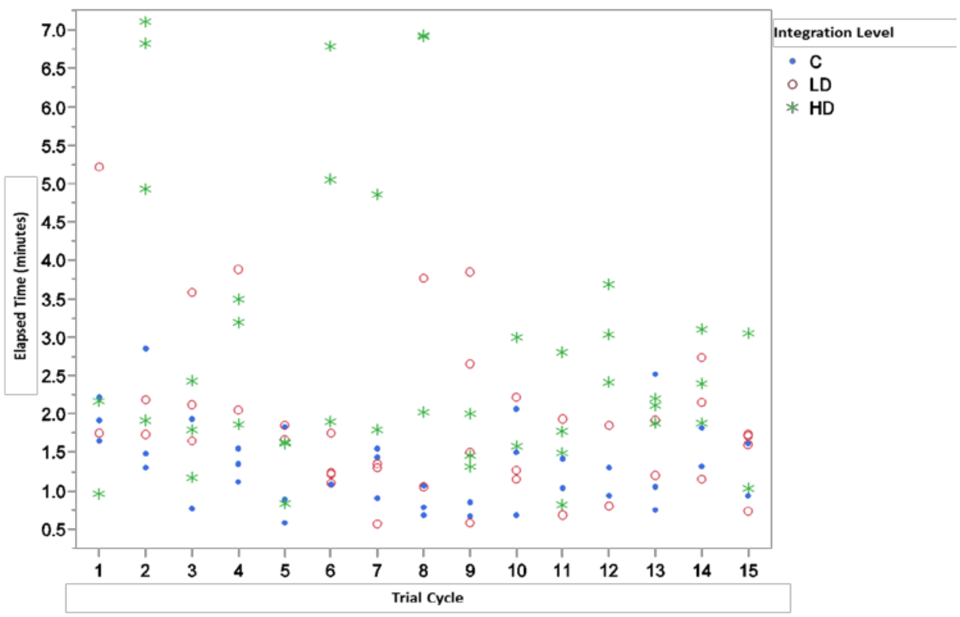
J.5.2.9.1 External CASEVAC of an Incapacitated Turret Gunner by Integration Level

See Section J.5.2.9 for task description

J.5.2.9.1.1 External CASEVAC of an Incapacitated Turret Gunner by Integration Level Scatterplots

Figure J-1 displays the results to complete the External CASEVAC of an Incapacitated Turret Gunner task by integration level. There were no outliers identified in the dataset.

Figure J-1. External CASEVAC from Turret by Integration Level



J.5.2.9.1.2 External CASEVAC of an Incapacitated Turret Gunner by Integration Level Data Table and Analysis

The tables below summarize the results of the task, External CASEVAC by Integration Level. The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-20. External CASEVAC of an Incapacitated Turret Gunner by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
External CASEVAC from Turret (minutes)*	C	38	1.33	0.54
	LD	41	1.86	0.99
	HD	44	2.81	1.79

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.

Table J-21. External CASEVAC of an Incapacitated Turret Gunner by Integration Level ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
External CASEVAC from Turret (minutes)	15.90 (2, 71.2)	< 0.01††	LD-C	0.53	39.63%	< 0.01*	0.30	0.76	0.23	0.83
			HD-C	1.48	110.96%	< 0.01*	1.11	1.84	1.00	1.95
			HD-LD	0.95	51.08%	< 0.01*	0.55	1.35	0.43	1.47

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances)

For this task, sample sizes are sufficiently large (n>30) to satisfy the normality assumption for ANOVA. However, the standard deviation for the HD group is more than twice that of the C group. Thus, we recommend using the robust ANOVA results presented in the table above.

The LD group was 39.63% slower than the C group, while the HD group was 110.97% slower than the C group; this difference is statistically significant. The control group had a mean time of 1.33 minutes, which is 39% faster than the LD group and 110% faster than the HD group.

J.5.2.9.1.3 External CASEVAC of an Incapacitated Turret Gunner by Integration Level Contextual Comments

When conducting the analysis of this task by just the concentration (HD, LD, C) without consideration to the critical billets and the order in which those billets executed the CASEVACS, there exists the potential to lose some understanding based on the concentration level alone. In these tasks, there existed the potential for low-density crews to have two males as the critical billet holders and high-density crews to have a FM or MF combination for the critical billet holders. Analysis using only concentration levels limits our visibility on what actually took place within the crews by masking gender impact. This skews the data analysis as there appears to be no significant effect of integration based on the similar mean times found in both CASEVACs when looking at just the crew concentration (between 1.33 minutes to 2.81 minutes or a difference of 1 minute and 29 seconds). By looking instead at the task by critical billets, the data show more significant differences between the mean times, as shown by the percent differences in time to execute the task. This breakout in the data is shown in subsequent paragraphs.

J.5.2.9.1.4 External CASEVAC of an Incapacitated Turret Gunner by Integration Level Additional Insights

The requirement of an AAV Rear Crewman and/or AAV Driver to execute a CASEVAC of an incapacitated AAV Turret Gunner is a T&R task for an AAV Crew. Unlike similar MOSs, the AAV T&R does not stipulate a specific time requirement. The HBCT manual requires a 1812 Tank Crewman MOS to conduct an evacuation out of an M1A1 Abrams Tank in 2 minutes. The AAV task designed in this experiment was similar to the Tank HBCT requirement. The ability of a Marine or Marines to extricate a casualty from the turret quickly could mean the difference between life and death for an injured or incapacitated AAV Turret Gunner. In the case of this task, seconds could be the determining factor. Unlike similar vehicle MOSs, the AAV T&R does not stipulate a specific time requirement to complete this task for it to be considered a display of mastery or non-mastery. The maximum time of 9 minutes was based on the goal of capturing all data points that could exist above a shorter time range. This data may help the AAV community determine an acceptable time with which to standardize the Assault Amphibious T&R Manual.

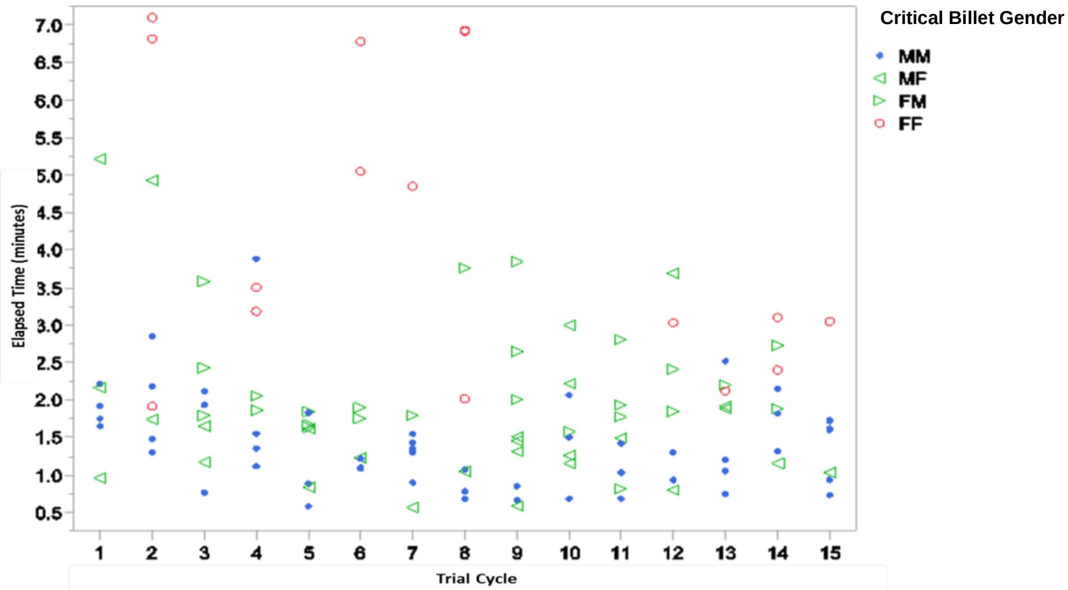
J.5.2.9.2 External CASEVAC of an Incapacitated Turret Gunner by Critical Billet

See Section J.5.2.9 for task description.

J.5.2.9.2.1 External CASEVAC of an Incapacitated Turret Gunner by Critical Billet Scatterplots

The scatter plots below display the results to complete the External CASEVAC of an Incapacitated Turret Gunner task by critical billet. There were no outliers identified in the dataset.

Figure J-11. External CASEVAC from Turret by Critical Billet



J.5.2.9.2.2 External CASEVAC of an Incapacitated Turret Gunner by Critical Billet Data Table and Analysis

The tables below summarize the results of the task, External CASEVAC (Analysis by Critical Billet). The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-22. External CASEVAC of an Incapacitated Turret Gunner by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
External CASEVAC from Turret (Rear Crewman/Driver) (minutes)*	MM	53	1.42	0.62
	MF	27	1.76	1.18
	FM	26	2.17	0.71
	FF	17	4.19	2.00

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level

Table J-23. External CASEVAC of an Incapacitated Turret Gunner by Critical Billet ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Diff	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
External CASEVAC from Turret (Rear Crewman/Driver) (minutes)	15.63 (3, 42.7)	< 0.01††	MF-MM	0.34	23.97%	0.17	0.02	0.66	-0.07	0.75
			FM-MM	0.75	52.56%	< 0.01*	0.54	0.96	0.47	1.02
			FF-MM	2.77	194.66%	< 0.01*	2.11	3.42	1.91	3.62
			MF-FM	-0.41	-18.74%	0.13	-0.75	-0.06	-0.85	0.04

Metric	F Statistic (df)	F Test P-Value	Compariso n	Diff	% Diff	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
			FF-MF	2.43	137.68%	< 0.01*	1.72	3.13	1.51	3.34
			FF-FM	2.02	93.14%	< 0.01*	1.35	2.69	1.15	2.89

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances)

Table J-22 display the Rear Crewman and Driver billets as a two-letter representation, where the first letter represents the gender of the Marine occupying the Rear Crewman billet and the second letter represents the gender of the Marine occupying the Driver billet, M for Male and F for female.

For this task, the MM, MF, FM and FF groups are all not normally distributed with Shapiro-Wilk Test p-values of less than 0.01. A Kruskal-Wallis test does confirm the results of the standard ANOVA. However, the standard deviation of the FF group is more than twice that of the MM group, thus we recommend using the results of the robust ANOVA presented in the table above.

The MM combination had a mean time of 1.42 minutes while the MF combination had a mean time of 1.76 minutes. The FM combination had a mean time of 2.17 minutes with a SD of 0.71 minutes. The FF combination had a mean time of 4.19 minutes with a SD of 2 minutes. The FM combination was 52.56% slower than the MM combination. The FF combination was 194.66% slower than the MM combination and 137.68% slower than the MF combination. Analytical results show the gender of the Marine in the first critical billet, the AAV Rear Crewman, was significant. When the AAV Rear Crewman (who had 60 seconds to attempt the evacuation alone) was male and AAV Driver was female, this task was executed 18.74% faster than when the AAV Rear Crewman was female and the AAV Driver was male.

In the cases where the female was in the Rear Crewman billet, the differences were statistically significant. The FF combination was 194.66% slower than the MM combination, and 137.68% slower than the MF combination. The FM combination was 52.56% slower than the MM combination while the FF combination was 93.14% slower than the FM combination.

J.5.2.9.2.3 External CASEVAC of an Incapacitated Turret Gunner by Critical Billet Contextual Comments

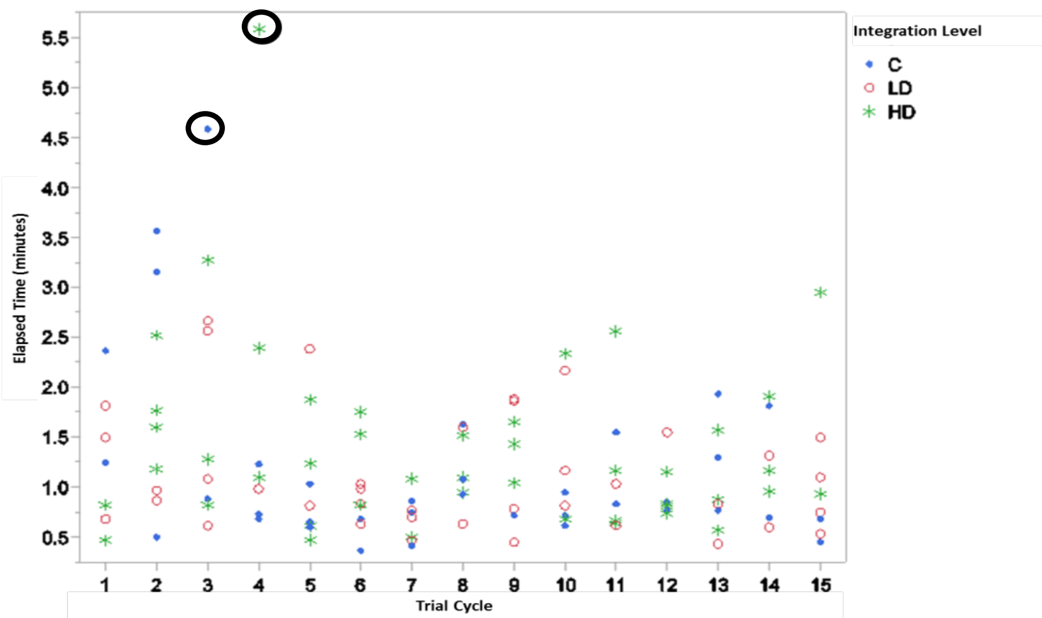
For experimental purposes, the AAV Rear Crewman participated solely in the evacuation for the first 60 seconds while the AAV Driver reported the incident to higher authority. At the 60-second mark, the AAV Driver was allowed to assist in the CASEVAC. The third crewmember (the Turret Gunner) could not assist until the 6-

minute mark. At 6 minutes, the Turret Gunner simulated an AAV crewman coming from a nearby vehicle crew to assist the Rear Crewman and Driver.

J.5.2.9.3 Internal CASEVAC of an Incapacitated Turret Gunner by Integration Level Scatterplots

The scatter plots below display the results to complete the Internal CASEVAC of an Incapacitated Turret Gunner task by integration level. On Trial Cycle days 3 and 4, two data points were excluded from analysis due to data entry errors. The data points are contained inside dark black circles on the scatterplot below.

Figure J-12. Internal CASEVAC of an Incapacitated Turret Gunner by Integration Level



J.5.2.9.4 Internal CASEVAC of an Incapacitated Turret Gunner by Integration Level Data Table and Analysis

The tables below summarize the results of the task, Internal CASEVAC by Integration Level. The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-24. Internal CASEVAC of an Incapacitated Turret Gunner by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Internal CASEVAC from Turret (minutes)	C	36	1.08	0.72
	LD	41	1.12	0.59
	HD	45	1.31	0.68

Table J-25. Internal CASEVAC of an Incapacitated Turret Gunner by Integration Level ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Internal CASEVAC from Turret (minutes)	1.48 (2, 119)	0.23	LD-C	0.04	3.64%	0.96	-0.22	0.30	-0.27	0.35
			HD-C	0.23	21.46%	0.26	-0.02	0.49	-0.07	0.54
			HD-LD	0.19	17.19%	0.37	-0.05	0.44	-0.10	0.49

**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

For this task, samples sizes are sufficiently large (n>30) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

The C group performed the task faster than the LD (2 seconds faster) and HD groups (17 seconds faster); these differences in mean times and percentage differences are not statistically significant.

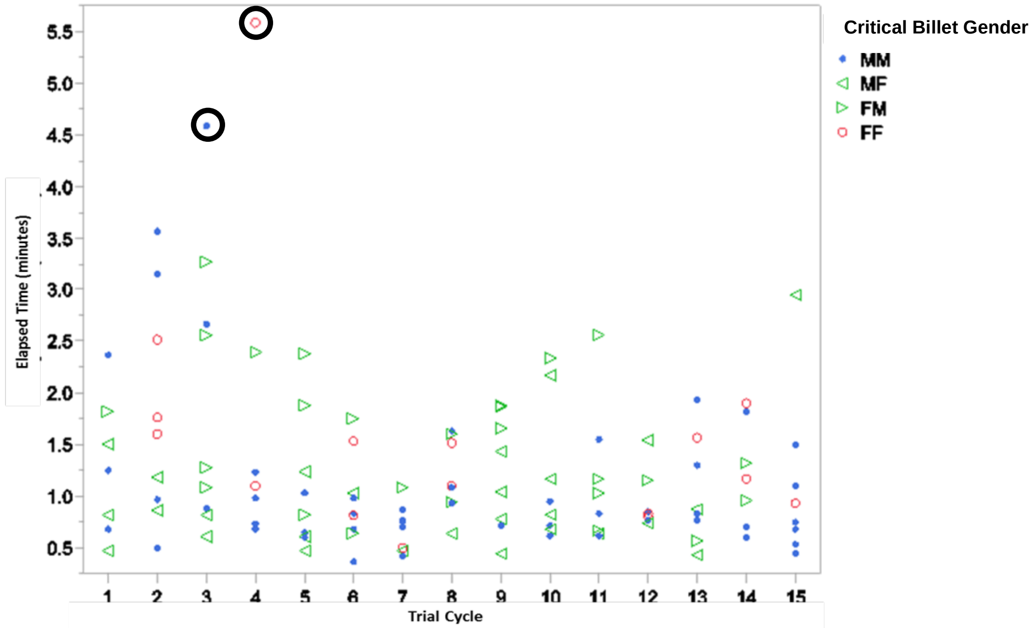
J.5.2.9.4.1 Internal CASEVAC of an Incapacitated Turret Gunner by Integration Level Contextual Comments

Operational relevance is negligible because of the small differences.

J.5.2.9.5 Internal CASEVAC of an Incapacitated Turret Gunner by Critical Billet

The scatter plots below display the results to complete the Internal CASEVAC of an Incapacitated Turret Gunner task by critical billet. On Trial Cycle days 3 and 4, two data points were excluded from analysis due to data entry errors. The data points are contained inside dark black circles on the scatterplot below

Figure J-13. Internal CASEVAC of an Incapacitated Turret Gunner by Critical Billet



J.5.2.9.6 Internal CASEVAC of an Incapacitated Turret Gunner by Critical Billet Data Table and Analysis

The tables below summarize the results of the task, Internal CASEVAC by Critical Billet. The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-26. Internal CASEVAC of an Incapacitated Turret Gunner by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Internal CASEVAC from Turret (Rear Crewman/Driver) (minutes)*	MM	51	1.05	0.66
	MF	28	0.98	0.56
	FM	27	1.54	0.71
	FF	16	1.33	0.52

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

Table J-27. Internal CASEVAC of an Incapacitated Turret Gunner by Critical Billet ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Internal CASEVAC from Turret (Rear Crewman/Driver) (minutes)	4.88 (3, 118)	< 0.01*	MF-MM	-0.07	-6.37%	0.97	-0.36	0.23	-0.41	0.28
			FM-MM	0.49	47.10%	< 0.01*	0.19	0.79	0.14	0.84
			FF-MM	0.28	27.20%	0.40	-0.08	0.64	-0.14	0.70
			MF-FM	-0.56	-36.35%	< 0.01*	-0.90	-0.22	-0.95	-0.16
			FF-MF	0.35	35.85%	0.29	-0.04	0.74	-0.11	0.81
			FF-FM	-0.21	-13.53%	0.73	-0.60	0.19	-0.67	0.25

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

For this task, the FM and FF groups are normally distributed with Shapiro-Wilk Test p-values of 0.12 and 0.57, respectively, while the MM and MF groups are not normally distributed with p-values less than 0.01. We proceed with presenting ANOVA results because they were confirmed by a Kruskal-Wallis test with a p-value of less than 0.01. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA.

Analysis shows the gender of the Marine in the first critical billet, the AAV Rear Crewman, matters. When the AAV Rear Crewman (who had 60 seconds to attempt the evacuation alone) was male and AAV Driver was female, this task was executed 36.35% faster than when the AAV Rear Crewman was female and the AAV Driver was male. See tables above for more detailed analytical results.

J.5.2.9.6.1 Internal CASEVAC of an Incapacitated Turret Gunner by Critical Billet Contextual Comments

For experimental purposes, the AAV Rear Crewman participated solely in the evacuation for the first 60 seconds while the AAV Driver reported the incident to higher authority. At the 60-second mark, the AAV Driver was allowed to assist in the CASEVAC. The third crewmember (the Turret Gunner) could not assist until the 6-minute mark. At 6 minutes, the Turret Gunner simulated an AAV crewman coming from a nearby vehicle crew to assist the Rear Crewman and Driver. The Internal CASEVAC is a much less physically demanding task when compared to the External CASEVAC.

J.5.2.10 Conduct Water Evacuation of an Incapacitated Rear Crewman

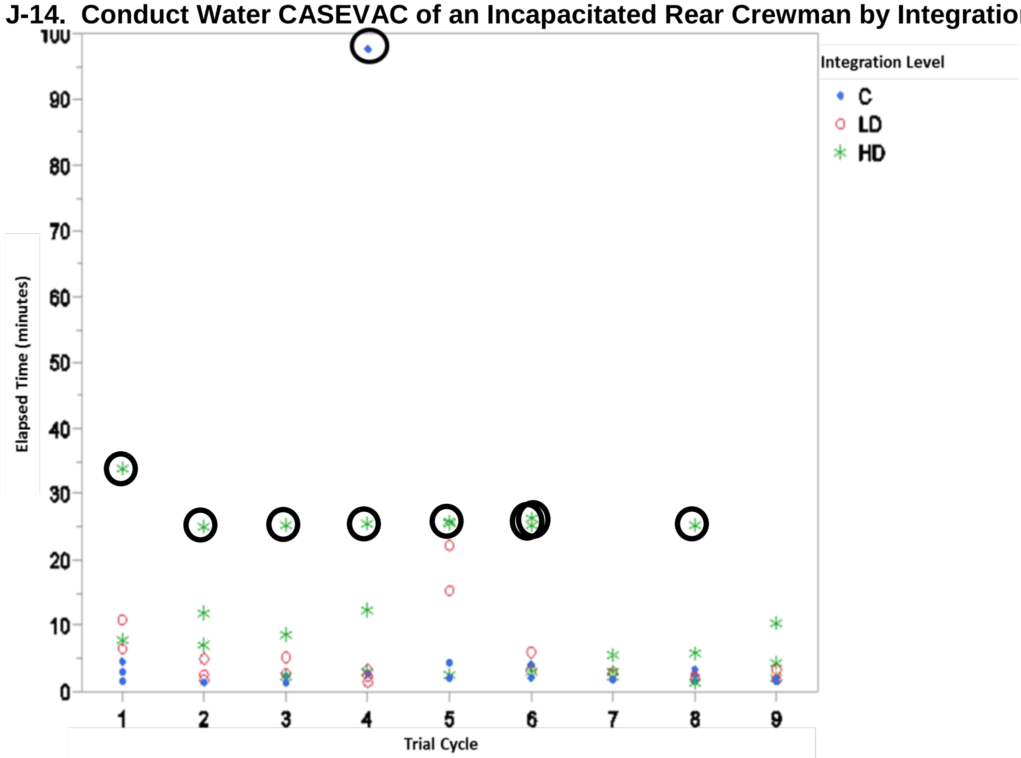
In this task, the AAV Turret Gunner and Driver lifted a simulated casualty from the floor of the troop compartment to the top of the vehicle through the starboard right cargo hatch, which is a T&R requirement for an 1833 AAV Crewman. The casualty was simulated by a dummy weighing approximately 205 lbs. The Turret Gunner and Driver had an experimental time limit of 25 minutes to accomplish this task which simulated conducting a CASEVAC from the troop compartment of the vehicle. The Data Collector, located in either the Troop Commander's seat or in the Troop Compartment of the AAV, initiated the task and recorded the start and stop time. The task was performed as a crew event, with the primary measure of performance being the time required to complete the task.

The data collected from this task was analyzed two ways. The first comparison looked for performance differences between the control group and the low-density group and the high-density group. The second comparison looked for performance differences between crews grouped by the gender of the two Marines serving in the critical billets and actively participating in the casualty extrication -- the AAV Driver and the AAV Turret Gunner.

J.5.2.10.1 Conduct Water CASEVAC of an Incapacitated Rear Crewman by Integration Level Scatterplots

Figure J-14 shows a scatter plot displaying the results to complete the Water CASEVAC of an Incapacitated Rear Crewman by Integration Level task. On trial cycle 4, a data point was removed from analysis due to a data entry error. On Experiment Days 1 through 6, and 8, a total of eight HD-group data points were removed from analysis as data entry errors.

Figure J-14. Conduct Water CASEVAC of an Incapacitated Rear Crewman by Integration Level



J.5.2.10.1.1 Conduct a Water CASEVAC of an Incapacitated Rear Crewman by Integration Level Data Table and Analysis

The tables below summarize the results of the task Water CASEVAC by Integration Level). The first table compares means across metrics and integration levels. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-28. Conduct Water CASEVAC of an Incapacitated Rear Crewman by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Conduct Water CASEVAC (minutes)*	C	21	2.38	0.97
	LD	23	4.75	5.03
	HD	19	6.48	5.63

*Indicates there is a statistically significant difference in a two-sided hypothesis test between the metric's mean values for the Integration Level.

Table J-29. Conduct Water CASEVAC of an Incapacitated Rear Crewman by Integration Level ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-Side P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Conduct Water CASEVAC (minutes)	6.96 (2, 28.1)	< 0.01††	LD-C	2.37	99.23%	0.04	0.96	3.78	0.53	4.20
			HD-C	4.09	171.70%	0.01*	2.36	5.83	1.83	6.36
			HD-LD	1.73	36.38%	0.31	-0.44	3.90	-1.08	4.54

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances)

For this task, the C, LD, and HD groups are all not normally distributed with Shapiro-Wilk Test p-values of less than 0.01. A Kruskal-Wallis test does confirm the results of the standard ANOVA. However, the standard deviation of the LD and HD groups are more than twice that of the C group, thus we recommend using the results of the robust ANOVA presented in the table above.

The control group executed the recovery with a mean time of 2.38 minutes; the LD group executed this task 99.23% slower (2:22 slower) than the C group and the HD group executed this task 171.70% slower (4:06 slower) than the C group. Although only the HD group's time (4:06 slower) is statistically significant, the time differences based on integration may be operationally significant. The LD group was 99.23% slower than the C group, while the HD group was 171.7% slower than the C group.

J.5.2.10.1.2 Conduct a Water CASEVAC of an Incapacitated Rear Crewman by Integration Level Contextual Comments

The ability of an AAV crew to extricate a casualty from the troop compartment quickly could mean the difference between life and death for an injured or incapacitated Marine. In the case of this task, seconds could be the determining factor, especially if the AAV is sinking or suffering from an engine or vehicle fire. Unlike similar vehicle MOSs, the AAV T&R does not stipulate a specific time requirement to complete this task for it to be considered a display of mastery or non-mastery. The maximum time of 25 minutes to complete was based on the goal of capturing all data points that could exist above a shorter time range. These data are available to help the AAV community determine an acceptable time with which to standardize the Assault Amphibious T&R Manual.

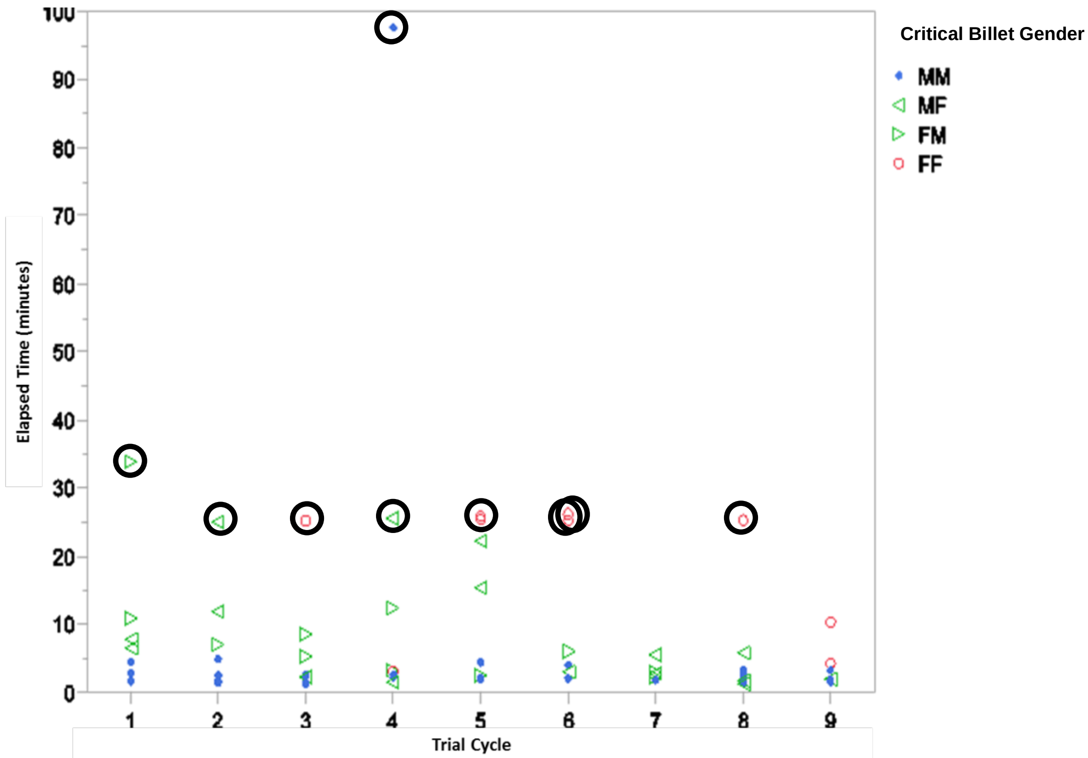
J.5.2.10.1.3 Conduct a Water CASEVAC of an Incapacitated Rear Crewman by Integration Level Additional Insights

The eight HD group data points that were excluded from the analysis as depicted on the scatterplot above (See section J.5.2.10.1) were valid data points from the perspective that each HD group failed to complete the task in the allotted time of 25 minutes. The Data Collector was following the experimental plan which asked DC to record a unit failure and stop the time once the crew reached the 25 minute mark. Unfortunately, we are not able to include the elapsed times for all these 8 data points because we cannot conclude when the group would have completed the task if the DC would have allowed the crew to continue. Table J-28 shows the HD crew with 19 complete trials vice 27, and a mean time of 6.48 minutes. If the excluded eight data points would have been accounted for in the analysis, the mean time for the HD group would have increased dramatically thereby producing a higher percentage difference between C and HD groups.

J.5.2.10.2 Conduct Water CASEVAC of an Incapacitated Rear Crewman by Critical Billet Scatterplots

The scatter plots below display the results to complete the Conduct Water CASEVAC of an Incapacitated Rear Crewman by Critical Billet task. On trial cycle 4, a data point was removed from analysis due to a data entry error. On Experiment Days 1 through 6, and 8, a total of eight HD-group data points were removed from analysis as data entry errors.

Figure J-15. Conduct Water CASEVAC of an Incapacitated Rear Crewman by Critical Billet



J.5.2.10.3 Conduct Water CASEVAC of an Incapacitated Rear Crewman by Critical Billet Data Tables and Analysis

The tables below summarize the results of the task, Water CASEVAC by Critical Billet. The first table compares means across metrics and critical billets. The second table presents ANOVA and other test results bringing to focus those metrics that resulted in statistical significance along with their percentage differences.

Table J-31. Conduct Water CASEVAC of an Incapacitated Rear Crewman by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD
Conduct Water CASEVAC (Driver/Turret Gunner (minutes)*	MM	31	2.43	0.97
	MF	17	7.08	7.34
	FM	12	5.74	3.35
	FF	3	5.96	3.88

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

Table J-30. Conduct Water CASEVAC of an Incapacitated Rear Crewman by Critical Billet ANOVA and other Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	2-Sided P-Value	80 % LCB	80% UCB	90% LCB	90% UCB
Conduct Water CASEVAC (Driver/Turret Gunner) (minutes)	5.11 (3, 59)	0.02††	MF-MM	4.65	191.60%	0.02	2.26	7.04	1.53	7.77
			FM-MM	3.32	136.54%	0.01*	1.98	4.65	1.56	5.07
			FF-MM	3.53	145.47%	0.25	-0.68	7.74	-2.97	10.03
			MF-FM	1.34	23.27%	0.52	-1.33	4.01	-2.13	4.81
			FF-MF	-1.12	-15.82%	0.71	-5.33	3.09	-6.86	4.62
			FF-FM	0.22	3.77%	0.94	-3.85	4.29	-5.69	6.13

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.
††Indicates statistical significance using a robust ANOVA (accounts for unequal variances

For this task, the FM and FF groups are normally distributed with Shapiro-Wilk Test p-value of 0.08 and 0.30, respectively, while the MM and FM groups are not normally distributed with p-values of less than 0.01. A Kruskal-Wallis test (p-value less than 0.01) does confirm the results of the standard ANOVA. However, the standard deviation of the MF, FM, and FF groups are more than twice that of the MM group, thus we recommend using the results of the robust ANOVA presented in the table above.

There is a significant difference in the mean times of the AAV crews' execution of this task in this method of comparison with the MF, FM, and FF critical billet combinations being slower by 191.6% (4:39) slower, 136.54% (3:21) slower, and 145.47% (3:31) slower, respectively, than the MM critical billet combination. Although only the MF combination showed it was statistically significant (4:39 slower), the differences in the ability of an integrated crew to conduct a water CASEVAC could be considered as a realistic and significant operational impact.

J.5.2.10.3.1 Conduct Water CASEVAC of an Incapacitated Rear Crewman by Critical Billet Contextual Comments

The analysis examines the task performed by critical billet: the AAV Driver and the AAV Turret Gunner. In this scenario, the Rear Crewman was incapacitated and the task accomplished by the two critical crewman billets. The order of gender by those two billets was not specifically important to the analysis, as the task is not scripted and the AAV Driver and the AAV Turret Gunner began the task at the exact same time, working as a team.

The ability of an AAV crew to extricate a casualty from the troop compartment quickly could mean the difference between life and death for an injured or incapacitated Marine. In the case of this task, seconds could be the determining factor, especially if the AAV is sinking or suffering from an engine or vehicle fire. Unlike similar vehicle MOSs, the AAV T&R does not stipulate a specific time requirement to complete this task for it to be considered a display of mastery or non-mastery. The maximum time of 25 minutes to complete was based on the goal of capturing all data points that could exist above a shorter time range. These data are available to help the AAV community determine an acceptable time with which to standardize the Assault Amphibious T&R Manual.

J.6 Statistical Modeling Results

J.6.1 Statistical Modeling Results Overview

The previous section discussed results only as they pertain to differences due to integration level and critical billets. The goal of statistical modeling as applied here is to estimate, simultaneously, the effect of gender integration levels and other relevant variables on the AAV crew performance. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For the selected tasks described in the previous section, this section presents an overview of the analysis and results, and then presents the modeling results for each task.

For each task, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates an increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time.

J.6.2 Selected Tasks Method of Analysis

A mixed effects model with all AAV crew members and all types of personnel variables was run for each task. For each result, we create a model using integration level and all types of personnel data for each crew member with a random effect for who filled each position within the crew. For example, age, height, and other variables for each member of the crew and integration level are modeled with the result (response time) as the response variable. For tasks that had critical billets, i.e. only one or two Marines were performing the task without help from the others, only their variables were included in the model, along with the genders of Marines in the critical billets. Where maximum likelihood estimation converged, AIC was used for variable selection. Otherwise, we comment on the significance of individual variables in the full model. Variables reported as significant are concluded to be significant based on at least a one-sided test.

J.6.3 Selected Tasks Overall Modeling Results

There are no personnel data variables that are both statistically significant and have a practical impact to the model. Each time personnel data variables are statistically significant in a model, their effects are practically negligible, conflicting, and/or incomplete for the AAV crew; i.e., there are no tasks for which a variable is significant for all, or even most members of the AAV crew.

Integration level, however, consistently appears as statistically significant in each of the tasks, and its effect is clear, causal, and practical. Therefore, integration level is the best variable to describe performance for each of these tasks. Refer to the Descriptive Statistics Section for the ANOVA summary for each of the below mentioned tasks.

J.6.3.1 External CASEVAC from Turret

We model elapsed time for the external CASEVAC as a function of the personnel variables and gender of the Marines performing the task in a mixed effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the AAV crew. We report statistically significant positive and negative correlations.

The results from the final model for External CASEVAC from Turret show that relative to two males performing the task, a Male in the Rear Crewman Billet and a Female in the Turret Gunner Billet, perform the task 0.39 minutes slower, and the difference is not statistically significant. If their billets are switched, the difference increases to 0.47 minutes and becomes statistically significantly slower (in a 1-sided test). However, if the task is performed by two females, then, holding all other variables constant, they perform it 2.52 minutes slower, on average, and this difference from a two-male crew is strongly significant. Although other variables in the model are observed to have a statistically significant relationship with elapsed time for this task, their impact on the time is negligible.

The models for the following variables do not run due to missing values:

- None.

The following variables are significant in the model and are positively correlated with the external CASEVAC:

- AFQT score of the Rear Crewman

The following Driver variables are significant in the model and are negatively correlated with the external CASEVAC:

- Rifle Range score of the Driver

Because the personnel variables do not have any patters and their effects are often negligible, the final model includes integration level only FF groups take 2.77 minutes longer than MM ($p < 0.01$), FM groups take 0.75 minutes longer than MM ($p < 0.01$), and MF groups take .34 minutes longer than MM (one-sided p-value = 0.08) . All these differences are statistically significant.

J.6.3.2 Interior Reload 1

We model elapsed time for the Interior Reload # 1 as a function of the personnel variables and gender of the Marines performing the task in a mixed effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the AAV crew. We report statistically significant positive and negative correlations.

For the task Interior Reload 1, gender of Marines in critical billets was not chosen by the AIC as one of the variables that helps predict elapsed times.

The models for the following variables do not run due to missing values:

- None.

The following variables are significant in the model and are positively correlated with the first interior reload:

- CFT MTC score for the Turret Gunner

The following variables are significant in the model and are negatively correlated with the first interior reload:

- Height of the Rear Crewman

There are no patterns for any variables the first interior reload. See the 1833 Descriptive Statistics section for the ANOVA summary of this task.

J.6.3.3 Interior Reload 2

We model elapsed time for the Interior Reload # 2 as a function of the personnel variables and gender of the Marines performing the task in a mixed effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the AAV crew. We report statistically significant positive and negative correlations.

For the task Interior Reload 2, gender of Marines in critical billets was not chosen by the AIC as one of the variables that helps predict elapsed times.

The models for the following variables do not run due to missing values:

- None.

The following variables are significant in the model and are positively correlated with the second internal reload:

- None.

The following variables are significant in the model and are negatively correlated with the second internal reload:

- Rifle score for Rear Crewman.

There are no patterns for any variables for the second interior reload. See the 1833 Descriptive Statistics section for the ANOVA summary of this task.

J.6.3.4 External Reload

We model elapsed time for the External Reload as a function of the personnel variables and gender of the Marines performing the task in a mixed effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the AAV crew. We report statistically significant positive and negative correlations.

In the External Reload task, the gender of Marines in the critical billet was eliminated from the model during variable selection using AIC. This means that the benefit of its relationship with the response variable was outweighed by the cost of including it in the model, or that it did not show a strong relationship with elapsed times once other variables were included.

The models for the following variables do not run due to missing values:

- None.

The following variables are significant in the model and are positively correlated with the external reload:

- CFT MTC time of the Rear Crewman
- PFT Crunch score of the Rear Crewman
- GT Score of the Turret Gunner.

The following variables are significant in the model and are negatively correlated with the external reload:

- Age of Rear Crewman
- Weight of Rear Crewman
- CFT MANUF time of Rear Crewman
- PFT run time of Rear Crewman
- AFQT score of Turret Gunner
- Weight of Turret Gunner
- Rifle Score of Turret Gunner.

There are no patterns for any variables for the external reload, and the effects described above are often negligible in size. See the 1833 Descriptive Statistics section for the ANOVA summary of this task.

J.6.3.5 Break / Reassemble Track

We model elapsed time for the Break/Reassemble track as a function of each personnel variable for each crew member and integration level in a mixed effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the AAV crew. We report statistically significant positive and negative correlations and whether we observe any patterns.

For the task Break / Reassemble Track, the model selection process eliminated integration levels, indicating that, ceteris paribus, integration level of the crew is not a good predictor of how long it takes to perform this task.

The models for the following variables do not run due to missing values:

- None.

The following variables are significant in the model and are positively correlated with the break and reassemble track:

- CFT MTC Score of the Rear Crewman
- PFT Run Score of the Turret Gunner.

The following variables are significant in the model and are negatively correlated with break and reassemble track:

- Age of the Rear Crewman
- Height of the Rear Crewman
- CFT MANUF score of the Rear Crewman
- Weight of the Turret Gunner
- CFT MANUF score of the Turret Gunner
- PFT Run time of the Turret Gunner.

There are no patterns for any variables for breaking and reassembling track, and the effects described above are often negligible in size. See the 1833 Descriptive Statistics section for the ANOVA summary of this task.

J.6.3.6 External Mount of M2 & Mk19 and Ammo

We model elapsed time for the External Mount of M2 and Mk19 as a function of each personnel variable for each crew member and integration level in a mixed effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the AAV crew. We report statistically significant positive and negative correlations and whether we observe any patterns.

For the task External Mount of M2 & Mk19 and Ammo, the model selection process eliminated integration levels, indicating that, ceteris paribus, integration level of the crew is not a good predictor of how long it takes to perform this task.

The models for the following variables do not run due to missing values:

- None.

The following variables are significant in the model and are positively correlated with the external mount of weapons and ammunition:

- CFT MTC time of the Turret Gunner
- PFT Crunch Score of the Turret Gunner.

The following variables are significant in the model and are negatively correlated with the external mount of weapons and ammunition:

- CFT MTC Score of the Driver
- Rifle Range score of the Driver
- Age of the Turret Gunner
- AFQT of Turret Gunner.

J.6.3.7 Internal CASEVAC from Turret

We model elapsed time for the Internal Casevac from Turret as a function of the personnel variables and gender of the Marines performing the task in a mixed effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the AAV crew. We report statistically significant positive and negative correlations.

Genders of Marines in critical billets were not selected by AIC into the final model of elapsed time for Internal CASEVAC from Turret, indicating that no difference was observed between genders for this task.

The models for the following variables do not run due to missing values:

- None.

The following variables are significant in the model and are positively correlated with the internal CASEVAC:

- GT of the Driver
- CFT MANUF time of the Driver.
- Rifle Range Score of the Driver
- CFT MTC time of the Rear Crewman
- Rifle Range Score of the Rear Crewman.

The following variables are significant in the model and are negatively correlated with the internal CASEVAC:

- AFQT of the Driver
- Weight for the Rear Crewman
- GT of the Turret Gunner
- Height of the Turret Gunner
- CFT MTC score of the Turret Gunner.

There are no patterns for any variables for the internal CASEVAC, the effects of most variables above are negligible. See the 1833 Descriptive Statistics section for the ANOVA summary of this task.

J.6.3.8 Water CASEVAC

We model elapsed time for the water casevac as a function of the personnel variables and gender of the Marines performing the task in a mixed effects model with a random effect for who filled each position. The covariates in each model are the values of each personnel variable for each volunteer member in the AAV crew. We report statistically significant positive and negative correlations.

Genders of Marines in critical billets were selected by AIC into the final model of elapsed time for Water CASEVAC. Compared to the reference group of two males performing the task, a male in the Rear Crewman billet and a female in the Turret Gunner billet conducted the task 4.9 minutes slower. This difference is statistically significant with a two-sided p-value of 0.06. Other gender compositions of critical billets did not show significant differences from the all-male control group.

The models for the following variables do not run due to missing values:

- None.

The following variables are significant in the model and are positively correlated with the water CASEVAC:

- AFQT of the Driver
- PFT Run time of the Driver
- CFT MTC Time of the Driver.

The following variables are significant in the model and are negatively correlated with the water CASEVAC:

- Weight of the Driver
- Rifle Range Score of the Driver
- CFT MANUF time of the Turret Gunner.

Because there are no patterns in the significance of the above variables and because their effects are often negligible, our final model includes only the genders of Marines in critical billets and the random effects for which volunteer fills each billet. In this model, it takes two female Marines 3.45 minutes longer than two male Marines to perform the task and the difference is on the border of statistical significance (two-sided p-value of 0.21), while a female Driver and male Turret Gunner take 3.11 minutes longer ($p = 0.06$) and a male Driver and female Turret Gunner take 4.28 minutes longer ($p < 0.01$) than two males. The last two differences are statistically significant.

Appendix to Annex J
1833 Supplemental Information

This appendix provides supplemental information for the 1833 portion of the GCE ITF experiment. It provides information regarding the additional descriptive and basic inferential statistics not described in Annex J.

Section 1: Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for additional 1833 tasks. Annex J contains the descriptive statistics for the remainder of the 1833 tasks. The words “metric” and “task” are used interchangeably throughout this Appendix. They both refer to the experimental task.

The tables in this appendix display results for the additional 1833 metrics, to include metrics and integration levels with their respective sample sizes, means, standard deviations, and percent difference between integration levels, and ANOVA results, including metrics and integration levels, and p-values suggesting statistical significance. For each task, an ANOVA and t-test were conducted to compare groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude there is statistical evidence that the response for the LD and HD groups are different from that in the C group.

Section 2: Additional Task Results

Conduct a modified Gun Table V Overview. The AAV crews engaged multiple targets from 400 to 1500 meters from multiple defensive and offensive positions. These distances and exposure times were dictated by the US Army HBCT Manual. Marines conducted four separate engagements: one offensive and three defensive, and engaged 12 vehicle or troop targets.

A modified Gun Table V was used to evaluate a vehicle crew’s ability to engage similar targets to those used during semi-annual crew and section gunnery qualifications. Employing both weapons systems in an AAV turret is a physically strenuous task for the Turret Gunner, especially in the case when immediate or remedial action is required. Although the other crew members are not directly involved in the employment of the Up-Gunned Weapons Station (UGWS), the driver’s performance spotting targets and maneuvering the vehicle to allow the turret gunner to safely engage those targets can have a significant effect on the Turret Gunner’s ability to identify and engage targets. Time for this task started when targets were presented and stopped when the Turret

Gunner placed effects on target or the target went down for time per the HBCT. The Master Gunner executed the range and modified gun table. A Data Collector was positioned alongside the Master Gunner to record target exposure times. The start time for each target presentation began when the target was in its exposed locked position. The Master Gunner identified when the target was in its locked positon and when it was down due to a hit or down for time. The stop time for each target engagement was when the Master Gunner assessed weapons effects on the presented target or that target went down for time based on the HBCT Unstabilized Platform Gunnery Tables. Subsequent paragraphs examine total engagement times for each type of engagement. Total engagement time for defensive engagements is the sum of 10 target exposure times; total engagement time for offensive engagements is the sum of 2 target exposure times. The data collected from this task was analyzed in two separate ways. The first method of comparison looked for the performance differences between the C, LD, and HD groups. The second method of comparison looked for performance differences between crews grouped by the gender of the Marine serving in the critical billet as the AAV Turret Gunner.

Fire Weapons – Defensive by Integration Level Data. Table J A summarizes the results of the task, Fire Weapons – Defensive by Integration Level.

Table J A – Fire Weapons – Defensive by Integration Level

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Fire Weapons - Defensive	C	35	7.14	1.40	4.64%	8.74%	3.92%
	LD	42	7.48	1.62			
	HD	42	7.77	1.48			

Table J A shows a difference of 20.4 seconds between the mean time for the C group's 7.14 minutes and the LD group's 7.48 minutes over 10 separate target exposures is only a 2.04-second difference per target. Considering all targets had a 70- to 78-second exposure time (vehicle targets had a 78-second exposure time to account for the 8 seconds it took to lift the target), the 2.04-second difference is negligible. Both tasks, "Fire Weapons – Defensive" and "Fire Weapons – Offensive", show no statistical significance per the ANOVA and the Tukey test.

Fire Weapons – Offensive by Integration Level. Table J B summarizes the results of the task, Fire Weapons – Offensive by Integration Level.

Table J B - Fire Weapons – Offensive by Integration Level

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Fire Weapons - Offensive	C	37	1.69	0.45	0.69%	8.42%	7.67%
	LD	42	1.7	0.43			
	HD	45	1.83	0.37			

Table J B shows the LD group was 0.69% slower than the C group. The HD group was 8.42% slower than the C group. The HD group was 7.67% slower than the LD group.

Fire Weapons – Defensive by Critical Billet - Turret Gunner. Table J C summarizes the results of the task, Fire Weapons – Defensive by Critical Billet.

Table J C - AAV Fire Weapons – Defensive (Turret Gunner) by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)
Fire Weapons - Defensive	M	78	7.17	1.4	12.75%
	F	41	8.08	1.57	

Table J C shows that female (F) turret gunners were 12.75% slower.

The analytical results for this task by critical billet, in this case the Turret Gunner, show female turret gunners (F) performed slower than their male counterparts (M).

- **Contextual Comments.** Providing Defensive Fires from an AAV is a combat task where time is most critical. Faster engagement times (even those measured in fractions of a second) during combat operations are highly desirable. Per the MCWP 3-13, “The side that brings effective fire to bear on the enemy first has a significant advantage. Firing first with accurate fire greatly increases chances of winning the engagement. If the initial burst misses, the first to fire can probably shoot again and hit the target before receiving return fire. If surprised by the enemy and unable to fire first, the AA unit should return fire as quickly as possible. The AA unit should continue to engage the enemy as fast as possible, because fire placed in the enemy’s area will lessen his effectiveness and give friendly weapons time to adjust. A wasted opportunity to engage a target may never be regained.” The analytical results show statistical significance when gender is taken into account for the critical billet in the performance of Defensive Fires, but it did not present results that were statistically significant for the integration level analysis, i.e., C, LD, or HD groups.

- **Additional Insights**. The difference in performance between the M and F groups may be a result of disparity in experience levels when employing the weapons station vice a physical disparity, as mentioned in the Experiment Limitations. The male volunteers were primarily Marines with several years' experience in the 1833 MOS, while the female volunteers were entry-level Marines based on their recent graduation from MOS school, as referenced in the AAV Experiment population description.

AAV Fire Weapons – Offensive (Turret Gunner) by Critical Billet. Table J D summarizes the results of the task, Fire Weapons – Offensive (Analysis by Critical Billet).

Table J D - AAV Fire Weapons – Offensive (Turret Gunner) by Critical Billet

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)
Fire Weapons - Offensive (Turret Gunner) (minutes)	M	81	1.66	0.43	14.29%
	F	43	1.9	0.36	

Table J D shows that female turret gunners were 14.29% slower on average; this difference is statistically significant in a one-sided hypothesis test and two-sided test.

- **Contextual Comments.** When gender is taken into account for the critical billet (Turret Gunner), the results show statistical significance. When analyzed by critical billet, results show female turret gunners (F) performed slower than their male counterparts (M). For offensive engagements where two targets were presented with 70- to 78-second exposure times, the female turret gunners achieved effects on targets 7.2 seconds slower than male turret gunners.
- **Additional Insights**
    - Faster engagement times (even those measured in fractions of a second) are highly desirable during combat operations. Per the MCWP 3-13, "The side that brings effective fire to bear on the enemy first has a significant advantage. Firing first with accurate fire greatly increases chances of winning the engagement. If the initial burst misses, the first to fire can probably shoot again and hit the target before receiving return fire. If surprised by the enemy and unable to fire first, the AA unit should return fire as quickly as possible. The AA unit should continue to engage the enemy as fast as possible, because fire placed in the enemy's area will

lessen his effectiveness and give friendly weapons time to adjust. A wasted opportunity to engage a target may never be regained."

- o The difference in performance between the M and F groups may be a result of disparity in experience levels when employing the weapons station vice a physical disparity, as mentioned in the Experiment Limitations. The male volunteers were primarily Marines with several years' experience in the 1833 MOS, while the female volunteers were entry-level Marines based on their recent graduation from MOS school, as referenced in the AAV Experiment population description.

Conduct a Manual Ramp Raise. In this Maintenance Action subtask, the AAV crew manually raised the AAV ramp with a ramp jack. The crews did not have responsibilities dictated for this task and were allowed to conduct the task as they saw fit until completion. Upon completion, an AAV Platoon Staff member inspected the ramp to ensure it had been properly secured. There was no maximum time limit for this task. The Data Collector stood to the rear of the vehicle and was the initiator of this task by verbal command and recorded start and stop times.

The task was performed as a crew event with the primary measure of performance being the time required to complete the task. The data collected from this task was analyzed by crew integration only, and not by critical billet, as each Marine in the AAV Crew had an active role in the execution.

Table J E - Conduct a Manual Ramp Raise by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Manual Raise of Ramp*	C	38	2.57	0.56
	LD	42	2.77	0.74
	HD	46	3.28	0.91

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

Table J F - Conduct a Manual Ramp Raise by Integration Level AAV ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value (one-sided)	80 % LCB	80% UCB	90% LCB	90% UCB
Manual Raise of Ramp (minutes)	9.95 (2, 123)	< 0.01*	LD-C	0.19	7.56%	0.49	-0.10	0.49	-0.16	0.55
			HD-C	0.71	27.70%	< 0.01*	0.42	1.00	0.37	1.06
			HD-LD	0.52	18.73%	< 0.01*	0.24	0.80	0.18	0.86

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

For this task, group sample sizes are sufficiently large (n ≥ 30) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. Table J-E and Table J-F show the LD group was 7.56% slower than the C group; this difference is not statistically significant in a Tukey test. The HD group was 27.7% slower than the C group; this difference is statistically significant in a Tukey test. The HD group was 18.73% slower than the LD group; this difference up is statistically significant in a Tukey test.

- **Contextual Comments**. Analysis looked for performance differences between the control group and the low-density group and the high-density group. The C group had the fastest mean time for the manual raising of the ramp (2.57 minutes), while the HD recorded an 18% to 27.7% difference when compared to the LD and C groups, respectively. This difference equates to slower times, and in the case of the HD groups, on average, a time 43 seconds slower than the C group. Although statistical significance exists from the analytical results, the small differences in times and percentages of raising an AAV Ramp can be considered operationally irrelevant.

Load Weapons and Ammunition Overview. The vehicle crew loaded the M2 .50-caliber machine gun, the MK-19 40-mm grenade launcher, and a full complement of ammunition (two cans of .50-caliber ammunition and three cans of 40-mm grenades) from the lowered ramp of the AAV. The crew executed this task from the interior of the vehicle. A fully assembled M2 weighs approximately 85 lb and the MK-19 weighs approximately 79 lb. Upon completion, an AAV Platoon Staff member inspected to ensure the weapons were installed correctly. The Data Collector was located on the near the Troop Compartment of the AAV, and initiated this task with a verbal start time, recording start and stop times. There was no maximum time limit for this task. This task simulated the AAV crew loading weapons and ammunition from various positions on the vehicle. The task was performed as a crew event, with the primary measure of performance being the time required to complete the task. The data collected from this task was analyzed by crew integration only, and not by critical billet, as each Marine in the AAV Crew had an active role in execution.

Internal Mount of M2 by Integration Level. Table J G summarizes the results of the task, Internal Mount of M2 by integration level.

Table J G - Internal Mount of M2 by Integration Level

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Int Mount of M2*	C	38	2.33	1.82	26.83%	40.48%	10.76%
	LD	41	2.96	1.61			
	HD	46	3.28	1.56			

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.

Table J G shows the LD group was 26.83% slower than the C group; this difference is not statistically significant in a Tukey test. The HD group was 40.48% slower than the C group; this difference is statistically significant in a Tukey test. The HD group was 10.76% slower than the LD group; this difference is not statistically significant in a Tukey test.

- **Contextual Comments**. In the analysis, the C group had a mean time of 2.33 minutes, which was 26% faster when compared to the LD group (38 seconds faster), and over 28% faster when compared to the HD group (57 seconds faster). Unless in an extreme situation (in combat or in training), a difference of approximately 1 minute in the internal mounting of the M-2 is operationally irrelevant.

Conduct Manual Turret Manipulation. In this subtask, the AAV Turret Gunner, without assistance, manually traversed, elevated, and depressed the vehicle turret, using the manual traverse and elevation controls within the turret. The Turret Gunner manually elevated to maximum, depressed to maximum, and then traversed the turret 360 degrees. A Data Collector captured the elapsed time for this task; there was no maximum time limit. This task simulated the Turret Gunner manually manipulating the turret to scan and engage targets. The task was performed as an individual event, with the primary measure of performance being the time required to complete the task.

The data collected from this task was analyzed two ways. The first comparison looked for performance differences between the integration level groups. The second comparison looked for performance differences between crews grouped only by the gender of the Marines serving in the critical billet (AAV Turret Gunner) and actively participating in manual manipulation of the Up-Gunned Weapon Station (UGWS).

Conduct Manual Turret Manipulation by Integration Level. The tables below summarize the results of the task, Manual Turret Manipulation by Integration Level.

Table J H - Conduct Manual Turret Manipulation by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Manual Manipulation of Gun (minutes)	C	38	1.82	0.24
	LD	42	1.96	0.38
	HD	45	2.13	0.42

Table J I - Conduct Manual Turret Manipulation by Integration Level

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value (one-sided)	80 % LCB	80% UCB	90% LCB	90% UCB
Manual Manipulation of Gun (minutes)	7.75 (2, 122)	< 0.01*	LD-C	0.14	7.56%	0.21	0.00	0.28	-0.03	0.31
			HD-C	0.31	17.17%	< 0.01*	0.17	0.45	0.15	0.48
			HD-LD	0.17	8.93%	0.07*	0.04	0.31	0.01	0.34

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

For this task, group sample sizes are sufficiently large (n ≥ 30) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. Table J H and Table J I show the LD group was 7.56% slower than the C group; this difference is not statistically significant in a Tukey test. The HD group was 17.17% slower than the C group; this difference is statistically significant in a Tukey test. The HD group was 8.93% slower than the LD group; this difference is statistically significant in a Tukey test. On Experiment Day 14, the HD-group data point was removed from analysis as a data entry error.

- **Contextual Comments.** In the analysis, the C group had a faster mean time of 1.82 minutes. During comparative analysis between all three integration-level groups, the C-HD comparison and the HD-LD comparison show statistically significant differences, with the HD group being 17% slower (19 seconds slower) than the C group in executing the task.

- **Additional Insights**. Manual manipulation of the UGWS to traverse and elevate the weapons systems onto targets is a required task that any Marine conducting gunnery training or participating in combat operations will execute. It requires approximately 8-10 lb of force to use the manual turret traverse mechanism. If the turret were to lose electrical traverse capability, the Turret Gunner would solely use the manual turret traverse mechanism. In training and combat, the ability to manipulate the weapons onto a target quickly is important. Executing this task in a combat environment is especially important because it increases the effectiveness and survivability of an AAV crew.

Conduct Manual Turret Manipulation by Critical Billet - AAV Turret Gunner. The tables below summarize the results of the task, Manual Turret Manipulation by Critical Billet.

Table J J - Conduct Manual Turret Manipulation by Critical Billet - AAV Turret Gunner

Metric	Integration Level	Sample Size	Mean	SD
Manual Manipulation of Gun (Turret Gunner) (minutes)	M	82	1.83	0.27
	F	43	2.26	0.41

Table J K - Conduct Manual Turret Manipulation by Critical Billet - AAV Turret Gunner ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	T-Test Statistic	P-Value (one-sided)	80 % LCB	80% UCB	90% LCB	90% UCB
Manual Manipulation of Gun (minutes)	49 (1, 123)	< 0.01*	F-M	0.43	23.34%	6.17 (61)	< 0.01* (< 0.01*)	0.34	0.52	0.31	0.55

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own

For this task, group sample sizes are sufficiently large ($n \geq 30$) to satisfy normality assumptions for ANOVA. Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. Table J J and Table J K show the female turret gunners were 23.34% slower on average; this difference is statistically significant in a one-sided hypothesis test and two-sided test.

- **Contextual Comments.** In the analysis, male Turret Gunners (M) had a mean of 1.83 minutes, while female Turret Gunners had a mean of 2.26 minutes. The differences between mean times produce a 23.34% difference (26 seconds slower) when comparing male and female Turret Gunners.
- **Additional Insights**. Manual manipulation of the UGWS to traverse and elevate the weapons systems onto targets is a required task that any Marine conducting gunnery training or participating in combat operations will execute. It requires approximately 8-10 lb of force to use the manual turret traverse mechanism. If the turret were to lose electrical traverse capability, the Turret Gunner would solely use the manual turret traverse mechanism. In training and combat, the ability to manipulate the weapons onto a target quickly is important. Executing this task in a combat environment is especially important because it increases the effectiveness and survivability of an AAV crew.

Perform Recovery Operations. In this task, the AAV crew conducted a water recovery and a land recovery of a simulated disabled AAV. The crews of the recovery and disabled AAVs worked together for each vehicle's portion of the recoveries.

The water recovery required the Turret Gunner and the Rear Crewman from each vehicle to either throw water tow ropes to the recovery vehicle or use a boat hook to secure the thrown ropes. The crews would then connect both vehicles stern-to-stern and tow the disabled vehicle in the water. The Data Collector, located in either the Troop Commander's Seat or the Troop Compartment, recorded a stop time on the commencement of the tow.

The land tow required both crews to manipulate the two vehicles and the land tow bar weighing approximately 150 lb, and then attach it to two points on the disabled vehicle. The operational vehicle backed up until the tow pintle was seated. Once the vehicle was attached, both crews loaded up into the recovery vehicle and the recovery vehicle would begin to tow. The Data Collector, located outside the AAV, captured the stop time for this task when the recovery vehicle had towed the disabled vehicle 125 ft up the beach to a designated area.

Upon completion of each tow, the crews swapped roles and executed the recovery again. There was no maximum time limit for this subtask. Water and land tow tasks are requirements outlined by the T&R Manual and are somewhat likely to occur during normal operations on land and in the water. Both tasks were performed as a crew event, with the primary measure of performance being the time required to complete the task.

The data collected from this task was analyzed by comparing the performance difference between the control group and the low-density group and the high-density group. As all three crew members of a single AAV were actively engaged in this task, there was no requirement to conduct analysis by critical billet. The water recovery data is separated into two groups, operational and disabled.

Conduct Water Recovery; Operational Vehicle. The tables below summarize the results of the task, Water Recover; Operational Vehicle by Integration Level.

Table J L- Conduct Water Recovery; Operational Vehicle by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Conduct Water Recovery; operational (minutes)	C	22	5.40	1.78
	LD	22	5.49	2.06
	HD	27	5.24	1.84

Table J M - Conduct Water Recovery; Operational by Integration Level ANOVA and Tukey

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value (one-sided)	80 % LCB	80% UCB	90% LCB	90% UCB
Conduct Water Recovery; operational (minutes)	0.12 (2, 68)	0.89	LD-C	0.09	1.70%	0.99	-0.90	1.08	-1.10	1.28
			HD-C	-0.17	-3.06%	0.95	-1.11	0.78	-1.30	0.97
			HD-LD	-0.26	-4.67%	0.88	-1.20	0.69	-1.39	0.88

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.

For the task above, the LD and HD groups are normally distributed as evidenced by Shapiro-Wilk Test p-values of 0.12 and 0.05, respectively, while the C group is not normally distributed with Shapiro-Wilk p-value of less than 0.01. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.95). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. Table J-L and Table J-M show the LD group was 1.7% slower than the C group; this difference is not statistically significant in a Tukey test. The HD group was 3.06% faster than the C group; this difference is not statistically significant in a Tukey test. The HD group was 4.67% faster than the LD group; this difference is not statistically significant in a Tukey test. On Experiment Day 1, the LD-group data point was removed from analysis as a data entry error.

- **Contextual Comments**. In the analysis, the C group had a mean time of 5.40 minutes and was 1.7% (5 seconds) faster than the LD group. The HD group was 3.06% (9 seconds) faster than the C group. Due to the means of the groups being so close, the determination could be made that the composition of an AAV Crew during the execution of this task may not have an operational impact. The land-recovery task analysis shows the C group had a mean time of 5.97 minutes and was 12.24% slower than the LD group. The HD group was 1.15% slower than the C group.

Conduct Water Recovery; Disabled Vehicle. The tables below summarize the results of the task, Water Recovery; Disabled Vehicle (Analysis by Integration Level).

Table J N - Conduct Water Recovery; Disabled Vehicle by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Conduct Water Recovery; disabled (minutes)	C	22	5.70	2.81
	LD	22	5.60	1.76
	HD	27	5.84	2.13

Table J O - Conduct Water Recovery; Disabled Vehicle by Integration Level ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value (one-sided)	80 % LCB	80% UCB	90% LCB	90% UCB
Conduct Water Recovery; disabled (minutes)	0.07 (2, 68)	0.93	LD-C	-0.11	-1.87%	0.99	-1.29	1.08	-1.53	1.32
			HD-C	0.14	2.41%	0.98	-0.99	1.27	-1.22	1.50
			HD-LD	0.24	4.36%	0.93	-0.88	1.37	-1.11	1.60

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.

For the overall time to mount and remove side panels, the C and HD groups are normally distributed as evidenced by Shapiro-Wilk Test p-values of 0.01 and 0.02, respectively, while the LD group is not normally distributed with Shapiro-Wilk p-value of less than 0.01. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.83). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. Table J-N and Table J-O show the LD group was 1.87% faster than the C group; this difference is not statistically significant in a Tukey test. The HD group was 2.41% slower than the C group; this difference is not statistically significant in a Tukey test. The HD group was 4.36% slower than the LD group; this difference is not statistically significant in a Tukey test.

- **Contextual Comments**. In the analysis, the C group had a mean time of 5.70 minutes. The LD group had a mean time of 5.60 minutes and was 6 seconds faster than the C group. The HD group had a mean time of 5.84 minutes that was 2.41% (8 seconds) slower than the C group. Due to the means of the groups being so close, the determination could be made that the composition of an AAV Crew during the execution of for this task may not have an operational impact.

Conduct Land Recovery by Integration Level. The tables below summarize the results of the task, Conduct Land Recovery by Integration Level.

Table J P - Conduct Land Recovery by Integration Level

Metric	Integration Level	Sample Size	Mean	SD
Conduct Land Recovery (minutes)	C	22	5.97	1.77
	LD	22	5.24	1.06
	HD	27	6.04	1.51

Table J Q - Conduct Land Recovery by Integration Level ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	% Difference	P-Value (one-sided)	80 % LCB	80% UCB	90% LCB	90% UCB
Conduct Land Recovery (minutes)	2.07 (2, 68)	0.13	LD-C	-0.73	-12.24%	0.24	-1.50	0.04	-1.66	0.20
			HD-C	0.07	1.15%	0.99	-0.67	0.81	-0.82	0.96
			HD-LD	0.80	15.26%	0.15	0.06	1.54	-0.09	1.69

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between the metric's mean values for the Integration Level.
**Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.

For the overall time to mount and remove side panels, the LD and HD groups are normally distributed as evidenced by Shapiro-Wilk Test p-values of 0.09 and 0.12, respectively, while the C group is not normally distributed with Shapiro-Wilk p-value of less than 0.01. We proceed with presenting ANOVA results since they are confirmed by a Kruskal-Wallis Test (p-value = 0.17). Additionally, group standard deviations are sufficiently similar to satisfy the equal variance assumption for ANOVA. Table J-P and Table J-Q show the LD group was 12.24% faster than the C group; this difference is not statistically significant. The HD group was 1.15% slower than the C group; this difference up is not statistically significant. The HD group was 15.26% slower than the LD group; this difference is not statistically significant.

Due to the means of the groups being so close, the determination could be made that the composition of an AAV Crew during the execution of for this task may not have an operational impact.

THIS PAGE INTENTIONALLY LEFT BLANK

Annex K.
Combat Engineer (MOS 1371)

This annex details the Combat Engineer (MOS 1371) portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment executed 2 March – 26 April 2015 at Ranges 107, 108, 110, and 114 aboard the Marine Corps Air-Ground Combat Center (MCAGCC), Twentynine Palms, CA. The sections outline the Combat Engineer Scheme of Maneuver (SOM), Limitations, Deviations, Descriptive Statistics, and Modeling Results.

K.1 Scheme of Maneuver

K.1.1 Experimental Cycle Overview

The Combat Engineer assessment of the GCEITF took place in a field environment aboard MCAGCC, Twentynine Palms, CA. The assessment consisted of 21 trial cycles, each of which was a 2-day test cycle, conducted over the course of 55 days. After every 4 days of trials, the Marines received 1 recovery day at Camp Wilson. Every squad consisted of eight volunteers and a direct-assignment squad leader. Each member was trained to fill each billet within the fireteam: fireteam leader, grenadier, automatic rifleman, and rifleman. The assessment was executed under the supervision of MCOTEA functional test managers and a range officer-in-charge/range safety officer from the GCEITF.

K.1.1.1 Experimental Details

The 2-day Combat Engineer assessment replicated offensive and defensive tasks. The 1371s began each cycle on the offensive tasks, and transitioned to the defensive tasks on the second day of the evolution. Three 1371 squads executed each trial cycle: a control (C) nonintegrated group, a low-density (LD) group with two females, and a high-density (HD) group with four females. For fireteam-level tasks, a C group was non-integrated, an LD group had one female, and an HD group had two females.

Day 1 was executed on Ranges 107 and 108, and consisted of a squad-reinforced attack, where engineer fireteams conducted a Bangalore breach. After the squad attack, the engineer squads departed to Range 108 to conduct a demolitions raid.

Day 2 was executed on Ranges 110 and 114, and consisted of squad-level defensive actions. The day started with a 7-km forced march from Range 107 to Range 110 wearing an approach load and carrying personal weapons. Then the squad departed on foot to the start of the established mine lanes in the vicinity of Range 110 to begin the dismounted route clearance. Upon completion of the dismounted route clearance portion, the squad was transported to Range 114 to conduct a cache reduction evolution.

K.1.1.2 Additional Context

Throughout the duration of the assessment, Marines bivouacked at Range 107, sleeping in two-man tents. During trial execution, Marines wore/carried prescribed loads for each task. Weighing packs each day prior to the 7-km forced march ensured consistency. After each trial cycle, the Marines operated under the guidance of the platoon leadership, performing minimal physically demanding tasks. The Marines not part of an assessed squad conducted the same experimental subtasks after the assessed squads to ensure all Marine volunteers were equally fatigued each day. Section K.1.3 discussed these tasks in detail.

K.1.2 SOM Experimental Tasks

K.1.2.1 1-km Movement

Assaulting an enemy position never starts from a static position: First, a movement must be conducted to the assault position (AP). The distance from the line of departure to the AP is dependent on myriad factors. Based on time and space constraints, this distance was set at just under 1 km. Each Combat Engineer fireteam moved this distance as quickly as possible while wearing the fighting load and carrying personal weapons and a predetermined number of Bangalore torpedo sections. This specific task put the Marines under moderate fatigue prior to commencing the attack.

K.1.2.2 Negotiate an Obstacle

The conduct of an attack often involves reducing or negotiating an obstacle. It is common in an urban environment to make entry through a window or over a wall—some type of obstacle. One of the more difficult tasks is climbing over a wall with a fighting load. Each Combat Engineer fireteam negotiated an 8-foot wall (obstacle), getting all Marines and equipment over as quickly as possible.

K.1.2.3 Squad Attack

After negotiating the obstacle, fireteams conducted a 425-meter movement in trace of the infantry squad to the Limit of Advance (LOA). Upon arrival at the LOA, the fireteams held in place until the Infantry squad completed casualty evacuation (CASEVAC) drills.

K.1.2.4 Bangalore Breach

The Combat Engineer fireteams conducted a hasty breach using Bangalore torpedoes to breach a concertina wire obstacle. The purpose of such a breach, as detailed in the Engineer and Utilities Training and Readiness (T&R) Manual, is to “quickly overcome unexpected or lightly defended tactical obstacles in order to maintain the momentum of the attack by denying the enemy time to mass forces at the breach sites." Near the conclusion of the squad attack evolution, on arrival at the LOA, the fireteams held in place until the infantry squad completed CASEVAC drills. When the signal was given, the fireteams sprinted 100 meters with their Bangalore sections to breach a concertina wire obstacle. Once the fireteam had all 10 Bangalore sections connected, the explosives were dually primed and the Combat Engineers returned to the LOA. Once all Marines were verified in a safe position, the explosives were initiated via Modern Demolition Initiators (MDI) and the wire obstacles breached.

K.1.2.5 Demolitions Raid

The demolitions raid portion replicated a friendly force seizing enemy terrain through offensive action, such as the squad attack, and then immediately disrupting an enemy's mobility to prevent loss of recently acquired terrain. Upon completion of the Bangalore breach, the Combat Engineers returned to the 8-foot obstacle/wall. From this point in the trial, the Combat Engineer fireteams came together for a squad-level evolution. Awaiting the squad at the 8-foot obstacle/wall was a vehicle delivering the packs and explosives required to execute the demolitions raid. Each squad was provided two 40-lb. shape charges and two 40-lb. cratering charges. The charges were placed in the assault packs in addition to the assault load equipment.

K.1.2.5.1 2.4-km Hike

The Marines began a 2.4-km movement along a prescribed route to Range 108. Four Marines carried charges at the start of the movement; those not carrying charges carried their assault load. At the approximate halfway point of the 2.4-km movement (a point established at which the engineer squad departed the surface danger zone of Range 107) the Marines transitioned the load among squad members within their individual fireteam. Those who began the movement not carrying a charge now carried a charge, and those carrying a charge at the beginning of the movement now carried only the assault load and weapon. Once the load change was complete, the squad continued along the prescribed route to Range 108.

K.1.2.5.2 Shape and Crater Charges

Upon arrival at the 2.4-km movement stop point on Range 108, the fireteam carrying the 40-lb. shaped charges continued movement for another 300 meters to place the shaped charges. Once the charges were in place and rigged for detonation, the Marines initiated a 7-minute time fuse and returned the 300 meters to the 2.4-km movement stop point. Once the shaped charges detonated and the area was cleared by the range safety officer (RSO), the fireteam carrying the cratering charges on the 2.4-km movement now moved the 300 m to the detonation site to place the explosives. Once the charges were in place and rigged for detonation, the Marines initiated a 7-minute time-fuse initiation system and returned the 300 meters to the 2.4-km movement stop point. Once the cratering charges detonated, each Marine took a Fatigue and Workload survey to assess overall fatigue and workload of the entire offensive task (see GCEITF Experimental Assessment Plan [EAP], Annex D) while the RSO verified full detonation. Once the range was complete, the Marines boarded a vehicle for a return movement to Range 107.

K.1.2.6 7-km Hike

Combat Engineers serving in the Combat Engineer Battalions (CEBs) work in direct support, or are often attached to, Infantry units at various levels. The Engineer and Utilities T&R Manual states that Combat Engineers must be prepared to “[f]ight as provisional infantry, participate in offensive operations such as attacks, raids, movement to contact, etc. Defensive operations that include withdrawal, patrolling, check point ops, convoy ops, and employment of organic weapons. Additional responsibilities may include operations other than war (civil disturbance, TRAP, cordon and search) and MOUT (attack, defend, patrol, clear a building, vehicle checkpoint).” As such, Combat Engineers must be able to move through all sorts of terrain by foot. Units often train by conducting a forced march with an approach load at a sustained rate of march. Due to the physically demanding nature of moving under load, each Marine took a Fatigue and Workload survey after completion of the 7-km hike. For the assessment, each Combat Engineer squad had to move a distance of 7 km as quickly as possible while carrying an approach load. At completion of the 7-km hike, Marines took a fatigue and workload survey to assess their fatigue and workload during execution (see GCEITF EAP, Annex D).

K.1.2.7 Dismounted Route Clearance

A common responsibility of Combat Engineers, particularly those in the CEB, is to ensure mobility of the supported maneuver element. One common method is to conduct dismounted route clearance operations. The Engineer and Utilities T&R Manual states that Combat Engineers will “[c]onduct dismounted route sweep operations to detect, investigate, mark, report, and reduce Explosive Hazards (EH) and other obstacles along a defined route to enable assured mobility.” Upon arrival to Range 110 at the conclusion of the 7-km hike, the squad took a 10-minute operational pause to remove their assault pack from the main pack, secure the main packs in a vehicle, and conduct an operational check of the mine detectors.

Once the operational check was complete, the squad departed on foot to the start of the established mine lanes in the vicinity of Range 110 to begin the dismounted route clearance. Each squad leader had a specific lane assigned, and for each evolution in which he led a squad, he had the squad sweep his assigned lane. This assignment ensured that there was no overlap by the squads on an individual lane. A randomized target roster determined the number of target mines placed within the lane prior to the research evolution. This ensured that the same number of mines was present across the three lanes per day, but that the number and location of the mines in each lane changed from trial to trial. Upon completion of the dismounted route clearance portion, the squad was transported to Range 114 to conduct the cache reduction evolution.

K.1.2.8 Destroy Captured Arms and Ammunition with Explosives

A common responsibility of Combat Engineers, particularly those in the CEB, is to ensure the destruction of enemy arms and ammunition to prevent their use against friendly personnel. The Engineer and Utilities T&R Manual describes this task as: "Destroy captured arms and ammunition with demolitions to ensure destruction. Examples include: confined gaseous, liquid, and solid propellants; explosives; pyrotechnics; chemical and riot-control agents; smokes and incendiaries (including bulk explosives); chemical warfare agents; chemical munitions; rockets; guided and ballistic missiles; bombs; warheads; mortar rounds; artillery ammunition; small arms ammunition; grenades; mines; torpedoes; depth charges; cluster munitions and dispensers; demolition charges; and devices and components of the above."

For this evolution, once instructed by the squad leader, the Marines loaded artillery rounds onto the vehicle. Once the rounds were loaded, the vehicle operator secured them for transport while the Combat Engineer squad loaded on vehicles for transport to the detonation site on Range 114. Once the Combat Engineer squad and ammunition vehicle were present at the point of detonation, the squad dug, emplaced, and rigged the artillery shells for detonation. Pioneer equipment was provided, and the squad did not begin digging until instructed to do so by their squad leader. On order, the Marines dug eight reduction pits in which they placed four artillery rounds. Time for this evolution ended when all 8 holes were dug and all 32 artillery shells were offloaded from vehicle and placed in reduction pits, rigged for detonation, and buried. Once the assigned tasks were completed, accountability of personnel and equipment was conducted, and the charges were initiated via a 7-minute time system. The Marines were transported to the safety bunker on Range 114 for the blasts, and once all blasts had detonated, Marines took a Fatigue and Workload survey to assess their fatigue and workload during execution (see GCEITF EAP, Annex D). At the completion of the 2-day cycle, Marines took a cohesion survey to record their cohesion during execution (see GCEITF EAP, Annex M).

K.1.3 Loading Events

The loading plan ensured, to the greatest extent possible, equity of physical activity among all volunteers throughout the duration of the experimental assessment. Every trial and task was conducted in the same manner and sequence to ensure consistency. Due to the number of volunteers, a handful of Marines were not part of an assessed squad each 2-day cycle. Collaboration with the platoon and company leadership determined that the best method of loading non-assessed Marines was to form them into a quasi-squad and have them perform the same tasks as an assessed squad to experience the same conditions and physical strain. Minor modifications were permitted due to the reduced size of the squad.

K.1.4 SOM Summary

The Combat Engineer assessment of the GCEITF took place in a field environment aboard MCAGCC, Twentynine Palms, CA. The assessment consisted of 21 2-day trial cycles conducted over the course of 55 days. Each trial cycles consisted of an offensive and defensive day. The offensive day involved five tasks based around supporting a squad-reinforced attack: 1-km movement, negotiate an obstacle, squad attack, Bangalore breach, and a demolitions raid. The defensive day involved three tasks: 7-km hike, dismounted route clearance, and destroy captured arms and ammunition. During trial execution, Marines rotated through every billet within the Combat Engineer squad.

K.2 Limitations

K.2.1 1371 Limitations Overview

The GCEITF experiment allowed operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a manner similar to those in an operational environment, but artificial limitations or interruptions were introduced that changed or altered normal performance of a task. While these limitations represent a degree of artificiality, they do not detract significantly from our ability to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for the 1371 Combat Engineer assessment.

K.2.2 Relative Difficulty of Record Test

The Combat Engineer GCEITF assessment gathered data associated with some of the most physically demanding tasks of the 1371 MOS within the Combat Engineer Battalion (CEB). These tasks in isolation do not fully replicate life experienced by a Marine during a typical CEB FEX, not to mention a combat environment. With the limited time available to conduct the assessment, only selective 1371 tasks were assessed. Due to specific experimental constraints and human-factor considerations, other tasks/duties outside of the assessment were minimized. During a typical FEX, it is common for Marines to conduct 24-hour operations that include day and nighttime operations/patrols, standing firewatch or security post, and conducting continuous tactical actions. The offensive day SOM took squads approximately 2 hours to complete, and the defensive day SOM took approximately 4 hours to complete. Outside the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

Another primary concern designing the Combat Engineer assessment was to ensure that it was achievable and sustainable for a 60-day period. The 7-km forced march distance was selected based on the amount of training time available prior to the assessment; however, many of the loads were decreased. The Combat Engineers did not carry ammunition, radios, batteries, or other equipment often required when operating in a tactical environment. The Marines were authorized 1 day off after every 4 days of training. This artificial recovery period is not achievable when conducting training or combat operations.

A final factor affecting the relative difficulty of the record test had to do with the intangible physiological impact of the volunteers being able to drop on request (DOR) at any point in time. Any time a volunteer dropped during a trial cycle, the squad/team performed the following subtasks with fewer personnel. This affected the cohesion of each squad and influenced their performance.

K.3 Deviations

Deviations to the execution of the Combat Engineer SOM were made and can be found in the Experimental Data Report signed May 2015; however, no deviations affected the analysis methodology outlined in the EAP.

K.3.1 7-km Hike Pack Weight

The Test Plan stated that the Combat Engineer Squad would carry the same weight for the 7-km hike as the 0311 squad, as established by the Infantry T&R Manual, Annex E. Upon further examination of subsequent tasks for the Combat Engineer Squad, it was decided to reduce the weight to make it more operationally relevant for the test scenario being executed.

K.3.2 Dismounted Route Clearance

The Test Plan stated that once the Combat Engineers completed the 7-km hike, they would be transported to the start point of a 1,500-meter dismounted route clearance lane. The original lanes were constructed so that each ended on Range 114, the range on which the Cache Reduction evolution would occur. After the first pilot test, the dismounted route clearance lanes were moved to the vicinity of Range 110 and reduced in distance.

The movement of the lanes from Range 114 to Range 110 was done to eliminate a safety concern of the volunteers transiting an area adjacent to an active demolitions range. Once the lanes were moved, they were subsequently reduced due to terrain limitations. The sweep lanes had to be reduced in distance in order to keep similar conditions for all squads. Beyond 500 meters, terrain on the three lanes became significantly different, so the lanes were reduced in distance to ensure that terrain variation was not a factor in the number of mines found by each squad.

K.3.3 Destroy Captured Arms and Ammo

The Test Plan stated that the Combat Engineer squad would dig 4 reduction pits and subsequently place 8 155mm artillery shells in each pit for reduction. With the appropriate amount of C-4 placed on the rounds for explosive reduction, the net explosive weight (NEW) for each of the 4 reduction shots was 196 pounds. In order to reduce the fragmentation distance, the Combat Engineer squad instead dug 8 reduction pits and placed 4 artillery shells in each reduction pit. This new configuration placed the NEW for each of the 8 reduction shots at 98 pounds and reduced the fragmentation distance.

K.4 Data Set Description

K.4.1 Data Set Overview

The 1371 portion of the experiment consisted of 2 pilot trial cycles and 21 record trial cycles. The pilot trial cycles were conducted from 2 March 2015 to 6 March 2015. Pilot trial cycle data are not used in analysis due to variations in the conduct of the test. We based all analysis on the 21 record trial cycles executed from 7 March 2015 to 26 April 2015.

K.4.2 Record Test Volunteer Participants

At the beginning of the first record trial cycle, there were 18 male 1371 volunteers and 8 female volunteers. There were several male Marines who voluntarily withdrew, or were involuntarily withdrawn, during the execution of the experiment. The final number of male 1371 volunteers was 13, and there were no female Marines who left the experiment.

K.4.3 Planned, Executed, and Analyzed Trial Cycles

Table K-1 displays the number of trial cycles planned, executed, and analyzed by task. The planned number of trial cycles for the 1371 MOS per Section 7.5.3 of GCEITF EAP is 60 trials, or 20 trial cycles per planned integration level (C, LD, and HD) at the squad level, and 120 trials or 40 trials per planned integration level at the fireteam level. We were able to plan for this number of trial cycles, but due to the number of Marines who voluntarily withdrew or were involuntarily withdrawn throughout the experiment, we were able to run only one integrated squad (LD or HD), which is why there are so few executed LD and HD trial cycles.

Of note, there are several occurrences of missing data for the 7-km hike by individual kilometer. GPS data analysis provided the individual kilometer times. Early in the experiment, the Garmin GPSs were set to record a volunteer’s position every second.

The GPS could not hold all of the data due to the limited storage capacity on the GPS, the length of the trial when volunteers executed the 7-km hike, and the follow-on tasks. Thus the GPS overwrote the hike data. Once the problem was discovered, the GPSs were corrected to record location every two seconds. The exact date of the change is unknown.

Table K-1. 1371 Planned, Executed and Analyzed Trial Cycles

Task and Metric Description	Integration Level	Number of Planned Trials	Number of trials conducted	Number of trials used in analysis	Notes
2.4-km Hike; Total Time	C	21	20	20	
	LD	21	14	14	
	HD	21	12	12	
2.4-km Hike; First Half	C	21	20	19	No data: Apr 16
	LD	21	14	13	No data: Mar 9
	HD	21	12	12	
2.4-km Hike; Last Half	C	21	20	19	
	LD	21	14	13	No data: Mar 9
	HD	21	12	12	
Dig Trench/Unload Captured Arms and Ammo; Total Time	C	21	20	19	Remove Mar 8
	LD	21	14	13	Remove Mar 8
	HD	21	12	11	Remove Mar 8
Employ Bangalore; Breech by FT	C	42	42	39	Remove both Mar 8; Mar 9 not valid
	LD	42	30	28	Remove both Mar 8
	HD	42	31	29	Remove both Mar 8
Load Captured Arms and Ammo	C	21	20	19	Remove Mar 8
	LD	21	14	13	Remove Mar 8
	HD	21	12	11	Remove Mar 8
Sweep Designated Route for Explosive Hazards; Percent Detected	C	21	20	20	
	LD	21	14	14	
	HD	21	12	12	
7-km Hike; Total Time	C	21	20	20	
	LD	21	14	14	
	HD	21	12	12	
Employ Bangalore; Breech by Squad	C	21	20	19	Remove Mar 8; Mar 9 data not valid
	LD	21	14	12	Remove Mar 8
	HD	21	12	11	Remove Mar 8
Dig Trench/Unload Captured Arms and Ammo; Dig Trench	C	21	20	19	Remove Mar 8
	LD	21	14	13	Remove Mar 8
	HD	21	12	11	Remove Mar 8
Dig Trench/Unload Captured Arms and Ammo; Place Ordnance in Trench	C	21	20	19	Remove Mar 8
	LD	21	14	13	Remove Mar 8
	HD	21	12	11	Remove Mar 8
Dig Trench/Unload Captured Arms and Ammo; Rig Explosives	C	21	20	19	Remove Mar 8
	LD	21	14	13	Remove Mar 8
	HD	21	12	11	Remove Mar 8
1-km Hike; by FT	C	42	42	42	
	LD	42	30	30	
	HD	42	31	31	

Task and Metric Description	Integration Level	Number of Planned Trials	Number of trials conducted	Number of trials used in analysis	Notes
Negotiate Obstacle; by FT	C	42	41	41	
	LD	42	30	30	
	HD	42	31	31	
7-km Hike; 1km Time	C	21	20	16	No data: Mar 8, Mar 13, Mar 15, Mar 18
	LD	21	14	10	No data: Mar 8, Mar 10, Mar 15, Mar 18
	HD	21	12	8	No data: Mar 8, Mar 13, Mar 15, Mar 18
7-km Hike; 2km Time	C	21	20	16	No data: Mar 8, Mar 13, Mar 15, Mar 18
	LD	21	14	10	No data: Mar 8, Mar 10, Mar 15, Mar 18
	HD	21	12	8	No data: Mar 8, Mar 13, Mar 15, Mar 18
7-km Hike; 3km Time	C	21	20	14	No data: Mar 8, Mar 13, Mar 15, Mar 18. High outlier (GPS error): Mar 20 and Apr 19
	LD	21	14	10	No data: Mar 8, Mar 10, Mar 15, Mar 18
	HD	21	12	8	No data: Mar 8, Mar 13, Mar 15, Mar 18
7-km Hike; 4km Time	C	21	20	14	No data: Mar 8, Mar 13, Mar 15, Mar 18. GPS error: Mar 30 and Apr 19
	LD	21	14	10	No data: Mar 8, Mar 10, Mar 15, Mar 18
	HD	21	12	8	No data: Mar 8, Mar 13, Mar 15, Mar 18
7-km Hike; 5km Time	C	21	20	16	No data: Mar 8, Mar 13, Mar 15, Mar 18
	LD	21	14	10	No data: Mar 8, Mar 10, Mar 15, Mar 18
	HD	21	12	8	No data: Mar 8, Mar 13, Mar 15, Mar 18
7-km Hike; 6km Time	C	21	20	15	No data: Mar 8, Mar 13, Mar 15, Mar 18. High outlier (GPS error): Mar 28
	LD	21	14	10	No data: Mar 8, Mar 10, Mar 15, Mar 18
	HD	21	12	8	No data: Mar 8, Mar 13, Mar 15, Mar 18
7-km Hike; 7km Time	C	21	20	15	No data: Mar 8, Mar 13, Mar 15, Mar 18. GPS error: Mar 28
	LD	21	14	10	No data: Mar 8, Mar 10, Mar 15, Mar 18
	HD	21	12	8	No data: Mar 8, Mar 13, Mar 15, Mar 18

K.5 Descriptive and Basic Inferential Statistics

K.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common combat engineer tasks and are indicative of unit-level proficiency during either field exercises or combat operations. Only 7 tasks out of 21 tasks are presented in this section. The Appendix to this Annex contains the descriptive statistics for the remainder of the 1371 tasks. The words “metric” and “task” are used interchangeably throughout this Annex, since both refer to the experimental task.

Each fireteam consisted of four volunteer Marines: the fireteam leader, automatic rifleman, grenadier, and rifleman. Each squad consisted of 8 volunteer Marines (two fireteams) with a direct assignment (nonvolunteer) squad leader. There were three integration levels for all tasks. For squad-level tasks, a Control (C) group was non–gender integrated, a Low Density (LD) group was gender-integrated with one female Marine, and a High Density (HD) group was gender-integrated with two female Marines. For fireteam-level tasks, a C group was non–gender integrated, a LD group contained one female, and a HD group contained two females.

This section includes experimental results based on descriptive statistics, analysis of variance (ANOVA), Tukey tests (or nonparametric tests as necessary), and scatterplots. The subsequent sections will cover each task in detail. Finally, contextual comments, additional insights, and subjective comments (as applicable) tying back to each experimental task are incorporated.

Special caution should be taken when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks such as distances, techniques, leadership, load carried, group size, and group composition.

K.5.2 1371 Selected Tasks Descriptive Statistics Results

The following two tables display the results for the seven selected 1371 metrics. Table K-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations (SD).

Table K-3 displays ANOVA and Tukey test results, including metrics and integration levels, p-values suggesting statistical significance, integration level elapsed time differences, and percentage differences between integration levels. For each task, an ANOVA was conducted to compare the three groups simultaneously, and Tukey tests were conducted to make pairwise comparisons. If nonparametric tests were needed, Table K-3 displays these results instead of ANOVA and Tukey test results. If p-values are less than the a priori-determined significance level of 0.10, we conclude that there is statistical evidence that the mean time or percent hits was not found the be the same across all three groups.

Table K-2. 1371 Selected Task Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD
2.4-km Hike (minutes)[a]	C	20	32.90	3.17
	LD	14	33.10	2.23
	HD	12	37.47	3.29
2.4-km Hike; First Half (minutes)[b]	C	19	17.19	2.93
	LD	13	17.88	3.91
	HD	12	20.50	3.38
2.4-km Hike; Last Half (minutes)[a]	C	19	16.01	1.30
	LD	12	15.63	0.78
	HD	12	16.63	1.03
Dig Trench, Unload Captured Arms & Ammo (minutes)	C	19	14.90	5.69
	LD	13	18.29	6.55
	HD	11	17.62	4.99
Employ Bangalore; Breach Movements by FT (minutes)[a]	C	39	4.52	1.08
	LD	28	5.40	1.57
	HD	29	5.69	1.14
Load Captured Arms & Ammo (minutes)	C	19	2.57	0.55
	LD	13	2.70	0.45
	HD	11	2.58	0.29
Sweep Designated Route for Explosive Hazards (proportion of detected mines)	C	20	0.74	0.20
	LD	14	0.82	0.14
	HD	12	0.83	0.18

[a].Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a nonparametric equivalent test.

[b].Indicates contradicting statistical significance results between ANOVA and a nonparametric equivalent test.

Table K-3. 1371 Selected Task ANOVA and Tukey Test Results

Metric	F Statistic (df)	F Test p-Value	Comparison	Difference	% Difference	p-Value	80 % LCB[b]	80% UCB[b]	90% LCB[b]	90% UCB[b]
2.4-km Hike	10.27 (2, 43)	< 0.01[a]	LD-C	0.20	0.60%	0.98	-1.60	1.99	-1.97	2.37
			HD-C	4.57	13.89%	< 0.01[a]	2.69	6.45	2.30	6.84
			HD-LD	4.37	13.22%	< 0.01[a]	2.35	6.40	1.92	6.82
2.4-km Hike; First Half[a]	3.71 (2, 41)	<0.01[b]	LD-C	0.69	4.02%	0.73[c]	-1.43[c]	0.82[c]	-1.77[c]	1.18[c]
			HD-C	3.31	19.27%	<0.01[c]	2.33[c]	4.83[c]	1.82[c]	5.08[c]
			HD-LD	-2.62	14.66%	0.02[c]	1.83[c]	4.62[c]	1.43[c]	4.82[c]
2.4-km Hike; Last Half	2.52 (2, 40)	0.09[a]	LD-C	-0.38	-2.37%	0.62	-1.09	0.33	-1.24	0.48
			HD-C	0.62	3.85%	0.29	-0.09	1.33	-0.24	1.48
			HD-LD	-1.00	6.38%	0.08[a]	0.21	1.78	0.05	1.95
Dig Trench, Unload Captured Arms & Ammo	1.54 (2, 40)	0.23	LD-C	3.39	22.77%	0.25	-0.26	7.04	-1.02	7.80
			HD-C	2.72	18.24%	0.44	-1.13	6.56	-1.93	7.36
			HD-LD	0.67	-3.69%	0.96	-4.83	3.48	-5.70	4.35
Employ Bangalore; Breach Movements by FT	8.02 (2, 93)	< 0.01[a]	LD-C	0.87	19.33%	0.02[a]	0.34	1.41	0.23	1.52
			HD-C	1.68	25.81%	< 0.01[a]	0.63	1.70	0.53	1.81
			HD-LD	-0.29	5.43%	0.65	-0.28	0.87	-0.40	0.99
Load Captured Arms & Ammo	0.36 (2, 40)	0.70	LD-C	0.14	5.28%	0.70	-0.16	0.43	-0.22	0.49
			HD-C	0.01	0.51%	1.00	-0.30	0.32	-0.36	0.39
			HD-LD	0.12	-4.53%	0.80	-0.46	0.21	-0.53	0.28
Sweep Designated Route for Explosive Hazards	1.35 (2, 43)	0.27	LD-C	0.08	11.43%	0.38	-0.03	0.19	-0.05	0.22
			HD-C	0.09	12.47%	0.35	-0.03	0.21	-0.05	0.23
			HD-LD	-0.01	0.93%	0.99	-0.12	0.13	-0.14	0.16

[a.] Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to ANOVA or a nonparametric equivalent test.
[b.] Tukey intervals have familywise confidence of the indicated percentage; each interval is not of the given confidence level on its own.
[c.] Results presented are from Kruskal-Wallis and Mann-Whitney nonparametric tests due to non-normality.

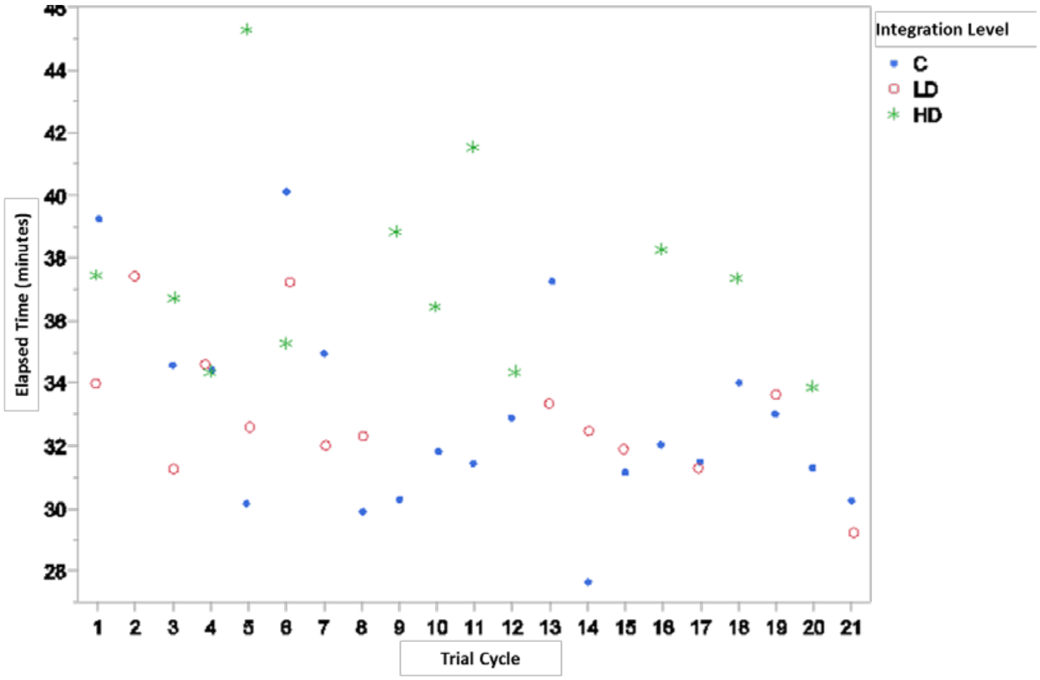
K.5.2.1 2.4-km Hike

K.5.2.1.1 2.4-km Hike Overview

The demolitions raid portion replicated a friendly force seizing enemy terrain through offensive action, and then disrupting an enemy's mobility to prevent loss of recently acquired terrain. For the assessment, each Combat Engineer squad executed a 2.4-km movement under load and employed demolitions upon arriving at the objective. The recorded time for this task started when the squad departed the Range 107 start point and stopped when the squad arrived at the Range 108 stop point.

Figure K-1 displays all 1371 2.4-km hike data. All data on the scatterplot are valid for analysis

Figure K-1. 2.4-km Hike



The data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.13 for the C group, 0.30 for the LD group, and 0.104 for the HD group.

The C group had a mean time of 32.90 minutes. This time is faster (but not statistically significantly) than the LD mean time of 33.10 minutes, and statistically significantly faster than the HD mean time of 37.47 minutes. These differences result in 0.60% (0.20-minute) and 13.89% (4.57-minute) degradations in time for the LD and HD groups, respectively. In addition, the LD groups had less variability and HD groups had greater variability, as shown by the 0.94-minute and 0.12-minute respective changes in SD (3.17 minutes for the C group, 2.23 minutes for the LD group, and 3.29 minutes for the HD group). The LD group was statistically significantly faster than the HD group. There was a 4.37% degradation in hike time from the LD to HD group. See Table K-2 and Table K-3 for detailed analytical results.

K.5.2.1.2 2.4-km Hike Contextual Comments

Analysis shows that the C group had a mean time of 32.90 minutes, while the LD and HD groups had mean times of 33.10 minutes (12 seconds slower than the C group) and 37.47 minutes (4:34 minutes slower than the C group), respectively.

K.5.2.1.3 2.4-km Hike Additional Insights

Time is critical when a unit is moving into position with the intent to deploy demolitions, and it can be the determining factor in the success of the mission. Marine Corps Doctrinal Publications (MCDP) consistently emphasize the importance of speed. MCDP 1-3: *Tactics* devotes an entire chapter to "Being Faster," which states: "Physical speed, moving more miles per hour, is a powerful weapon in itself." MCDP-6: *Command and Control* also speaks to speed relative to the enemy: "The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results." A 2.4-km hike movement is meant to be executed quickly, without delay. Faster results are favorable, because reaching the objective to employ demolitions is critical to the mission.

K.5.2.1.4 2.4-km Hike Subjective Comments

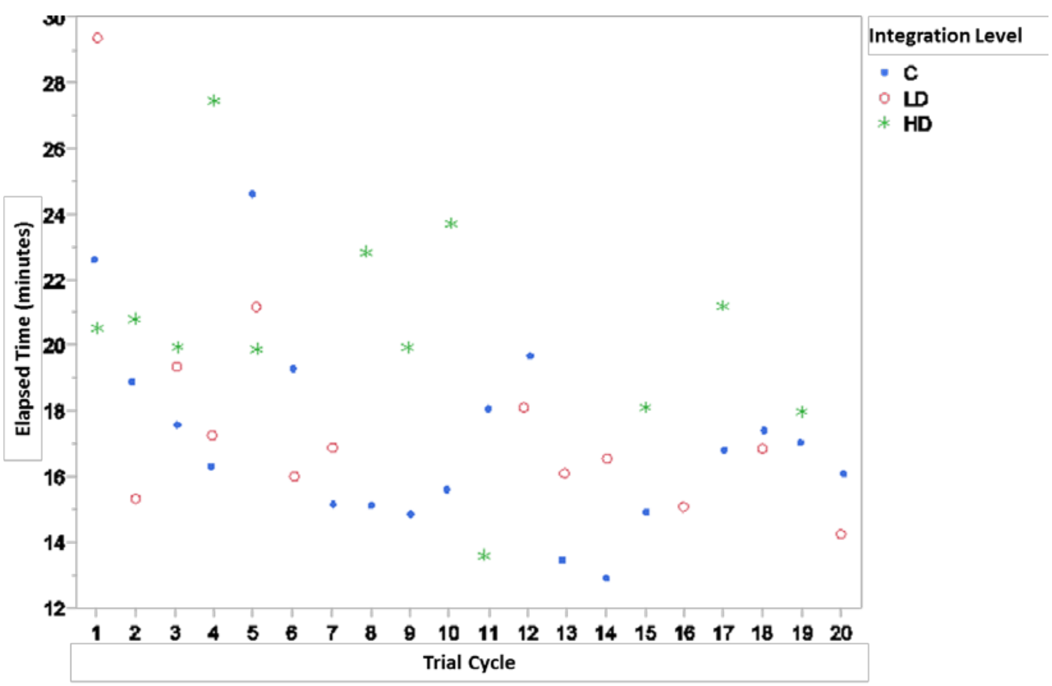
For subjective comments relating to this task, see the Appendix.

K.5.2.2 2.4-km Hike First Half

K.5.2.2.1 2.4-km Hike First Half Overview

Figure K-2 displays all 2.4-km hike/first data. All data on the scatterplot are valid for analysis.

Figure K-2. 2.4-km Hike First Half



The C group and HD group data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.17 for the C group and 0.62 for the HD group. However, the LD data are not normally distributed (p-value <0.01).

The C group had a mean time of 17.19 minutes. This time is faster (but not statistically significantly) than the LD mean time of 17.88 minutes, and statistically significantly faster than the HD mean time of 20.50 minutes. These differences result in 4.02% (0.69-minute) and 19.27% (3.31-minute) degradations in time for the LD and HD groups, respectively. The LD group was statistically significantly faster than the HD group only in a one-sided t-test, but in both one- and two-sided Mann-Whitney tests. There was a 14.66% (2.62-minute) degradation in hike time from the LD to HD group. Because of a lack of normality, we recommend using the Mann-Whitney test results (reported in Table K-3). However, the conflict of parametric and nonparametric results suggests that further study of this task is warranted.

In addition, the LD and HD groups had greater variability, as shown by the 0.98-minute and 0.45-minute respective changes in SD (2.93 minutes for the C group, 3.91 minutes for the LD group, and 3.38 minutes for the HD group). See Table K-2 and Table K-3 for detailed analytical results.

K.5.2.2.2 2.4-km Hike First Half Contextual Comments

None.

K.5.2.2.3 2.4-km Hike First Half Additional Insights

None.

K.5.2.2.4 2.4-km Hike First Half Subjective Comments

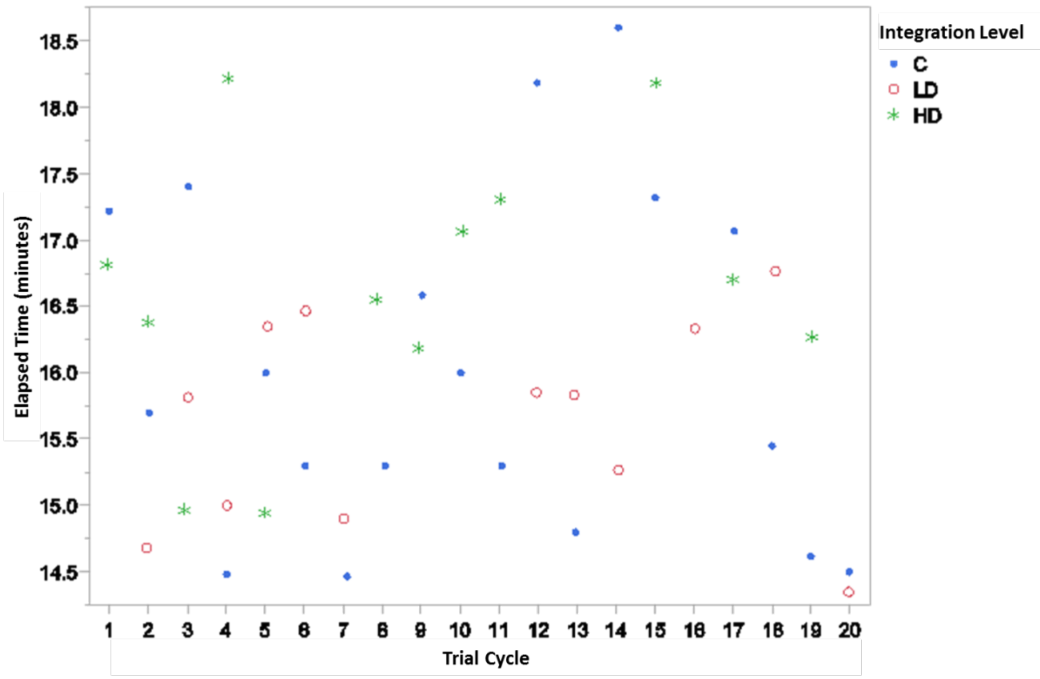
For subjective comments relating to this task, see the Appendix.

K.5.2.3 2.4-km Hike Last Half

K.5.2.3.1 2.4-km Hike Last Overview

Figure K-3 displays all 2.4-km hike last half data. All data on the scatter plot are valid for analysis.

Figure K-3. 2.4-km Hike/Last Half



The data are normally distributed as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.12 for the C group, 0.58 for the LD group, and 0.43 for the HD group.

The C group had a mean time of 16.01 minutes. This time is slower (but not statistically significantly) than the LD mean time of 15.63 minutes, and faster (but not statistically significantly) than the HD mean time of 16.63 minutes. These differences result in 2.37% (0.38-minute) improvement in time for the LD group and 3.85% (0.62-minute) degradation in time for the HD group. In addition, the LD and HD groups had less variability, as shown by the 0.52-minute and 0.27-minute respective changes in SD (1.30 minutes for the C group, 0.78 minutes for the LD group, and 1.03 minutes for the HD group). The LD group was statistically significantly faster than the HD group. There was a 6.38% (1.00-minute) degradation in hike time from the LD to HD group. See Table K-2 and Table K-3 for detailed analytical results.

K.5.2.3.2 2.4-km Hike Last Half Contextual Comments

None.

K.5.2.3.3 2.4-km Hike Last Half Additional Insights

None.

K.5.2.3.4 2.4-km Hike Last Half Subjective Comments

For subjective comments relating to this task, see the Appendix.

K.5.2.4 Dig Trench, Unload Captured Arms and Ammo, Rig for Detonation

K.5.2.4.1 Dig Trench, Unload Captured Arms and Ammo, Rig for Detonation Overview

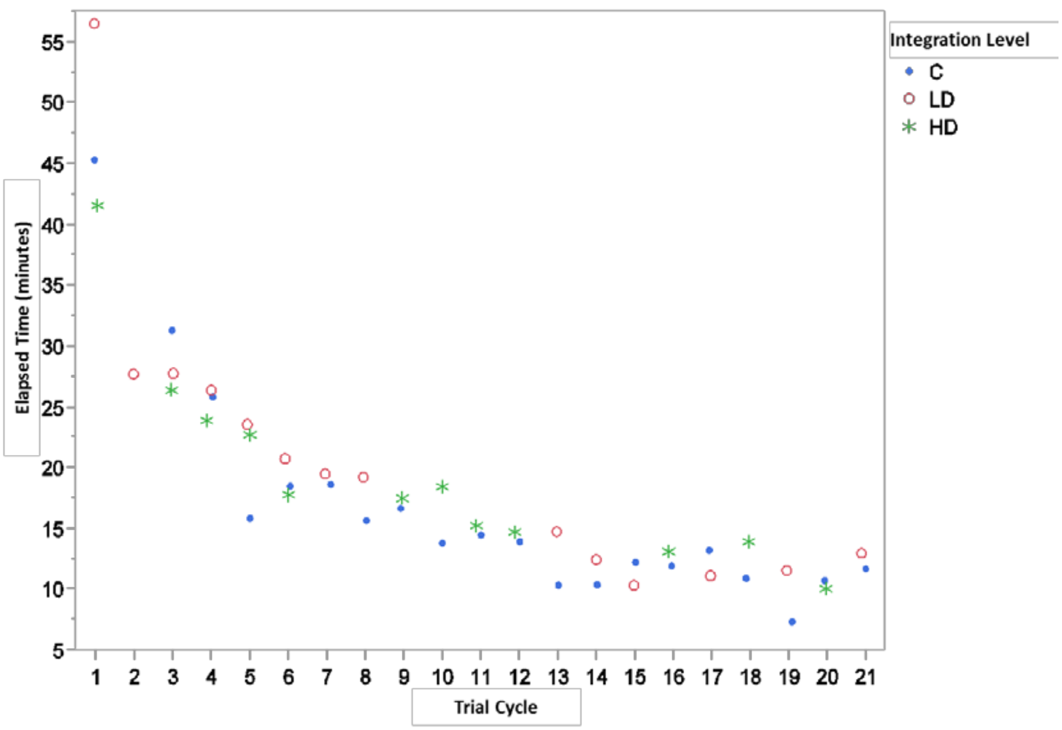
A common responsibility of Combat Engineers—particularly those in the currently closed CEB—is to ensure the destruction of enemy arms and ammunition to prevent their use against friendly personnel. For the assessment, each Combat Engineer squad had to dig 8 reduction pits to specified dimensions, unload 32 155-mm artillery shells from the MTVR, place them in the reduction pits, rig for detonation, and bury the charges as quickly as possible. Time began when the squads began digging the trenches and ended once the 155-mm shells were buried.

This task is broken down into the following sub-tasks:

- Dig Trench
- Place Explosives
- Rig for Detonation.

Figure K-4 displays all destroy-captured-arms-and-ammo-with-explosives data. All data on the scatterplot are valid for analysis.

Figure K-4. Dig Trench, Unload Captured Arms and Ammo, Rig for Detonation



The LD and HD group data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.105 for the LD group and 0.75 for the HD group. However, the results for the C group do not appear normal with a p-value < 0.01.

Because nonparametric results are consistent with ANOVA results for this task, we present only the ANOVA results below.

The C group had a mean time of 14.90 minutes. This time is faster (but not statistically significantly) than the LD mean time of 18.29 minutes and the HD mean time of 17.62 minutes. These differences result in 22.77% (3.39-minute) and 18.24% (2.72-minute) degradations in time for the LD and HD groups, respectively. In addition, the LD group had greater variability and the HD groups had less variability, as shown by the 0.86-minute and 0.70-minute respective changes in SD (5.69 minutes for the C group, 6.55 minutes for the LD group, and 4.99 minutes for the HD group). The LD group was slower (but not statistically significantly) than the HD group. There was a 3.69% (0.67-minute) improvement in time from the LD to HD group. See Table K-2 and Table K-3 for detailed analytical results.

K.5.2.4.2 Dig Trench, Unload Captured Arms and Ammo, Rig for Detonation Contextual Comments

None.

K.5.2.4.3 Dig Trench, Unload Captured Arms and Ammo, Rig for Detonation Additional Insights

In an operational environment, the time it takes a unit to accomplish this type of task corresponds with time in a relatively static position, in the open and away from cover from indirect fires. Faster times are favorable in the performance of this task to limit exposure and the possibility of taking casualties. Loading and transferring captured ammunitions in an expeditious manner is a key aspect of mission accomplishment.

K.5.2.4.4 Dig Trench, Unload Captured Arms and Ammo, Rig for Detonation Subjective Comments

For subjective comments relating to this task, see the Appendix.

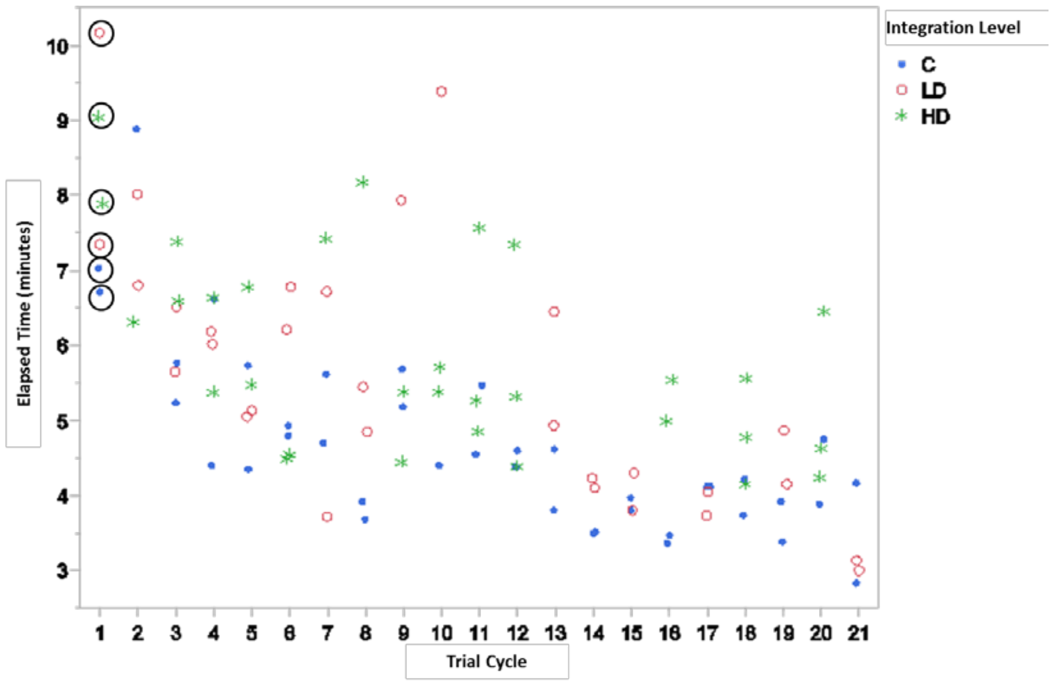
K.5.2.5 Employ Bangalore

K.5.2.5.1 Employ Bangalore Overview

The Combat Engineer fireteams used Bangalore torpedoes to breach a concertina wire obstacle. For the assessment, each four-Marine Combat Engineer fireteam executed a hasty Bangalore breach as quickly as possible. The recorded time began on the command “Breach, Breach, Breach,” and ended when the last Marine was prone at the LOA.

Figure K-5 displays all 1371 Employ Bangalore data. Data for all groups on March 7 were invalid because these data points were high outliers due to the learning curve. In addition, data for the C group on March 9 were invalid due to a test incident. With the exception of these data points, all data on the scatterplot are valid for analysis.

Figure K-5. Employ Bangalore



For this task, sample sizes are sufficiently large (n > 30) to satisfy the normality assumption for ANOVA.

The C group had a mean time of 4.50 minutes. This time is statistically significantly faster than the LD mean time of 5.40 minutes and the HD mean time of 5.69 minutes. These differences result in 19.33% (0.87-minute) and 25.81% (1.68-minute) degradations in time for the LD and HD groups, respectively. In addition, the LD group had greater variability, as shown by the 0.49-minute increase in SD (1.08 minutes for the C group, 1.57 minutes for the LD group). The LD group was faster, on average, than the HD group. There was a 5.43% (0.29-minute) degradation in hike time from the LD to the HD group, but this difference is not statistically significant. See Table K-2 and Table K-3 for detailed analytical results.

K.5.2.5.2 Employ Bangalore Contextual Comments

Although only the HD group showed a statistical significance, the execution of a breach is a critical point in combat operations and time is critical. Employing a Bangalore and executing a breach requires a unit to overcome enemy obstacles rapidly and maintain friendly momentum. According to Sun Tzu: “Speed is the essence of war. Take advantage of the enemy’s unpreparedness; travel by unexpected routes and strike him where he has taken no precautions” (MCDP-1, p. 69). Speed and surprise are crucial to success.

Faster results are more desirable because the longer a unit spends conducting a breach, the longer it is exposed to the effects of concentrated enemy fires and the greater the propensity of increased casualties or potential mission failure. The ANOVA and other statistical tests support the claim that integrated Combat Engineer units, on average, would take longer to execute a breach when compared to a nonintegrated Combat Engineer unit.

K.5.2.5.3 Employ Bangalore Additional Insights

None.

K.5.2.5.4 Employ Bangalore Subjective Comments

For subjective comments relating to this task, see the Appendix.

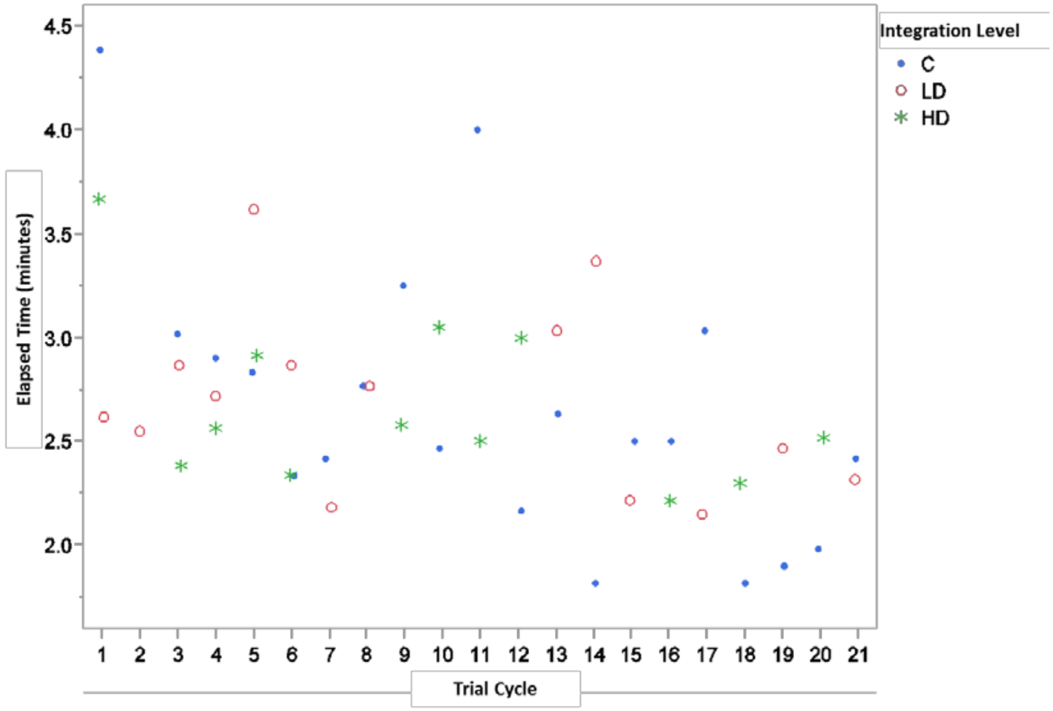
K.5.2.6 Load Captured Arms and Ammo

K.5.2.6.1 Load Captured Arms and Ammo Overview

A common responsibility of Combat Engineers—particularly those in the closed CEB—is to ensure the destruction of enemy arms and ammunition to prevent their use against friendly personnel. Stockpiles of enemy arms require movement to a reduction site to prevent unnecessary collateral damage. For the assessment, each Combat Engineer squad loaded 32 155-mm artillery shells from ground level to the back of a 7-ton Medium Tactical Vehicle Recovery (MTVR) vehicle as quickly as possible. Time for this task began when the first Marine began moving rounds and ended when all rounds were on the MTVR.

Figure K-6 displays all load-captured-arms-and-ammo data. All data on the scatterplot are valid for analysis.

Figure K-6. Load Captured Arms and Ammo



The data are normally distributed as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.29 for the C group, 0.43 for the LD group, and 0.17 for the HD group.

The C group had a mean time of 2.57 minutes. This time is faster (but not statistically significantly) than the LD mean time of 2.70 minutes and the HD mean time of 2.58 minutes. These differences result in 5.28% (0.14-minute) and 0.51% (0.01-minute) degradations in time for the LD and HD groups, respectively. In addition, the LD and HD group had less variability, as shown by the 0.10-minute and 0.26-minute respective changes in SD (0.55 minutes for the C group, 0.45 minutes for the LD group, and 0.29 minutes for the HD group). The LD group was slower (but not statistically significantly) than the HD group. There was a 4.53% (0.12-minute) improvement in time from the LD to the HD group. See Table K-2 and Table K-3 for detailed analytical results.

K.5.2.6.2 Load Captured Arms and Ammo Contextual Comments

None.

K.5.2.6.3 Load Captured Arms and Ammo Additional Insights

In an operational environment, the time it takes a unit to accomplish this type of task corresponds with time in a relatively static position, likely in the open and away from cover from indirect fires. Faster times are favorable in the performance of this task to limit exposure and the possibility of taking casualties. Loading and transferring captured ammunitions in an expeditious manner is a key aspect of accomplishing this mission.

K.5.2.6.4 Load Captured Arms and Ammo Subjective Comments

Although the times are not statistically significant, it is important to note that during the execution of this particular task there were numerous observations made of males in the integrated groups shouldering a disproportionate amount of the physical load to accomplish the mission. There were 34 total GCEITF direct assignment subjective comments regarding this task. Of the 34 comments, there were 16 comments regarding males in an integrated squad (either LD or HD) compensating for other Marines or doing the majority of the work. Of the 34 comments, there were 6 comments noting females who required assistance because they were unable to lift and push the rounds up the height of the vehicle bed by themselves.

There was no prescribed method for the squad to accomplish this task to prevent the introduction of operational artificiality, but it was observed that males assumed the more physically demanding task of lifting the rounds to the bed of the vehicle, and that females, when attempting this same task, could not do so on their own.

For additional subjective comments relating to this task, see the Appendix.

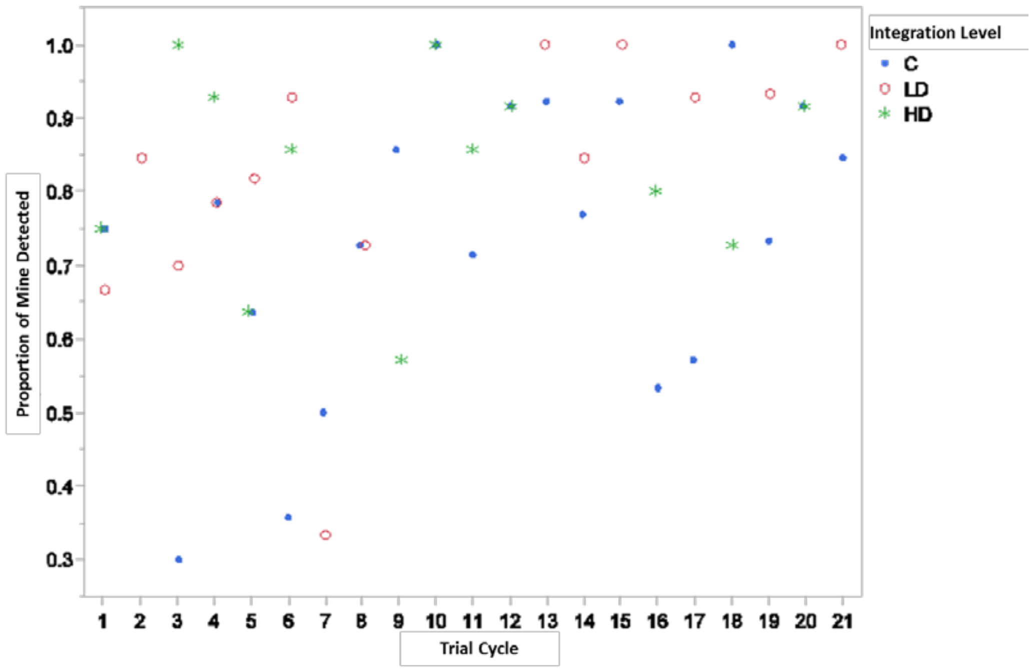
K.5.2.7 Sweep Designated Route for Explosive Hazards

K.5.2.7.1 Sweep Designated Route for Explosive Hazards Overview

A common responsibility of Combat Engineers—particularly those in the closed Combat Engineer Battalion—is to ensure the mobility of the supported maneuver element. One method is to conduct dismounted route clearance operations. For the assessment, each Combat Engineer squad swept a 500-meter route and located buried inert mines. There was no time component to this task.

Figure K-7 displays all sweep-designated-route-for-explosive-hazards data. All data on the scatterplot are valid for analysis.

Figure K-7. Sweep Designated Route for Explosive Hazards



The data are normally distributed, as evidenced by the Shapiro-Wilk test that resulted in a p-value of 0.17 for the C group, 0.02 for the LD group, and 0.48 for the HD group.

The C group had a mean proportion of mines found of 0.74. This proportion is lower (but not statistically significantly) than the LD mean proportion of 0.82 and the HD mean proportion of 0.83. These differences result in 11.43% (0.08–percentage point) and 12.47% (0.09–percentage point) degradations in time for the LD and HD groups, respectively. The LD group proportion was lower (but not statistically significantly) than the HD group. There was a 0.93% (0.01–percentage point) increase in proportion from the LD to the HD group. See Table K-2 and Table K-3 for detailed analytical results.

K.5.2.7.2 Sweep Designated Route for Explosive Hazards Contextual Comments

From an operational standpoint, the most favorable results are the ones with the highest find percentage. If an explosive hazard is not identified and mitigated by the route clearance element, a follow-on unit may trigger it, resulting in injury or death. The ability for Combat Engineers to locate and eliminate explosive hazards is critical to the mission; a higher find percentage is important as a contributor to the success of the mission.

K.5.2.7.3 Sweep Designated Route for Explosive Hazards Additional Insights

None.

K.5.2.7.4 Sweep Designated Route for Explosive Hazards Subjective Comments

For subjective comments relating to this task, see the Appendix.

K.6 Statistical Modeling Results

K.6.1 Statistical Modeling Overview

The previous section discussed results only as they pertain to differences due to integration level alone. The goal of statistical modeling as applied here is to estimate simultaneously the effect of gender integration levels and other relevant variables on Engineer squad performance. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For the same seven selected tasks described in the previous section, this section presents an overview of the analysis and results, and then presents the modeling results for each of the tasks.

For each task, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates that an increase in that variable will result in a decrease in the response variable, which is a desired outcome for elapsed time but not a desired outcome for the percent of mines found outcome. The results report where certain patch numbers are significant for a given variable. The experiment tracked Marines within the Engineer squad by a patch number that associated their random position within the squad to a specific billet. Table K-4 displays the patch numbers and associated billet titles for the Combat Engineer squad.

Table K-4. Patch Numbers and Billet titles for the Combat Engineer Squad

Patch Number	Billet Title
1	FT 1 Fireteam Leader
2	FT 1 Automatic Rifleman
3	FT 1 Grenadier
4	FT 1 Rifleman
5	FT 2 Fireteam Leader
6	FT 2 Automatic Rifleman
7	FT 2 Grenadier
8	FT 2 Rifleman

K.6.2 1371 Method of Analysis

Due to the small number of trials, a mixed-effects model with all engineer squad members and all types of personnel data did not work for the 1371 dataset. Thus we model each personnel variable with integration level separately with a random effect for who filled each position within the engineer squad. For example, age for each member of the engineer squad (eight variables) and integration level are modeled with the result (response time or percentage of mines found) as the response variable. Where maximum likelihood estimation converged, the Akaike information criterion (AIC) was used for variable selection. Otherwise, we comment on the significance of individual variables in the full model. Variables reported as significant are concluded to be significant based on at least a one-sided test.

K.6.3 1371 Selected Tasks Overall Modeling Results

There are no personnel data variables that are both statistically significant and have a practical impact to the model. Each time personnel data variables are statistically significant in a model, their effects are practically negligible, conflicting, and/or incomplete for the Engineer squad (i.e., there are no tasks for which a variable is significant for all members of the Combat Engineer squad).

Integration level is significant in the final model for the following selected tasks: 2.4-km hike, 2.4-km hike second half, Bangalore breach, and load captured arms and ammo. For each of these tasks, modeling the random effects for the individuals participating in the task results in changes from the initial results in the descriptive statistics. Each respective task paragraph describes these changes. For all other tasks, AIC chose the intercept model. See Section K.5 for descriptive statistics for these tasks.

K.6.3.1 2.4-km Hike

We model elapsed time for the 2.4-km hike as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position on the squad. For each model, we report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not converge:

- PFT crunches
- CFT MANUF
- Rifle score.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Age

- Height
- Weight
- AFQT score
- GT score
- PFT 3-mile run.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- GT score
- CFT MTC.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 2.4-km hike time:

- Age of patch 1
- Height of patches 2 and 3
- AFQT score of patch 2
- GT score of patch 3
- CFT MTC of patches 1, 2, 5, 6, and 7.

The following variables are significant in their respective models and are negatively correlated with the 2.4-km hike time:

- Age of patches 2 and 4
- Weight of patch 4
- AFQT score of patches 6 and 8
- GT score of patches 2, 4, 6 and 7.

The following personnel variables have no significant variables in their respective models:

- Number of female carries.

Because we did not identify any discernable patterns in the effects of the personnel variables, and because their effects are often negligible, our final model includes integration level with only HD significant with a difference of 4.04 minutes when compared to the C group. The comparison yields a statistically significant p-value of <0.01. This difference is a decrease from the 4.57 difference found in the descriptive statistics, which is a 11.59% change.

K.6.3.2 2.4-km Hike First Half

We model elapsed time for the 2.4-km hike first half as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position on the squad. For each model, we report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not converge:

- Age
- PFT crunches
- PFT 3-mile run
- Rifle score.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Height
- AFQT score
- GT score.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- GT score.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 2.4-km hike first half time:

- Height of patch 3
- GT score of patch 3
- CFT MTC of patches 1, 5, and 7

- CFT MANUF of patches 1 and 2.

The following variables are significant in their respective models and are negatively correlated with the 2.4-km hike first half time:

- Weight of patches 1, 2, 3, 4, and 7
- AFQT score of patch 8
- GT score of patches 4, 6, and 7
- CFT MANUF of patch 5.

The following personnel variables have no significant variables in their respective models:

- None.

Because we did not identify any discernable patterns in the effects of the personnel variables, and because their effects are often negligible, our final model includes integration level only where HD has a difference of 3.18 minutes when compared to the C group. The comparison yields a statistically significant p-value of <0.01. This difference is a decrease from the 3.31-minute difference identified in the descriptive statistics, which is a 3.93% change. However, AIC prefers the intercept model. Refer to Section K.5.2.2 to see the ANOVA summary for this task.

K.6.3.3 2.4-km Hike/Last Half

We model elapsed time for the 2.4-km hike/second half) as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position on the squad. For each model, we report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not converge:

- Age
- PFT crunches
- PFT 3-mile run.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Height
- Weight

- AFQT score
- GT score
- Rifle score.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- GT score.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the 2.4-km hike/last half time:

- GT score of patches 3 and 8
- CFT MTC of patches 1, 2, and 6
- Rifle score of patches 1 and 6.

The following variables are significant in their respective models and are negatively correlated with the 2.4-km hike/last half time:

- Weight of patch 4
- GT score of patches 4 and 7
- CFT MANUF of patches 1 and 3
- Rifle score of patch 2.

The following personnel variables have no significant variables in their respective models:

- None.

Because we did not identify any discernable patterns in the effects of the personnel variables, and because their effects are often negligible, our final model includes integration level only where HD has a difference of 0.73 minutes when compared to the C group. The comparison yields a statistically significant p-value of 0.08. This difference is an increase from the 0.62-minute difference identified in the descriptive statistics, which is a 17.74% change.

K.6.3.4 Dig Trench, Unload Captured Arms and Ammo

We model elapsed time for the dig trench, unload captured arms and ammo as a function of each personnel variable and integration level in a separate mixed model.

The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position on the squad. For each model, we report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not converge:

- Height
- CFT MTC
- PFT crunches.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Age
- GT score
- CFT MANUF
- PFT 3-mile run
- Rifle score.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the dig trench, unload captured arms and ammo time:

- Squad leader
- CFT MANUF of patch 6.

The following variables are significant in their respective models and are negatively correlated with the dig trench, unload captured arms and ammo time:

- Age of patches 4, 7 and 8
- Weight of patches 2, 4, and 7
- AFQT score of patches 1 and 7

- GT score of patch 1
- CFT MANUF time of patch 7.

The following personnel variables have no significant variables in their respective models:

- None.

Because we did not identify any discernable patterns in the effects of the personnel variables, and because their effects are often negligible, our final model includes integration level only where LD has a difference of 2.62 minutes when compared to the C group. The comparison yields a statistically significant p-value of 0.15 in a one-way test. However, AIC prefers the intercept model. Refer to Section K.5.2.4 to see the ANOVA summary for this task.

K.6.3.5 Employ Bangalore

We model elapsed time for the Bangalore breach as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position on the squad. For each model, we report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not converge:

- None.

Both the HD and the LD integration levels are significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Age
- AFQT score
- GT score
- PFT crunches
- PFT three-mile
- Rifle score.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the Bangalore breach time:

- Age of patch 2
- CFT MTC of patches 1, 2, and 3
- CFT MANUF of patches 3 and 4
- PFT 3-mile run of patch 1.

The following variables are significant in their respective models and are negatively correlated with the Bangalore breach time:

- Age of patch 4
- Height of patches 2, 3, and 4
- Weight of patches 1, 2, 3, and 4
- GT score of patch 4
- PFT crunches of patch 1
- Rifle score of patch 1.

The following personnel variables have no significant variables in their respective models:

- None.

Because we did not identify any discernable patterns in the effects of the personnel variables, and because their effects are often negligible, our final model includes integration level only where HD has a difference of 1.08 minutes when compared to the C group. The comparison yields a statistically significant p-value of <0.01. This difference is a decrease from the 1.68 minutes identified in the descriptive statistics, which is a 35.71% change. The LD integration level has a difference of 0.74 minutes when compared to the C group and a p-value of 0.01. This difference is a decrease from the 0.87 difference found in the descriptive statistics, which is a 14.94% change.

K.6.3.6 Load Captured Arms and Ammo

We model elapsed time for the load captured arms and ammo as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position on the squad. For each model, we report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not converge:

- GT score
- PFT 3-mile run.

The HD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Weight.

The LD integration level is significant and positively correlated with the response for the models that include the following personnel variables:

- Squad leader
- Age
- Height
- AFQT score
- CFT MTC
- Rifle score.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- PFT crunches.

The following personnel variables are significant in their respective models and are positively correlated with the load captured arms and ammo time:

- Age of patches 4, 6, and 7
- CFT MANUF of patches 2, 4, and 6
- PFT crunches of patches 4, 6, and 7.

The following variables are significant in their respective models and are negatively correlated with the load captured arms and ammo time:

- Age of patch 8
- Height of patches 7 and 8
- Weight of patches 1, 3, and 5
- AFQT score of patches 1, 5, 6 and 8
- CFT MANUF of patch 8
- PFT crunches of patches 2 and 5
- Rifle score of patch 5.

The following personnel variables have no significant variables in their respective models:

- None.

Because we did not identify any discernable patterns in the effects of the personnel variables, and because their effects are often negligible, our final model includes integration level only where LD has a difference of 0.30 minutes when compared to a C group. The comparison yields a statistically significant p-value of 0.04. This difference is greater than the 0.14-minute difference found in the descriptive statistics, which is a 16% change.

K.6.3.7 Sweep Designated Route for Explosive Hazards

We model proportion sweep designated route for explosives as a function of each personnel variable and integration level in a separate mixed model. The covariates in each model are the values of each personnel variable for each patch number, integration level, and a random effect of who filled each position on the squad. For each model, we report statistically significant positive and negative correlations, and whether we observe any patterns.

The models for the following variables do not converge:

- Weight
- PFT crunches.

Both the HD and the LD integration levels are significant and positively correlated with the response for the models that include the following personnel variables:

- Age.

Both the HD and the LD integration levels are significant and negatively correlated with the response for the models that include the following personnel variables:

- None.

The following personnel variables are significant in their respective models and are positively correlated with the proportion of mines found:

- Age of patches 3, 7 and 8
- GT score of patches 6 and 8
- CFT MANUF for patch 3
- PFT 3-mile run of patches 3, 4, 6 and 8.

The following variables are significant in their respective models and are negatively correlated with the proportion of mines found:

- GT score of patches 4 and 7
- Height of patches 1 and 3.

The following personnel variables have no significant variables in their respective models:

- Squad leader
- CFT MTC
- Rifle score.

Because we did not identify any discernable patterns in the effects of the personnel variables, and because their effects are often negligible, our final model includes integration level only. AIC prefers only the intercept in the final model. Refer to Section K.5.2.7 to see the ANOVA summary for this task.

Appendix to Annex K
1371 Supplemental Information

This appendix provides supplemental information for the 1371 portion of the GCE ITF experiment. It provides additional descriptive and basic inferential statistics not described in Annex K.

Section 1: GCE ITF Leadership Subjective Comments

The GCE ITF leadership provided comments on their observations of the experiment throughout its execution. Table K A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table K A – Summary of GCE ITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement				Requesting extra breaks				Requires extra assistance				Needs no assistance				Compensating for another Marine				Gear pass off				Other				No category				Total
		C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	C	LD	HD	Total	
2.4-km Hike	M	6	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
	F	0	0	3	3	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Load Captured Arms and Ammo	M	0	0	0	0	0	0	0	0	0	0	0	0	11	5	1	17	0	7	3	10	0	0	0	0	0	0	0	0	1	0	0	1	28
	F	0	0	0	0	0	0	0	0	0	2	3	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
Sweep Designated Route for Explosive Hazards	M	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	3
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dig Trench/Unload Captured Arms and Ammo	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	F	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
7-km Hike	M	4	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	6
	F	0	1	9	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10

Section 2. Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for 13 additional 1371 tasks. Annex K contains the descriptive statistics for the remainder of the 1371 tasks. The words “metric” and “task” are used interchangeably throughout this appendix; they both refer to the experimental task.

The two tables below display the results for 13 additional 1371 metrics. Table K B displays the metrics and integration levels with their respective sample sizes, means, and standard deviations. Table K C displays ANOVA and Tukey Test results, including metrics and integration levels, p-values suggesting statistical significance, integration-level elapsed-time differences, and percentage differences between integration levels. For each task, an ANOVA was conducted to compare the three groups and Tukey Tests were conducted to compare each pair of two groups. If non-parametric tests were needed, Table K C displays these results instead of ANOVA and Tukey Test results. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the result was not found to be the same across all three groups. We present basic inferential statistics for three additional 1371 tasks.

Table K B. 1371 Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
7-km Hike	C	20	84.06	6.17	5.07%	10.95%	5.59%
	LD	14	88.32	3.91			
	HD	12	93.26	6.86			
7-km Hike; First km	C	16	9.65	0.46	2.11%	3.21%	1.08%
	LD	10	9.85	0.66			
	HD	8	9.96	0.39			
7-km Hike; Second km	C	16	9.65	0.54	5.80%	6.62%	0.78%
	LD	10	10.21	0.64			
	HD	8	10.29	0.61			
7-km Hike; Third km	C	14	9.31	0.46	5.05%	4.61%	-0.42%
	LD	10	9.78	0.69			
	HD	8	9.74	0.58			
7-km Hike; Fourth km	C	14	9.69	0.43	5.15%	10.80%	5.38%
	LD	10	10.19	0.71			
	HD	8	10.74	1.37			
7-km Hike; Fifth km	C	16	12.73	3.94	14.67%	57.65%	37.49%
	LD	10	14.60	5.72			
	HD	8	20.07	3.90			
7-km Hike; Sixth km	C	15	15.98	4.08	11.85%	-18.60%	-27.22%
	LD	10	17.87	4.78			
	HD	8	13.01	4.03			
7-km Hike; Seventh km	C	15	11.64	1.74	-2.14%	2.23%	4.47%
	LD	10	11.39	0.71			
	HD	8	11.90	0.85			
1-km Hike; by FT	C	42	9.42	1.18	6.29%	9.63%	3.14%
	LD	30	10.01	0.65			
	HD	31	10.33	0.81			
Dig Trench/Unload Captured Arms & Ammo; Dig Trench	C	19	5.07	2.76	23.12%	19.62%	-2.84%
	LD	13	6.24	3.42			
	HD	11	6.07	2.88			
Dig Trench/Unload Captured Arms & Ammo; Place Ordnance in Trench	C	19	5.73	1.92	36.35%	13.59%	-16.69%
	LD	13	7.81	3.17			
	HD	11	6.50	1.87			

Metric	Integration Level	Sample Size	Mean	SD	% Difference (LD-C)	% Difference (HD-C)	% Difference (HD-LD)
Dig Trench/Unload Captured Arms & Ammo; Rig for Detonation	C	19	4.10	1.91	11.76%	20.03%	10.08%
	LD	12	4.59	1.70			
	HD	11	5.05	1.60			
Negotiate Obstacle; by FT	C	41	1.50	0.40	31.87%	52.48%	15.63%
	LD	28	1.97	0.89			
	HD	30	2.28	1.24			

Table K C. Engineers ANOVA and Tukey Test Results.

Metric	F Statistic (df)	F Test P-Value	Comparison	Difference	P-Value	80 % LCB**	80% UCB**	90% LCB**	90% UCB**
7-km Hike	9.58 (2, 43)	< 0.01*	LD-C	4.26	0.10*	0.74	7.78	0.01	8.51
			HD-C	9.20	< 0.01*	5.51	12.89	4.75	13.65
			HD-LD	4.94	0.09*	0.96	8.91	0.14	9.74
1-km Hike; by FT	8.70 (2, 100)	< 0.01*	LD-C	0.59	0.03*	0.20	0.98	0.12	1.06
			HD-C	0.91	< 0.01*	0.52	1.29	0.44	1.37
			HD-LD	0.31	0.40	-0.10	0.73	-0.19	0.82
Negotiate Obstacle; by FT	3.80† (2)	0.15†	LD-C	0.48	0.02†	0.15†	0.55†	0.08†	0.63†
			HD-C	0.76	< 0.01†	0.33†	0.62†	0.30†	0.67†
			HD-LD	0.31	0.28†	-0.02†	0.40†	-0.12†	0.48†

*Indicates there is a statistically significant difference between Integration Levels
**Tukey intervals have familywise confidence of the indicated percentage, each interval is not of the given confidence level on its own.
†Results presented are from Kruskal-Wallis and Mann-Whitney non-parametric tests due to non-normality.

Additional Task Results:

7-km Hike. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.27 for the C group, 0.18 for the LD group, and 0.30 for the HD group.

The C group had a mean time of 84.06 minutes, which is statistically significantly faster than the 88.32-minute average for the LD group and significantly faster than the 93.26-minute average for the HD group. The difference between the LD and HD group is statistically significant, as well. There was a 5.07% degradation in performance from the C to the LD group, and a 10.95% degradation in performance from the C to the HD group. No obvious pattern emerges from looking at the hiking times, by kilometer.

- **Contextual Comments**. The analytical results show that integrated Combat Engineer units would move at a slower rate than an all-male unit. The operational relevance is the potential negative impact on mission accomplishment during a unit moving into position to conduct operations.

1-km Hike; by FT. The data are normally distributed, as evidenced by the Shapiro-Wilk Test that resulted in a p-value of 0.55 for the C group, 0.36 for the LD group, and 0.95 for the HD group.

The C group had a mean time of 9.42 minutes, which is statistically significantly faster than the LD group mean time of 10.01 minutes and also statistically significantly faster than the HD group mean time of 10.33 minutes. On average, the LD group was 6.29% slower than the C group, and the HD group was 9.63% slower than the C group. On average, the HD group was 3.14% slower than the LD group and the difference between the LD and HD group is not statistically significant in a Tukey test.

- **Contextual Comments**. Although operational situations may differ, the time a unit takes to move into position can be the determining factor in the success of the mission. This 1 Km hike movement is meant to be executed quickly and without delay. Therefore, faster results favorable because reaching the objective in a timely manner is critical to the mission

Negotiate Obstacle; by FT. The data are not normally distributed for the HD and LD groups, as evidenced by the Shapiro-Wilk Test p-values less than 0.01. We thus analyze these data using the Kruskal-Wallis and Mann-Whitney-U non-parametric procedures.

The C group had a mean time of 1.50 minutes, the LD group had a mean time of 1.97 minutes, and the HD group had a mean time of 1.97 minutes. Although the Kruskal-Wallis test does not show that the three groups are statistically significantly different at a 10% level, the pairwise comparisons indicate that differences do exist. On average, the LD group was 31.87% slower than the C group and the difference between the LD and C group is statistically significant when compared to a Bonferroni-adjusted significance level of 0.03. On average, the HD group was 52.48% slower than the C group and the difference between the HD and C group is statistically significant when compared to a Bonferroni-adjusted significance level of 0.03. On average, the HD group was 15.63% slower than the LD group and the difference between the LD and HD group is not statistically significant.

- **Contextual Comments**. It is operationally realistic to expect that a Combat Engineer unit will have to negotiate an obstacle such the one provided for this research this rather than breaching or bypassing it. One such example may be a compound, or security wall in a MOUT environment. In most all scenarios, it is more beneficial to negotiate the obstacle in a shorter amount of time in order to maintain movement rate or to reduce exposure time to potential enemy fires.

Annex L.

Mountaineering - Closed MOSs

This annex details the Mountaineering portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment that consisted of Marine volunteers who participated in the closed combat arms MOS (0311, 0331, 0341, 0351, 0352) portion of the experiment. The Mountaineering experiment was executed 4 May – 18 May 2015 at the Leavitt Training Area (LTA), aboard the Mountain Warfare Training Center (MWTC), Bridgeport, CA. The sections outline the Mountaineering assessment (closed MOSs) Scheme of Maneuver (SOM), Limitations, Deviations, and Descriptive and Basic Inferential Statistics.

L.1 Scheme of Maneuver

L.1.1 Experimental Cycle Overview

The Mountaineering assessment (closed MOSs only) of the GCEITF took place aboard MWTC, Bridgeport, CA. The assessment continued over 15 days, in which the closed MOS squads conducted trials every other day. Every Marine received 1 day of recovery after each execution day. All squads consisted of 12 volunteers and a direct-assignment squad leader who was not assessed. Every volunteer was assigned a patch number, which dictated their order in conducting two of the assessed subtasks. The assessment was executed under the supervision of MCOTEA functional test managers and a range officer in charge/range safety officer from the GCEITF unit. The technical sites (gorge crossing and cliff climb) were established each morning by mountaineering subject matter experts, referred to as “red hats,” from the MWTC staff.

L.1.2 Experimental Details

The mountaineering assessment simulated conducting a logistical resupply of a forward-staged squad while moving in a mountainous environment. Each squad departed an assembly area located at the Lower Base Camp (LBC) of MWTC. Four closed MOS squads executed a trial cycle: two control (C) non-integrated squads, and two high-density (HD) integrated squads with six females in each squad.

They hiked 4.6 km with a 75-lb pack and personal weapon (M-4) to an objective rally point (ORP). They then made a short non-assessed movement to a gorge, where they tied a Swiss seat. The second assessed task consisted of the 12-person squad crossing a 200-ft gorge via two single-rope bridges. After all 12 Marines had crossed the gorge, they pulled the packs across and made another short movement to the LTA. The third assessed task consisted of the 12-Marine squad climbing a 40-ft cliff using two climbing lanes. The Marines executed the gorge crossing and cliff climb per a prescribed order. The final assessed task was a 5-km hike back to LBC carrying the same load. At the conclusion of the trial day, the composition of the squad was dissolved and a new sample of males and females were randomly assigned for the following cycle.

L.1.3 Additional Context
Throughout the duration of the assessment, Marines lived in squad bays at LBC. During trial execution, Marines wore/carried prescribed loads. Weighing packs each day prior to the initial hike ensured consistency. After each experimental cycle, the Marines operated under Company leadership, performing minimal physically demanding tasks. Marines who were not part of an assessed squad conducted the same experimental subtasks to ensure equity between individuals participating in a trial cycle and those not chosen for that particular cycle.

L.1.4 Experimental Tasks

L.1.4.1 4.6-km Movement

Primary functions of Marine Corps maneuver elements are shoot, move, and communicate. Movement is a task common to all MOSs and applies to a variety of terrain and climate types. Special consideration must be given to certain factors while moving through a mountainous environment, to include distance, load, pace, technical skill, and special equipment. During the assessment, squads conducted a 4.6-km foot movement with a 75-lb pack and their personal weapon through highly restrictive terrain, starting at the LBC and ending at a gorge. Squads moved as quickly as they could without any Marines falling behind. This task determined the squad's rate of movement while carrying a specified load. Due to the physically demanding nature of moving under load, each Marine took a fatigue and workload survey after completion of the 4.6-km hike (see GCEITF Experimental Assessment Plan [EAP], Annex D).

L.1.4.2 Gorge Crossing

Crossing gorges is a specific challenge when moving through mountainous terrain. Mountain leaders develop the skill and train with the equipment to establish gorge-crossing sites. Once established, Marines must pull themselves across and continue their mission. During the assessment, each squad crossed a 200-ft gorge using two single-rope bridges. Each Marine crossed the gorge one at a time in a specified order using a specific route. Once all Marines had crossed the gorge, a "mule team" (working party) loaded their packs onto the rope bridge and the volunteers pulled them across. A "red hat" instructor provided a safety brief to the volunteers, checked their knots and Swiss seats, hooked up each volunteer to the rope bridge, consistently reset the safety line after each Marine, and supervised the entire gorge-crossing evolution. This task was chosen due to the upper body strength required to pull oneself across the gorge. This task determined the time for a squad to cross a gorge.

L.1.4.3 Cliff Ascent

Climbing a cliff (rock climbing) is another specific challenge when moving through mountainous terrain. Mountain leaders develop the skill and train with the equipment to establish a climbing site. Once the site has been established, Marines must safely and efficiently climb a rock-climbing route and continue their mission. During the assessment, each squad climbed a 40-ft cliff using two climbing routes. Each Marine climbed in a specified order using a specific route. Upon reaching the top of the route, they were lowered to the bottom of the climbing route by a belay Marine. A "red hat" instructor provided a safety brief to the volunteers, checked their knots and Swiss seats, and supervised the climbing and belaying techniques. This task was chosen due to the upper body and lower body strength required to climb a cliff. This task determined the time for a squad to climb a cliff.

L.1.4.4 5-km Movement

During the assessment, squads conducted a 5-km foot movement with a 75-lb pack and their personal weapon through highly restrictive terrain to return from the cliff where they had conducted the climb back to LBC in order to complete the trial.

At the completion of the trial, each Marine took a fatigue and workload survey to assess their fatigue and workload during the execution of the gorge crossing, cliff ascent, and 5-km movement (see GCEITF EAP, Annex D). Marines took a cohesion survey to record their cohesion during the execution of all tasks (see GCEITF EAP, Annex M).

L.1.5 Loading Plan

The loading plan ensured, to the greatest extent possible, equity of physical activity among all volunteers throughout the duration of the experimental assessment. Trials and tasks were conducted in the same manner and sequence to ensure consistency. Due to the number of volunteers, a handful of Marines were not part of an assessed squad each cycle. Collaboration with Company leadership determined that the best method of loading non-assessed Marines was to have them perform the same tasks as an assessed squad in order to experience the same conditions and physical strain. In some instances, the loading Marines formed a quasi-squad and conducted the trial after all the assessed squads were done for the day. When the quasi-squad was too small, the loading Marines were attached to an assessed squad, in which case they operated at the rear of the assessed squad. At no point in time did a loading Marine aid or interfere with an assessed Marine/squad.

L.1.6 Scheme of Maneuver Summary

The mountaineering experiment consisted of a 1-day trial cycle executed every other day that simulated moving through a mountainous environment to conduct a logistical resupply. Each trial consisted of the following tasks: 4.6-km movement, gorge crossing, cliff ascent, and 5-km movement. During the course of the experiment, the closed MOS squads executed five trial cycles.

L.2 Limitations

L.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while unbiased measurements were gathered. Most tasks were performed in a manner similar to the way they would be performed in an operational environment. Under certain situations, artificial limitations or interruptions were introduced that changed or altered the way a task would normally be performed. While these limitations represent a degree of artificiality, they do not detract significantly from our abilities to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for the Mountaineering assessment.

L.2.2 Relative Difficulty of Record Test

The Mountaineering assessment was designed to gather data associated with some of the most physically demanding tasks that Marines of all MOSs could perform in a mountainous environment. Marines are required to conduct continuous operations in a variety of environments and climates, to include mountainous terrain. The difficulty of the assessed tasks in isolation do not fully replicate life experienced by a unit during a Mountain Training Exercise (MTX) or combat operations in a mountainous location. Due to the limited amount of time available to conduct the assessment, only selected mountaineering tasks were assessed. Due to specific experimental constraints and consideration of human factors, tasks/duties outside of the assessment were minimized. During a typical field exercise, it is common for Marines to conduct 24-hour operations that include performing daytime and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. The Marines in the experiment were not required to bivouac in the elements, but rather slept in squad bays every night. Outside of the assessed trials, there were minimal tasks required of the volunteers that demanded any degree of physical strain.

Lastly, the Marines only conducted trials every other day, with a day-on/day-off schedule. This artificial recovery period (or delay in operations) is not realistic when conducting training or combat operations.

L.2.3 Experimental Timing Requirements

Another primary concern in designing the mountaineering assessment was throughput for the Marines. Only four squads were supportable, given the time and space requirements of the SOM and the limited amount of daylight. Initially, it took a single squad approximately 5 hours to complete a single trial. To prevent a squad from influencing another, there was a 1.5-hour delay between squads.

L.2.4 Execution Artificialities

Several artificialities were present as the volunteers completed the SOM. Once the squad completed the initial 4.6-km hike, the unit arrived at the gorge to an already-established high-tension crossing site to expedite the throughput. This is realistic in the sense that a commander or unit leader, if supportable, could send an element forward of the main body and have the technical sites established. There was a fire team of non-volunteers emplaced at the gorge-crossing site prior to the squad's arrival. This team assisted with loading the individual packs onto the rope system once the squad completed the gorge-crossing movement, allowing the squad to complete the crossing unimpeded and the test team to minimize the random error inherent in the experiment created by the administrative movement.

Another artificiality that existed pertained to the requirement to lower Marines after climbing the cliff. During a typical cliff ascent, once Marines reach the top of the cliff they climb over the crest and continue with the mission. Due to experimental and safety constraints, after the volunteers topped out, they had another volunteer in the squad belay for them and lower them to the ground. The belay-man's technical skill level affected the time for total execution by the squad. An experienced individual was comfortable lowering the climber back down to the bottom quickly, while an inexperienced one lowered the climber more slowly. These artificialities impacted total performance time.

L.2.5 Limitations Summary

The mountaineering assessment was designed to replicate realistic training in a field environment in mountainous terrain. The endstate was to create an experiment in which the volunteers felt they were conducting realistic and operationally relevant tasks. Certain unavoidable limitations to both the assessed tasks and the non-assessed operating environment introduced a level of artificiality that would not normally have been present in a field training or combat environment.

L.3 Deviations

L.3.1 Gorge Crossing

During initial planning and coordination with MWTC, it was established that all squads would utilize one high-tension system to cross the gorge. However, during the "dry runs," the MWTC instructors and subject matter experts, or "red hats," advised MCOTEA and the GCEITF leadership that they could establish two high-tension systems at the gorge-crossing site. The high-tension systems would be the same distance and level of difficulty, with minimum variations. Since there was an initial concern about how long the entire evolution would take, MCOTEA decided to approve the deviation in order to gain efficiencies in the overall time and reduce the likelihood of a backlog at the gorge.

Additionally, the packs that the Marines carried to the gorge crossed the high-tension rope bridge. After all Marines crossed the gorge and the assessment time stopped, there was an administrative movement in which the packs were tied together and loaded for the squad to pull across. Groups of three Marines pulled three packs across in order to ensure that all Marines were under equal loading conditions.

L.3.2 Failure Criteria for the Gorge Crossing

Annex D of the Experimental Test Plan is the Data Collection Execution Matrix. As written, the failure criteria for the gorge-crossing portion of the assessment stated that failure would occur once the unit took 2 hours to complete the crossing. This correlated to each Marine in the 12-Marine squad having 10 minutes to complete the gorge crossing, at which point the “red hat” would pull the Marine to the far side of the gorge. After discussion with the MWTC staff, the consensus was that if it took a Marine more than 5 minutes to cross the gorge, there would be a safety concern requiring the “red hat” to facilitate completion. It was agreed that 5 minutes was acceptable for an individual failure on the gorge crossing. Additionally, the overall time for the squad to complete the gorge would be reduced to 1 hour for squad task failure.

L.3.3 Concentration of Squad

The Test Plan stated that MCOTEA would assess three levels of integration for the Rifle Squad: the control group, the low-integration group, and the high-integration group. However, before the record trial cycle began, the ITF population could no longer support a low-integration group. The Infantry test team decided to eliminate the low-concentration group and began the record trial cycle on 4 May 2015 with the control group and the high-concentration group of six females per squad.

L.3.4 Two-Day Test Event

Based on the number of Marines and squads that were expected to train at Bridgeport, the initial plan outlined a 3-day cycle to support 12 squads. The plan was for each squad to execute the trial once every 3 days. However, due to volunteer attrition by the commencement of the Bridgeport assessment, the task force was only able to fill four closed MOS squads. This reduction allowed for the trials to be conducted over a 2-day cycle (1 day for open MOS squads, and 1 for closed MOS squads).

L.4 Data Set Description

L.4.1 Data Set Overview

The closed MOS portion of the Bridgeport experiment consisted of 1 pilot trial cycle day and 13 trial cycles. The pilot trial was conducted on 4 May 2015. Because the pilot trial data and execution did not deviate from record trial execution, we use all pilot trial data for analysis. The final data set includes data collected 4 May – 18 May 2015.

L.4.2 Record Test Volunteer Participants

At the beginning of the first pilot trial, there were 43 male closed MOS volunteers and 19 female closed MOS volunteers. All Marines who began this portion of the experiment completed it.

L.4.3 Planned, Executed, and Analyzed Trial Cycles

Table L-1 displays number of trial cycles planned, executed, and analyzed by task. The planned number of trial cycles for the closed MOS portion of Bridgeport per Section 7.5.3 of GCEITF EAP is 60 trials, or 20 trials per planned integration level (C, low-density, and HD). Due to the number of Marines who voluntarily withdrew or were involuntarily withdrawn from the experiment prior to the execution of the first record trial cycle, only two squads of the C and HD integration levels remained.

The planned number of trial cycles in Table L-1 reflects 14 planned trial cycles for each integration level. There were no trials on 7 May – 9 May 2015 or on 15 May 2015, resulting in the loss of four trials per integration level. We used 1 make-up day to make up two of those trials for each integration level.

Table L-1. Bridgeport Closed MOS Planned, Executed, and Analyzed Trial Cycles

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
4.6-km Hike	C	14	12	12	Did not execute 4 May due to weather
	HD	14	13	13	
5-km Return Hike	C	14	12	12	Did not execute 4 May due to weather
	HD	14	12	12	
Gorge Crossing	C	14	12	12	Did not execute 4 May due to weather
	HD	14	12	12	
Cliff Ascent	C	14	11	11	Did not execute 4 May due to weather
	HD	14	12	12	
4.6-km Hike; 1km Time	C	14	12	10	Missing data: 17 May
	HD	14	13	11	Missing data: 17 May
4.6-km Hike; 2km Time	C	14	12	10	Missing data: 17 May
	HD	14	13	11	Missing data: 17 May
4.6-km Hike; 3km Time	C	14	12	10	Missing data: 17 May
	HD	14	13	11	Missing data: 17 May
4.6-km Hike; 4km Time	C	14	12	10	Missing data: 17 May
	HD	14	13	11	Missing data: 17 May
4.6-km Hike; 5km Time	C	14	12	10	Missing data: 17 May
	HD	14	13	11	Missing data: 17 May
5-km Return Hike; 1km Time	C	14	12	10	Missing data: 17 May
	HD	14	12	10	Missing data: 17 May
5-km Return Hike; 2km Time	C	14	12	10	Missing data: 17 May
	HD	14	12	10	Missing data: 17 May
5-km Return Hike; 3km Time	C	14	12	10	Missing data: 17 May
	HD	14	12	10	Missing data: 17 May
5-km Return Hike; 4km Time	C	14	12	10	Missing data: 17 May
	HD	14	12	10	Missing data: 17 May

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
5-km Return Hike; 5km Time	C	14	12	10	Missing data: 17 May
	HD	14	12	10	Missing data: 17 May

L.5 Descriptive and Basic Inferential Statistics

L.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common infantry tasks and are indicative of unit-level proficiency during either field exercises or combat operations. Four tasks out of 14 are presented in this section. Bridgeport Closed MOS Appendix contains the descriptive statistics for the remainder of the Bridgeport tasks. The words "metric" and "task" are used interchangeably throughout this Annex. They both refer to the experimental task.

Each fireteam consisted of four volunteer Marines: the fireteam leader, automatic rifleman, grenadier, and rifleman. Each squad consisted of 12 volunteer Marines (three fireteams) with a direct assignment (nonvolunteer) squad leader. There were two integration levels for all tasks. A C group was non-gender-integrated and a HD group was gender integrated with six female Marines.

This section includes experimental results based on descriptive statistics, Mann-Whitney tests, and scatter plots. The subsequent sections will cover each task in detail. Lastly, contextual comments, additional insights, and subjective comments (as applicable) tying back to each experimental task are incorporated.

Caution must be used when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks, such as distances, techniques, leadership, load carried, group size, and group composition.

L.5.2 Bridgeport Closed MOS Selected Tasks Descriptive Statistic Results

The two tables below display the results for the four selected Bridgeport closed MOS metrics. Table L-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations. Table L-3 displays Mann-Whitney results, including metrics and integration levels, p-values suggesting statistical significance, elapsed-time differences between integration levels, and percentage differences between integration levels. For each task, a Mann-Whitney test (due to the small sample size) was conducted to compare the two groups. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the mean time or percent hits for the HD group is different from that for the C group.

Table L-2. Bridgeport Closed MOS Selected Task Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD
4.6-km Hike (minutes)*	C	12	59.39	4.97
	HD	13	70.36	6.34
5-km Return Hike (minutes)*	C	12	56.45	6.07
	HD	12	67.59	6.56
Gorge Crossing (minutes)*	C	12	16.62	2.64
	HD	12	19.51	3.40
Cliff Ascent (minutes)*	C	11	34.92	7.46
	HD	12	40.69	9.61

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to a Mann-Whitney test.

Table L-3. Bridgeport Closed MOS Mann-Whitney Test Results

Metric	Comparison	Difference	% Difference	2-sided P-Value	1-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
4.6-km Hike (minutes)	HD-C	10.97	18.47%	< 0.01*	<0.01*	7.56	13.85	6.58	14.93
5-km Return Hike (minutes)	HD-C	11.14	19.74%	< 0.01*	<0.01*	6.86	15.25	6.27	15.59
Gorge Crossing (minutes)	HD-C	2.89	17.37%	0.01*	0.01*	1.43	4.00	0.91	4.32
Cliff Ascent (minutes)	HD-C	5.77	16.51%	0.12	0.06*	0.87	10.14	-0.39	12.02

*Indicates there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to the Mann-Whitney test.

L.5.2.1 4.6-km Hike

L.5.2.1.1 4.6-km Hike Overview

This experimental task assessed a squad of 12 Marines moving 4.60 km while each Marine carried a 75-lb pack (resupply load) and an M-4. The route for this movement was on an unimproved surface and very hilly; the terrain was hard and rocky. This movement consisted primarily of elevation gain (uphill) as shown in Figure L-1. The recorded time for this task started when the squad departed the LBC start point and stopped when the squad arrived at the gorge-crossing site. Each squad moved as fast as the slowest person and could take as many breaks as necessary.

Figure L-1. 4.6-km Hike Elevation Profile

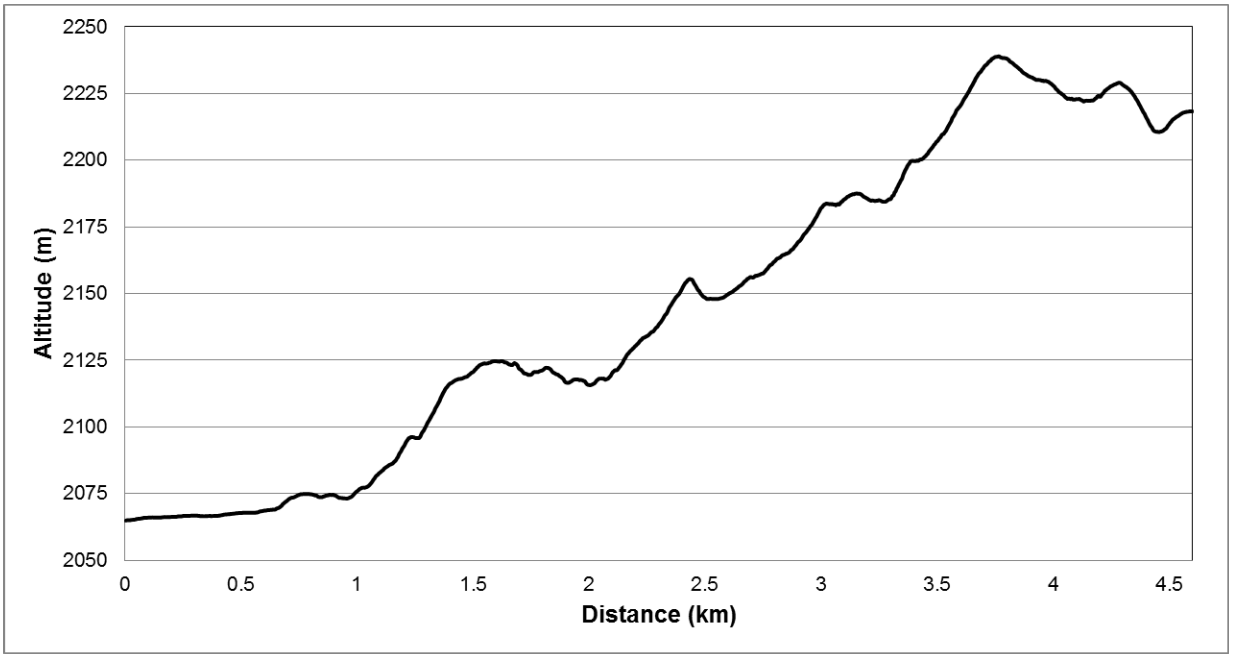
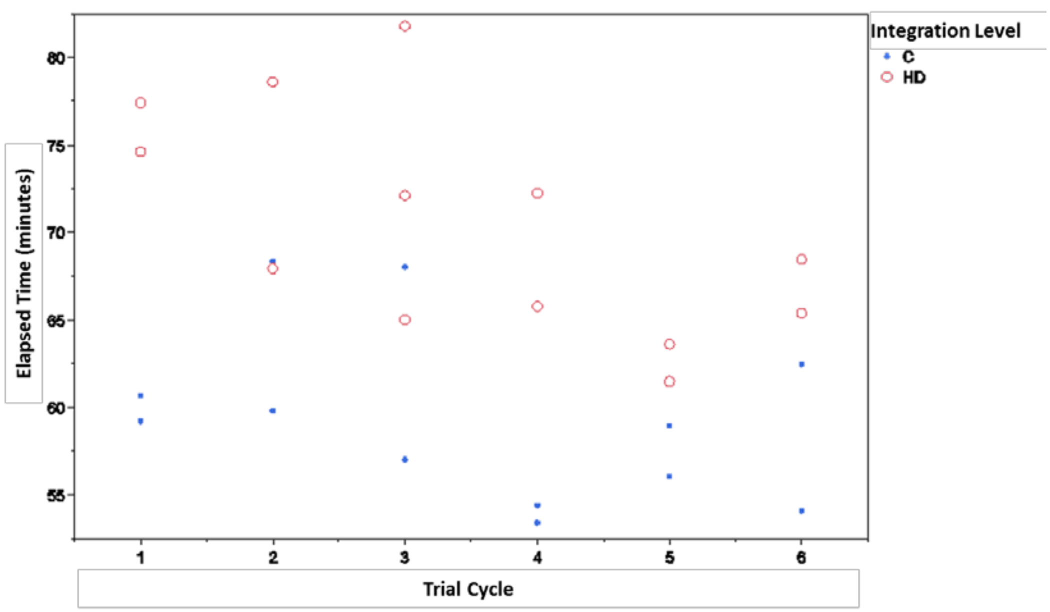


Figure L-2 displays all closed MOS 4.6-km hike time data. All data on the scatter plot are valid for analysis.

Figure L-2. 4.6-km Hike



The C group had a mean time of 59.39 minutes. This time is statistically significantly faster than the HD mean time of 70.36 minutes. This difference results in an 18.47%, or a 10.97-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 1.37-minute increase in standard deviation (SD) (4.97 minutes for the C group and 6.34 minutes for the HD group).

L.5.2.1.2 4.6-km Hike Contextual Comments

The difference in hike times is relevant to both the training and combat environments as it will take integrated squads more time to conduct foot marches. Per the tactical march standards noted in the 0311 Descriptive Statistics, the Marine Corps standard of hiking is 4.0 km/h. The HD group failed to meet this standard. The C group’s average pace was 4.65 km/h and the HD group’s average pace was 3.92 km/h. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), the HD group would finish 48 minutes behind the C group. Furthermore, on any given day (under the same environmental conditions), the fastest group was always a C group and the slowest group was always an HD group.

L.5.2.1.3 4.6-km Hike Additional Insights

Based on the USMC standard of a 4.0-km/h pace over a 4.6-km route (which would result in a 69-minute 4.6-km completion time), the HD groups was 1.36 minutes slower than that standard. In a battlefield situation, in which speed is essential, this delay is advantageous for the enemy. Marine Corps Doctrinal Publications (MCDPs) consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to “Being Faster” and states, “Physical speed, moving more miles per hour, is a powerful weapon in itself.” MCDP-6 Command and Control also speaks to speed relative to the enemy and states, “The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results.”

L.5.2.1.4 4.6-km Hike Subjective Comments

For subjective comments relating to this task, see the BP Closed-MOS Appendix.

L.5.2.2 5.0-km Return Hike

L.5.2.2.1 5.0-km Return Hike Overview

This experimental task assessed a squad of 12 Marines moving 5.0 km while each Marine carried a 75-lb pack (resupply load) and an M-4. The route for this movement was on an unimproved surface and very hilly; the terrain was hard and rocky. This movement consisted primarily of elevation loss (downhill) as shown in Figure L-3. The recorded time for this task started when the squad departed the LTA and stopped when the squad arrived at the LBC. Each squad moved as fast as the slowest person and could take as many breaks as necessary.

Figure L-3. 5.0-km Return Hike Elevation Profile

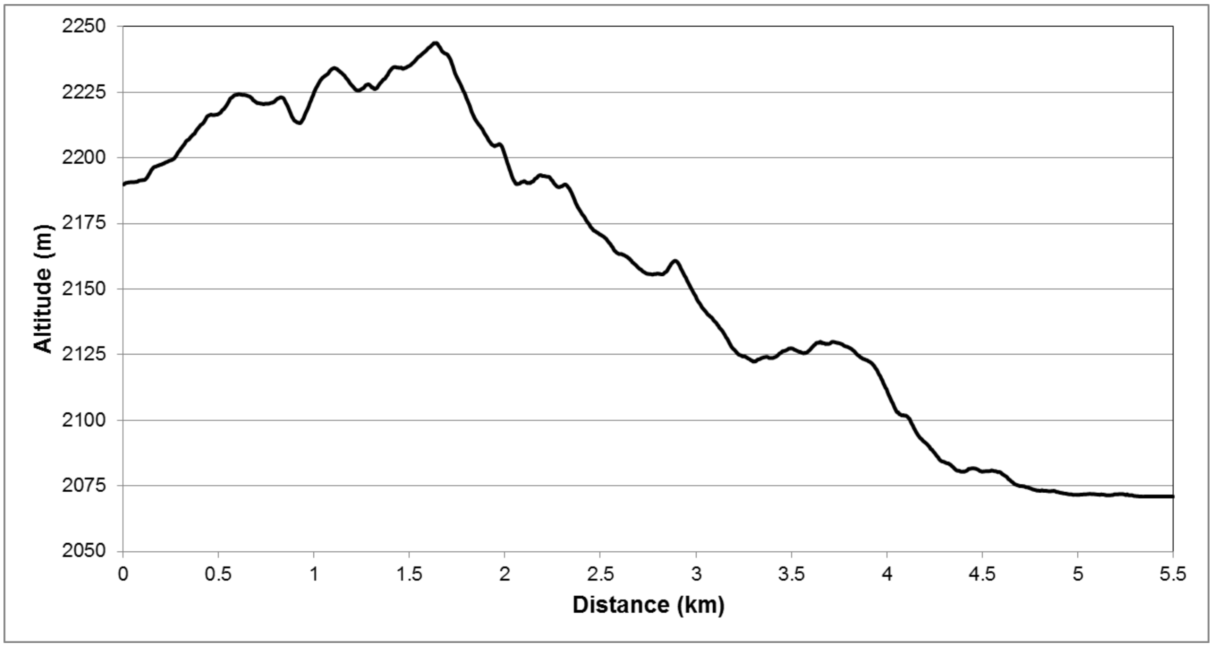
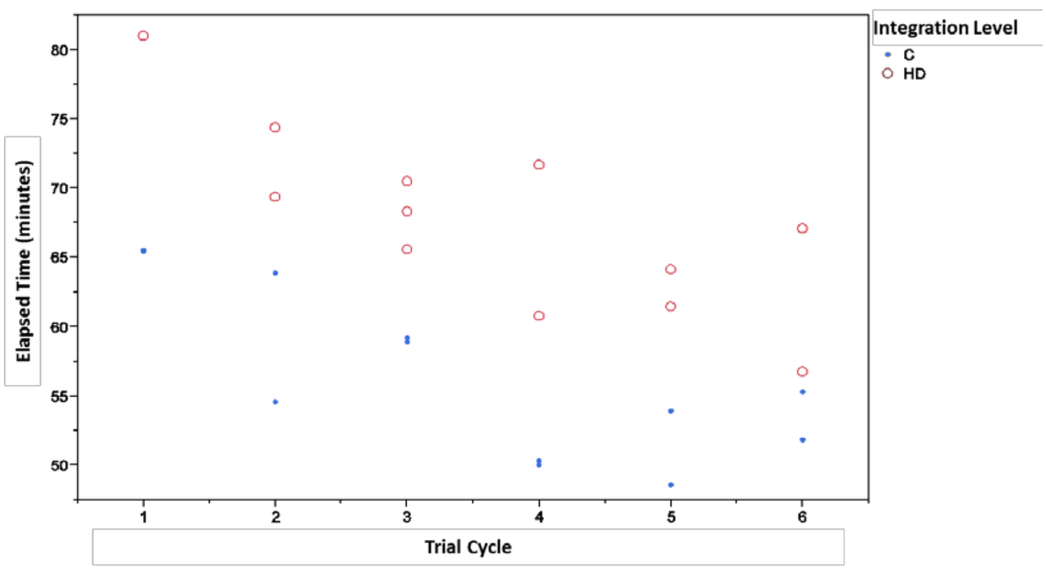


Figure L-4 displays all Bridgeport closed MOS 5.0-km hike time data. All data on the scatter plot are valid for analysis.

Figure L-4. 5.0-km Hike



The C group had a mean time of 56.45 minutes. This time is statistically significantly faster than the HD mean time of 67.59 minutes. This difference results in a 19.74%, or an 11.14-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 0.49-minute increase in SD (6.07 minutes for the C group and 6.56 minutes for the HD group).

L.5.2.2.2 5.0-km Return Hike Contextual Comments

The difference in hike times is relevant to both the training and combat environments as it will take integrated squads more time to conduct foot marches. Per the tactical march standards noted in the 0311 Descriptive Statistics, the Marine Corps standard of hiking is 4.0 km/h. Both the C group and HD group met this standard. The C group's average pace was 5.31 km/h and the HD group's average pace was 4.44 km/h. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), the HD group would finish 44.3 minutes behind the C group. Furthermore, on any given day (under the same environmental conditions), the slowest C group was faster than the fastest HD group.

L.5.2.2.3 5.0-km Return Hike Additional Insights

MCDPs consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to "Being Faster" and states, "Physical speed, moving more miles per hour, is a powerful weapon in itself." MCDP-6 Command and Control also speaks to speed relative to the enemy and states, "The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results."

L.5.2.2.4 5.0-km Return Hike Subjective Comments

For subjective comments relating to this task, see the BP Closed-MOS Appendix.

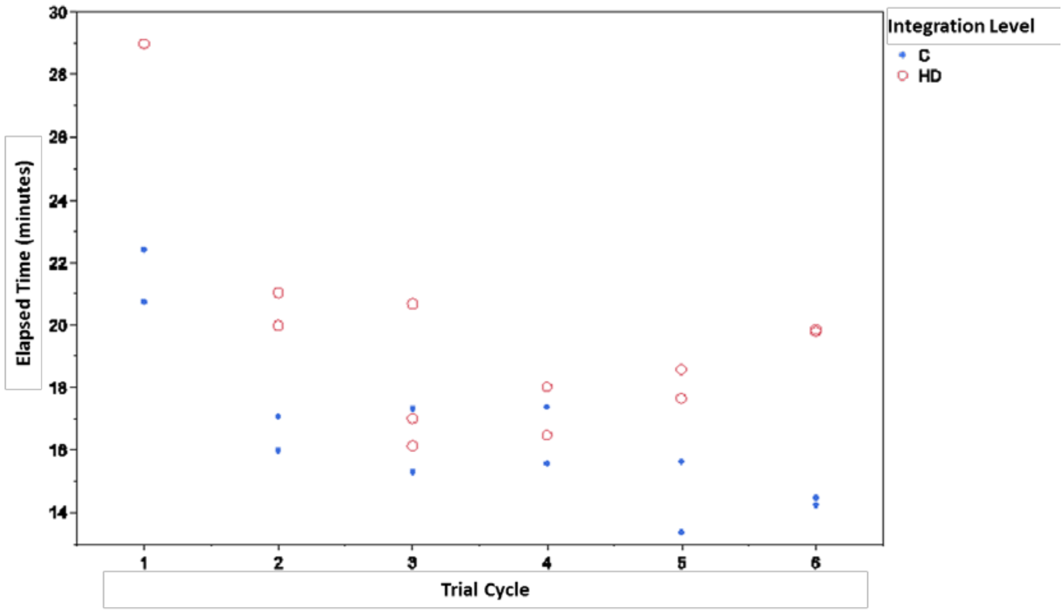
L.5.2.3 Gorge Crossing

L.5.2.3.1 Gorge Crossing Overview

This experimental task assessed a squad of 12 Marines moving across two 200-foot single-rope bridges. A mountaineering expert ("red hat") was the only one authorized to verify knots and hook up each Marine to the rope bridge. Crossing the gorge was an individual task, but the recorded time was for the squad as a whole. After each individual reached the far side of the gorge, a squad leader reset each lane by pulling the tag-line back to the near side of the gorge for the next Marine. The recorded time for this task started when the first Marine of the squad touched the rope bridge and stopped when last member of the squad was completely disconnected from the rope bridge on the far side.

Figure L-5 displays all Bridgeport closed MOS gorge-crossing time data. All data on the scatter plot are valid for analysis.

Figure L-5. Gorge Crossing



The C group had a mean time of 16.62 minutes. This time is statistically significantly faster than the HD mean time of 19.51 minutes. This difference results in a 17.37%, or a 2.89-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 0.76-minute increase in SD (2.64 minutes for the C group and 3.40 minutes for the HD group).

L.5.2.3.2 Gorge Crossing Contextual Comments

Sun Tzu stated, “Speed is the essence of war. Take advantage of the enemy’s unpreparedness; travel by unexpected routes and strike him where he has taken no precautions” (MCDP-1, p. 69). Gaining this advantage in a mountainous environment may involve crossing a gorge. Crossing a single-rope bridge by pulling oneself across is a physical strain on the upper body. While no purely objective standard can be set for crossing a gorge, any decrement in speed translates into increased exposure to enemy fires and greater risk for friendly casualties. On average, the HD group took 2.89 minutes longer than the C group. Furthermore, on any given day (under the same environmental conditions), the C group was always the fastest squad to finish and the HD group was always the slowest squad to finish.

L.5.2.3.3 Gorge-Crossing Additional Insights

A purely objective evaluation of 2.89-minute difference is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated unit. Considering an enemy unit moving at 4.0 km/h on foot to defend a piece of key terrain, the enemy would be able to move an additional 193 meters as the HD group was still crossing the gorge. Far worse would be if an enemy observer could call for fire with mortars, artillery, or aviation assets against the integrated group, given the time delay.

Additionally, one source of masking occurred during this task that was not anticipated during the design. The physically demanding aspect of this task involved pulling oneself across the gorge, but the reset time was generally twice that of the execution time. For instance, many Marines were able to cross the gorge within 60 seconds; however, it consistently took approximately 2.5 minutes to unhook a Marine, pull the tag-line back to the near side, and hook up the next Marine. Since six Marine went across each line, a time of approximately 6 minutes was devoted to Marines crossing the gorge while a time of approximately 12.5 minutes was devoted to the reset process. Strictly considering the performance of each Marine would magnify the results and reveal the gender difference as being even greater.

L.5.2.3.4 Gorge Crossing Subjective Comments

For subjective comments relating to this task, see the BP Closed-MOS Appendix.

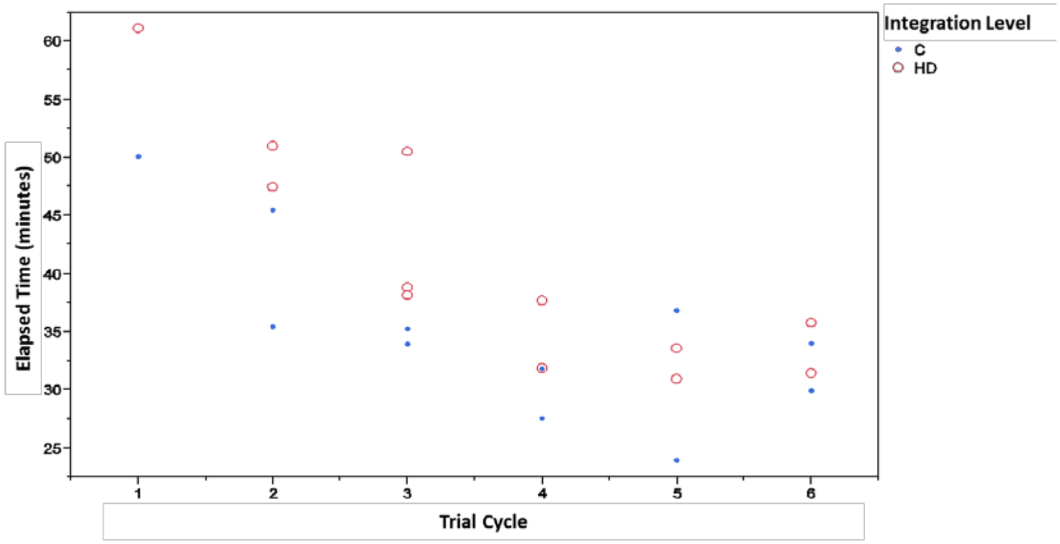
L.5.2.4 Cliff Ascent

L.5.2.4.1 Cliff Ascent Overview

This experimental task assessed a squad of 12 Marines climbing two 40-foot technical rock-climbing routes. A mountaineering expert (“red hat”) was the only one authorized to verify knots and hook up each Marine to the belay line. Climbing the cliff was an individual task, but the recorded time was for the squad as a whole. After each individual reached the top of the cliff, a fellow Marine on belay lower the climber to the ground. The recorded time for this task started when the first Marine of the squad touched the rope and stopped when last member of the squad was completely disconnected from the rope.

Figure L-6 displays all Bridgeport closed MOS cliff ascent time data. All data on the scatter plot are valid for analysis.

Figure L-6. Cliff Ascent



The C group had a mean time of 34.92 minutes. This time is statistically significantly faster than the HD mean time of 40.69 minutes. This difference results in a 16.51%, or a 5.77-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 2.51-minute increase in SD (7.46 minutes for the C group and 9.61 minutes for the HD group).

L.5.2.4.2 Cliff Ascent Contextual Comments

Sun Tzu stated, “Speed is the essence of war. Take advantage of the enemy’s unpreparedness; travel by unexpected routes and strike him where he has taken no precautions” (MCDP-1, p. 69). Gaining this advantage in a mountainous environment may involve climbing a cliff. Technical rock-climbing is a physical strain on the upper and lower body. While no purely objective standard can be set for climbing a cliff, any decrement in speed translates into an advantage to the enemy and a greater risk of friendly casualties. On average, the HD group took 5.77 minutes longer than the C group. Furthermore, on any given day (under the same environmental conditions), the C group was always the fastest squad to finish and, with the exception of one trial, the HD group was always the slowest squad to finish.

L.5.2.4.3 Cliff Ascent Additional Insights

A purely objective evaluation of 5.77 minutes is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated unit. Considering an enemy unit moving at 4.0 km/h on foot to defend a piece of key terrain, the enemy would be able to move an additional 385 meters as the HD group was still climbing the cliff. Far worse would be if an enemy observer could call for fire with mortars, artillery, or aviation assets against the integrated group, given the time delay.

The obvious "learning curve" (trial cycles 1 and 2) that occurred can be explained in two ways: familiarity of the climbing route and the masking effect of belaying. First, there were two separate but similar routes that Marines had to climb. Each subsequent opportunity to climb involved learned techniques/behaviors that resulted in a faster climb time. Second, a source of masking occurred during this task that was not anticipated during the design. The physically demanding aspect involved climbing the cliff; however, the belay/lowering time was a significant part of the overall time. The belay time also got drastically shorter as the assessment progressed, because belay Marines became more comfortable with this skill. Strictly considering the performance of each Marine would have better informed the gender difference involved in this task.

L.5.2.4.4 Cliff Ascent Subjective Comments

For subjective comments relating to this task, see the BP Closed MOS Appendix.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix to Annex L
Mountaineering Closed MOSs Supplemental Information

This appendix provides supplemental information for the Mountaineering Closed MOSs portion of the GCEITF experiment. It provides information regarding the GCEITF leadership subjective comments and additional descriptive statistics not described in Annex L.

Section 1. GCEITF Leadership Subjective Comments

The GCEITF leadership provided comments on their observations of the experiment throughout its execution. Table L A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table L A – Summary of GCEITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement			Requesting extra breaks			Requires extra assistance			Needs no assistance			Compensating for another Marine			Gear pass off			Other			No category			Total
		C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	
4.6-km Hike	M	16	3	19	15	1	16	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	37
	F	0	34	34	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	38
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
5-km Return Hike	M	3	1	4	0	0	0	3	0	3	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	8
	F	0	17	17	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	20
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	3	0	0	0	3
Gorge Crossing	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	0	0	5	5	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	6
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cliff Ascent	M	0	0	0	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	F	0	0	0	0	0	0	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1

Section 2. Additional Task Basic Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for 10 additional Closed MOS tasks. Annex L contains the descriptive statistics for the remainder of the Closed MOS tasks. The words “metric” and “task” are used interchangeably throughout this Appendix. They both refer to the experimental task.

The table below displays the results for 10 additional Closed MOS metrics. Table L B displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent differences between integration levels.

Table L B – Bridgeport 03XX Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
4.6-km Hike; first km (minutes)	C	10	9.54	0.39	11.06%
	HD	11	10.60	0.45	
4.6-km Hike; second km (minutes)	C	10	10.53	0.24	21.82%
	HD	11	12.83	1.03	
4.6-km Hike; third km (minutes)	C	10	12.29	1.14	17.63%
	HD	11	14.45	1.86	
4.6-km Hike; fourth km (minutes)	C	10	17.86	2.83	25.50%
	HD	11	22.41	3.04	
4.6-km Hike; fifth km (minutes)	C	10	9.31	0.90	14.70%
	HD	11	10.68	1.03	
Return Hike; first km (minutes)	C	10	12.05	0.98	21.36%
	HD	10	14.62	1.52	
Return Hike; second km (minutes)	C	10	11.96	2.05	23.43%
	HD	10	14.77	1.61	
Return Hike; third km (minutes)	C	10	10.19	1.50	14.12%
	HD	10	11.63	0.90	
Return Hike; fourth km (minutes)	C	10	10.07	1.29	17.42%
	HD	10	11.83	1.59	
Return Hike; fifth km (minutes)	C	10	12.74	1.28	24.52%
	HD	10	15.87	1.48	

Additional Task Results:

4.6-km Hike by km. The general trend for the 4.6-km hike was that the difference between the HD and C groups increased over the course of the hike. These differences are largest during the points of the largest elevation change.

Return Hike by km. The general trend for the return hike was that the difference between the HD and C groups increased over the course of the hike. These differences are largest during the points of the largest elevation change.

Annex M.

Mountaineering - Open MOSs

This annex details the mountaineering portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment that consisted of Marine volunteers who hold an open MOS (in this experiment, all open MOSs consisted of Marines trained as MOS 1371 Combat Engineers, provisional infantry riflemen, and provisional infantry machine gunners). The mountaineering experiment was executed from 4 May to 18 May 2015 at the Leavitt Training Area (LTA), aboard the Mountain Warfare Training Center (MWTC), Bridgeport, CA. The sections outline the mountaineering assessment (open MOSs) Scheme of Maneuver (SOM), Limitations, Deviations, and Descriptive and Basic Inferential Statistics.

M.1 Scheme of Maneuver

M.1.1 Experimental Cycle Overview

The mountaineering assessment (open MOSs only) of the GCEITF took place aboard MWTC, Bridgeport, CA. The assessment continued over 15 days during which the open MOS squads conducted trials every other day. Every Marine received 1 day of recovery after each execution day. All squads consisted of 12 volunteers and a direct-assignment squad leader who was not assessed. All volunteers were assigned patch numbers, which dictated their order in conducting two of the assessed subtasks. The assessment was executed under the supervision of MCOTEA functional test managers and a range officer in charge/range safety officer from the GCEITF unit. The technical sites (gorge crossing and cliff climb) were established each morning by mountaineering subject matter experts, referred to as “Red Hats,” from the MWTC staff.

M.1.2 Experimental Details

The mountaineering assessment simulated conducting a logistical resupply of a forward-staged squad while moving in a mountainous environment. Each squad departed an assembly area located at the Lower Base Camp (LBC) of MWTC. Four open MOS squads executed a trial cycle: two control (C) nonintegrated squads and two high-density (HD) integrated squads with six females in each squad.

They hiked 4.6 km with 75-lb packs and personal weapons (M-4s) to an objective rally point (ORP). They then made a short nonassessed movement to a gorge where they tied a Swiss seat. The second assessed task consisted of the 12-person squad crossing a 200-ft gorge via two single-rope bridges. After all 12 Marines crossed the gorge, they pulled the packs across and made another short movement to the LTA. The third assessed task consisted of the 12-Marine squad climbing a 40-ft cliff using two climbing lanes. The Marines executed the gorge crossing and cliff climb per a prescribed order.

The final assessed task was a 5-km hike back to LBC carrying the same load. At the conclusion of the trial day, the composition of the squad was dissolved and a new sample of males and females were randomly assigned for the following cycle.

M.1.3 Additional Context

Throughout the duration of the assessment, Marines lived in squad bays at LBC. During trial execution, Marines wore/carried prescribed loads. Weighing packs each day prior to the initial hike ensured consistency. After each experimental cycle, the Marines operated under company leadership, performing minimal physically demanding tasks. Marines who were not part of an assessed squad conducted the same experimental subtasks to ensure equity between individuals participating in a trial cycle and those not chosen for that particular cycle.

M.1.4 Experimental Tasks

M.1.4.1 4.6-km Movement

Primary functions of Marine Corps maneuver elements are shoot, move, and communicate. Movement is a common task to all MOSs and applies to a variety of terrain and climate. Special consideration must be made while moving through a mountainous environment to include distance, load, pace, technical skill, and special equipment. During the assessment, squads conducted a 4.6-km foot movement with 75-lb packs and personal weapons through highly restrictive terrain, starting at the LBC and ending at a gorge. Squads moved as quickly as they were able without any Marines falling behind. This task determined the squad’s rate of movement while carrying a specified load. Due to the physically demanding nature of moving under load, each Marine took a fatigue and workload survey after completion of the 4.6-km hike (see GCEITF Experimental Assessment Plan [EAP], Annex D).

M.1.4.2 Gorge Crossing

Crossing gorges is a specific challenge when moving through mountainous terrain. Mountain leaders develop the skill and train with the equipment to establish gorge-crossing sites. Once established, Marines must pull themselves across and continue their mission. During the assessment, each squad crossed a 200-ft gorge using two single-rope bridges. Marines crossed the gorge one at a time in a specified order using a specific route. Once all Marines had crossed the gorge, a mule team (working party) loaded their packs onto the rope bridge and the volunteers pulled them across. A Red Hat instructor provided a safety brief to the volunteers, checked their knots and Swiss seats, hooked up each volunteer to the rope bridge, consistently reset the safety line between Marines, and supervised the entire gorge-crossing evolution. This task was chosen for the upper body strength required to pull oneself across the gorge. This task determined the time for a squad to cross a gorge.

M.1.4.3 Cliff Ascent

Climbing a cliff (rock climbing) is another specific challenge when moving through mountainous terrain. Mountain leaders develop the skill and train with the equipment to establish a climbing site. Once established, Marines must safely and efficiently climb a rock-climbing route and continue their mission. During the assessment, each squad climbed a 40-ft cliff using two climbing routes. Each Marine climbed in a specified order using a specific route. Upon reaching the top of the route, he or she was lowered to the bottom of the climbing route by a belay Marine. A Red Hat instructor provided a safety brief to the volunteers, checked their knots and Swiss seats, and supervised the climbing and belaying techniques. This task was chosen for the upper body and lower body strength required to climb a cliff. This task determined the time for a squad to climb a cliff.

M.1.4.4 5-km Movement

During the assessment, squads conducted a 5-km foot movement with 75-lb packs and personal weapons through highly restrictive terrain to return from the cliff where they conducted the climb back to LBC to complete the trial.

At the completion of the trial, each Marine took a fatigue and workload survey to assess their fatigue and workload during the execution of the gorge crossing, cliff ascent, and 5-km movement (see GCEITF EAP, Annex D). Marines took a cohesion survey to record their cohesion during the execution of all tasks (see GCEITF EAP, Annex M).

M.1.5 Loading Plan

The loading plan ensured, to the greatest extent possible, equity of physical activity among all volunteers throughout the duration of the experimental assessment. Trials and tasks were conducted in the same manner and sequence to ensure consistency. Due to the number of volunteers, a handful of Marines were not part of an assessed squad each cycle. Collaboration with company leadership determined that the best method of loading nonassessed Marines was to have them perform the same tasks as an assessed squad to experience the same conditions and physical strain. In some instances, the loading Marines formed a quasi-squad and conducted the trial after all the assessed squads were done for the day. When the quasi-squad was too small, the loading Marines were attached to an assessed squad, in which case they operated at the rear of the assessed squad. At no point in time did a loading Marine aid or interfere with an assessed Marine/squad.

M.1.6 Scheme of Maneuver Summary

The mountaineering experiment consisted of a 1-day trial cycle executed every other day that simulated moving through a mountainous environment to conduct a logistical resupply. Each trial consisted of the following tasks: 4.6-km movement, gorge crossing, cliff ascent, and 5-km movement. During the course of the experiment, the open MOS squads executed four test cycles.

M.2 Limitations

M.2.1 Limitations Overview

The GCEITF experiment was designed to allow operationally relevant tasks to occur as naturally as possible, while gathering unbiased measurements. Most tasks were performed in a manner similar to those in an operational environment. Under certain situations, artificial limitations or interruptions were introduced that altered the way a task would normally be performed. While these limitations represent a degree of artificiality, they do not detract significantly from our abilities to generalize the conclusions of this experiment to the performance of Marines in a field environment. The following limitations were observed for the mountaineering assessment.

M.2.2 Relative Difficulty of Record Test

The mountaineering assessment was designed to gather data associated with some of the most physically demanding tasks that Marines of all MOSs could perform in a mountainous environment. Marines are required to conduct continuous operations in a variety of environments and climates, including mountainous terrain. The difficulty of the assessed tasks in isolation do not fully replicate life experienced by a unit during a Mountain Training Exercise (MTX) or combat operations in a mountainous location. Because of the limited amount of time available to conduct the assessment, only selected mountaineering tasks were assessed. Due to specific experimental constraints and human factor considerations, other tasks/duties outside the assessment were minimized. During a typical field exercise, it is common for Marines to conduct 24-hour operations that include daytime and nighttime operations/patrols, standing firewatch or a security post, and conducting continuing tactical actions. The Marines in the experiment were not required to bivouac in the elements, but rather slept in squad bays every night. Outside the assessed trials, minimal tasks were required of the volunteers that demanded any degree of physical strain.

Lastly, the Marines only conducted trials every other day, with a day on/day off schedule. This artificial recovery period (or delay in operations) is not realistic when conducting training or combat operations.

M.2.3 Experimental Timing Requirements

Another primary concern in designing the mountaineering assessment was throughput for the Marines. Only four squads were supportable given the time and space requirements of the SOM and the limited amount of daylight. Initially, it took a single squad approximately 5 hours to complete a single trial. To prevent one squad from influencing another, there was a 1.5-hour delay between squads.

M.2.4 Execution Artificialities

Several artificialities were present as the volunteers completed the SOM. Once the squad completed the initial 4.6-km hike, the unit arrived at the gorge to an already-established high-tension crossing site to expedite the throughput. This is realistic in the sense that a commander or unit leader, if supportable, could send an element forward of the main body and have the technical sites established. A fire team of nonvolunteers was emplaced at the gorge-crossing site prior to the squad's arrival. This team assisted with loading the individual packs onto the rope system once the squad completed the gorge-crossing movement, allowing the squad to complete the crossing unimpeded and the test team to minimize the random error inherent in the experiment created by the administrative movement.

Another artificiality that existed pertained to the requirement to lower Marines after climbing the cliff. During a typical cliff ascent, once Marines reach the top of the cliff, they climb over the crest and continue with the mission. Due to experimental and safety constraints, after the volunteers topped-out, they had another volunteer in the squad belay for them and lower them to the ground. The belay Marine's technical skill level affected the time for total execution by the squad. An experienced individual was comfortable lowering the climber back down to the bottom quickly, while an inexperienced one lowered the climber more slowly. These artificialities affected total performance time.

M.2.5 Limitations Summary

The mountaineering assessment was designed to replicate realistic training in a field environment in mountainous terrain. The end state was to create an experiment in which the volunteers felt they were conducting realistic and operationally relevant tasks. Certain unavoidable limitations to both the assessed tasks and nonassessed operating environment introduced a level of artificiality that would not normally have been present in a field training or combat environment.

M.3 Deviations

M.3.1 Gorge Crossing

During initial planning and coordination with MWTC, it was established that all squads would utilize one high-tension system to cross the gorge. During the dry runs, however, the MWTC instructors and subject matter experts, or Red Hats, advised MCOTEA and the GCEITF leadership that they could establish two high-tension systems at the gorge crossing site. Both high-tension systems were the same distance and level of difficulty, with minimal variations. Since there was an initial concern about how long the entire evolution would take, MCOTEA decided to approve the deviation to gain efficiencies in the overall time and reduce the likelihood of a backlog at the gorge.

Additionally, the packs that the Marines carried to the gorge crossed the high tension rope bridge. After all Marines crossed the gorge and the assessment time stopped, there was an administrative movement in which the packs were tied together and loaded for the squad to pull across. Groups of three Marines pulled three packs across in order to ensure that all Marines were under equal loading conditions.

M.3.2 Failure Criteria for the Gorge Crossing

Annex D of the Experimental Test Plan is the Data Collection Execution Matrix. As written, the failure criteria for the gorge crossing portion of the assessment stated that failure would occur once the unit took 2 hours to complete the crossing. This correlated to each Marine in the 12-Marine squad having 10 minutes to complete the gorge crossing, at which time the red hat would pull the Marine to the far side of the gorge. After discussion with the MWTC staff, the consensus was that, if it took a Marine more than 5 minutes to cross the gorge, there would be a safety concern requiring the red hat to facilitate completion. It was agreed that 5 minutes was acceptable for an individual failure on the gorge crossing. Additionally, the overall time for the squad to complete the gorge would be reduced to 1 hour for squad task failure.

M.3.3 Concentration of Squad

The Test Plan stated that MCOTEA would assess three different levels of integration for the Rifle Squad: the control group, low-integration, and high-integration. However, prior to the record trial cycle beginning, the ITF population could no longer support a low-integration group. The Infantry test team decided to eliminate the low-concentration group and began the record trial cycle on 5 May 2015 with the control group and high-concentration of six females per squad.

M.3.4 Two-Day Test Event

Based on the number of Marines and squads that were expected to train at Bridgeport, the initial plan outlined a 3-day cycle to support 12 squads. The plan was for each squad to execute the trial once every 3 days. Due to volunteer attrition, however, by the commencement of the Bridgeport assessment, the task force was only able to fill four open MOS squads. This reduction allowed for the trials to be conducted over a 2-day cycle (1 day for open MOS squads and 1 for closed MOS squads).

M.4 Data Set Description

M.4.1 Data Set Overview

The open MOS portion of the Bridgeport experiment consisted of 1 pilot trial cycle day and 13 trial cycles. The pilot trial cycle was conducted on 5 May 2015. Because the pilot trial data and execution did not deviate from record trial execution, we use all pilot trial data for analysis. The final data set includes data collected from 4 May to 18 May 2015.

M.4.2 Record Test Volunteer Participants

At the beginning of the first pilot trial cycle, there were 46 male open MOS volunteers and 17 female open MOS volunteers. All Marines who began this portion of the experiment completed it.

M.4.3 Planned, Executed, and Analyzed Trial Cycles

Table M-1 displays the number of trial cycles planned, executed, and analyzed by task. The planned number of trial cycles for the open-MOS portion of Bridgeport per Section 7.5.3 of GCEITF EAP is 60 trials, or 20 trials per planned integration level (C, LD, and HD). Due to the number of Marines who voluntarily withdrew or were involuntarily withdrawn from the experiment prior to the execution of the first record test, only two squads of the C and HD integration levels remained.

The planned number of trials in Table M-1 reflects 14 planned trial cycles for each integration level. There were no trials on 7–9 May 2015 or on 15 May 2015, resulting in the loss of 4 trials per integration level.

Table M-1. Bridgeport Open MOS Planned, Executed, and Analyzed Trial Cycles

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
4.6-km Hike	C	14	10	10	
	HD	14	10	10	
5-km Return Hike	C	14	10	9	Did not execute 14 May due to weather. Remove 5 May run 2 high outlier
	HD	14	10	10	
Gorge Crossing	C	14	10	10	
	HD	14	10	9	Missing data 17 May
Cliff Ascent	C	14	9	9	Did not execute 14 May due to weather
	HD	14	10	10	
4.6-km Hike; 1-km Time	C	14	10	10	
	HD	14	10	10	
4.6-km Hike; 2-km Time	C	14	10	10	
	HD	14	10	10	
4.6-km Hike; 3-km Time	C	14	10	10	
	HD	14	10	10	
4.6-km Hike; 4-km Time	C	14	10	10	
	HD	14	10	10	
4.6-km Hike; 5-km Time	C	14	10	10	
	HD	14	10	10	
5-km Return Hike; 1-km Time	C	14	10	10	
	HD	14	10	10	
5-km Return Hike; 2-km Time	C	14	10	10	
	HD	14	10	10	
5-km Return Hike; 3-km Time	C	14	10	10	
	HD	14	10	10	
5-km Return Hike; 4-km Time	C	14	10	9	Remove influential point
	HD	14	10	10	
5-km Return	C	14	10	10	

Task and Metric Description	Integration Level	Number of Planned Trials	Number of Trials Conducted	Number of Trials Used in Analysis	Notes
Hike; 5-km Time	HD	14	10	10	

M.5 Descriptive and Basic Inferential Statistics

M.5.1 Descriptive Statistics Overview

Performance metrics were measured during the conduct of common infantry tasks and are indicative of unit-level proficiency during either field exercises or combat operations. This section presents 4 out of 14 tasks. Bridgeport Open MOS Appendix contains the descriptive statistics for the remainder of the Bridgeport tasks. The words “metric” and “task” are used interchangeably throughout this annex. They both refer to the experimental task.

Each fireteam consisted of four volunteer Marines: the fireteam leader, automatic rifleman, grenadier, and rifleman. Each squad consisted of 12 volunteer Marines (three fireteams) with a direct assignment (nonvolunteer) squad leader. There were two integration levels for all tasks. A (C) group was not gender integrated and a HD group was gender integrated with six female Marines.

This section includes experimental results based on descriptive statistics, Mann-Whitney Tests, and scatter plots. The subsequent sections will cover each task in detail. Lastly, contextual comments, additional insights, and subjective comments (as applicable) tying back to each experimental task are incorporated.

Use caution when comparing similar tasks executed by different MOSs within the GCEITF experiment. Comparative analysis may be misleading due to differing factors between MOS tasks, such as distances, techniques, leadership, load carried, group size, and group composition.

M.5.2 Bridgeport Open MOS Selected Task Descriptive Statistics Results

The two tables that follow display the results for the four selected Bridgeport Open MOS metrics. Table M-2 displays the metrics and integration levels with their respective sample sizes, means, and standard deviations. Table M-3 displays Mann-Whitney results, including metrics and integration levels, p-values suggesting statistical significance, integration level elapsed-time differences, and percentage differences between integration levels. For each task, a Mann-Whitney test (due to the small sample size) was conducted to compare the two groups. If p-values are less than the a priori determined significance level of 0.10, we conclude that there is statistical evidence that the mean time or percent hits for the HD group is different from that in the C group.

Table M-2. Bridgeport Open MOS Selected Task Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD

Metric	Integration Level	Sample Size	Mean	SD
4.6-km Hike (minutes)*	C	10	63.38	5.99
	HD	10	70.83	3.61
5-km Return Hike (minutes)*	C	9	61.95	4.15
	HD	10	69.03	5.46
Gorge Crossing (minutes)*	C	10	17.09	1.95
	HD	9	19.76	2.71
Cliff Ascent (minutes)*	C	9	37.78	6.02
	HD	10	45.88	9.26
Cliff Ascent (minutes) (influential point removed)*	C	9	37.78	6.02
	HD	9	43.57	6.06

*Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to a Mann-Whitney test.

Table M-3. Bridgeport Open MOS Mann-Whitney Test Results

Metric	Comparison	Difference	% Difference	2-sided P-Value	1-sided P-Value	80% LCB	80% UCB	90% LCB	90% UCB
4.6-km Hike (minutes)	HD-C	7.45	11.75%	0.01*	0.01*	3.85	10.38	2.85	11.95
5-km Return Hike (minutes)	HD-C	7.08	11.43%	<0.01*	<0.01*	3.70	10.23	2.95	11.82
Gorge Crossing (minutes)	HD-C	2.67	15.62%	0.02*	0.01*	1.00	3.97	0.80	4.25
Cliff Ascent (minutes)	HD-C	8.10	21.45%	0.04*	0.02*	3.20	12.10	1.65	13.18
Cliff Ascent (minutes) (potential influential point removed)	HD-C	5.79	15.34%	0.08*	0.04*	1.73	9.40	0.28	11.05

*Indicates that there is a statistically significant difference, in either a one-sided or two-sided hypothesis test, between integration levels according to the Mann-Whitney test.

M.5.2.1 4.6-km Hike

M.5.2.1.1 4.6-km Hike Overview

This experimental task assessed a squad of 12 Marines moving 4.60 km while each Marine carried a 75-lb pack (resupply load) and an M-4. The route for this movement was on an unimproved surface and very hilly; the terrain was hard and rocky. This movement consisted primarily of elevation gain (uphill) as shown in Figure M-1. The recorded time for this task started when the squad departed the LBC start point and stopped when the squad arrived at the gorge crossing site. Each squad moved as fast as the slowest person and could take as many breaks as necessary.

Figure M-1. 4.6-km Hike Elevation Profile

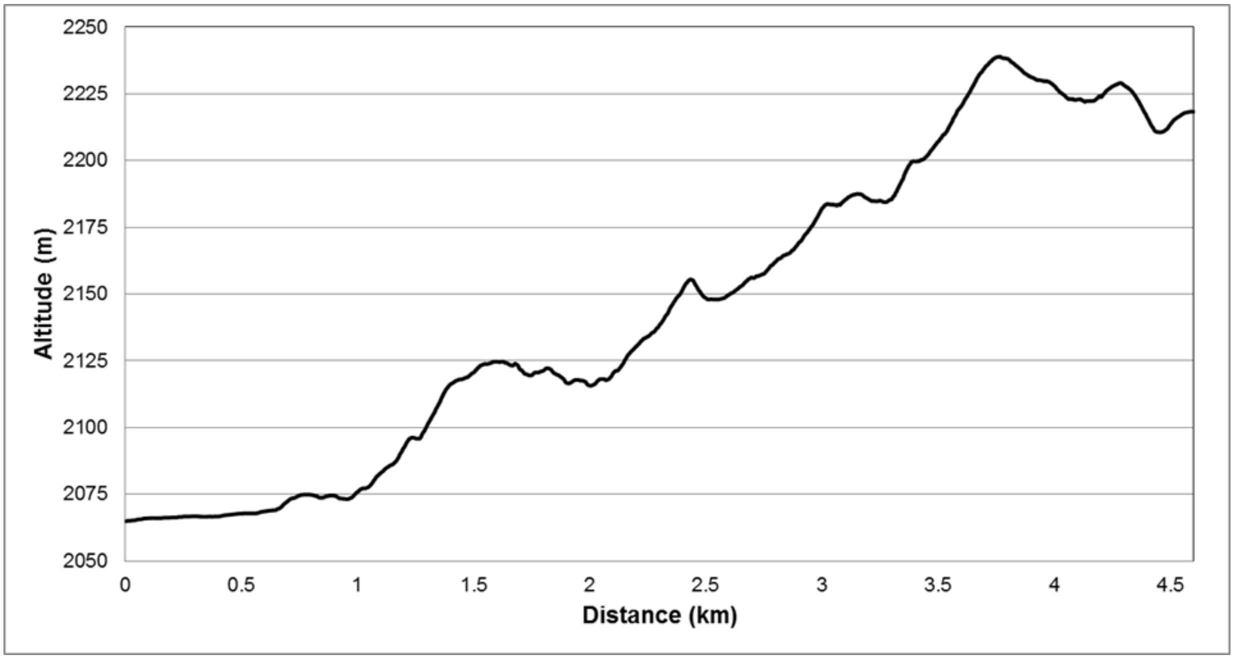
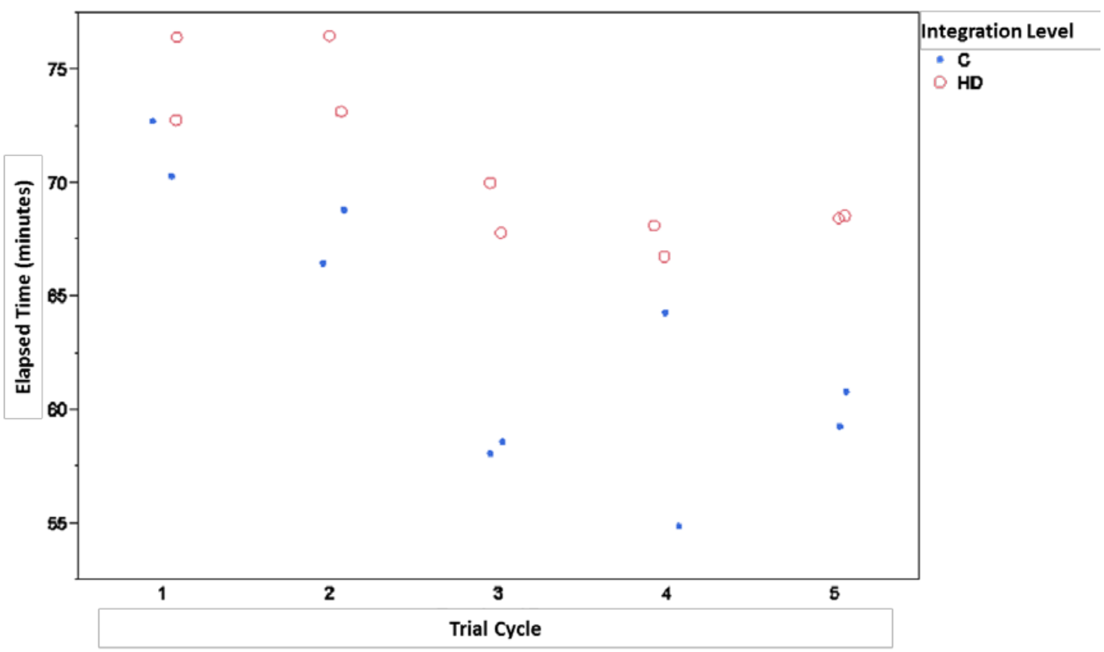


Figure M-2 displays all Bridgeport open MOS 4.6-km hike data. All data on the scatter plot are valid for analysis.

Figure M-2. 4.6-km Hike



The C group had a mean time of 63.68 minutes. This time is statistically significantly faster than the HD mean time of 70.83 minutes. This difference results in an 11.75%, or 7.45-minute, degradation in hike time between the groups. The HD group has less variability, as shown by the 2.38-minute decrease in standard deviation (SD) (5.99

minutes for the C group and 3.61 minutes for the HD group). See Table M-2 and Table M-3 for detailed analytical results.

M.5.2.1.2 4.6-km Hike Contextual Comments

The difference in hike times is relevant to both the training and combat environment as it will take integrated squads more time to conduct foot marches. Per the tactical march standards noted in the 0311 Descriptive Statistics, the Marine Corps standard of hiking is 4.0 km/h. The HD group failed to meet this standard. The average C group pace was 4.65 km/h, and the average HD group's pace was 3.92 km/h. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), the HD group would finish 48 minutes behind the C group. Furthermore, on any given day (under the same environmental conditions), the fastest group was always a C group and the slowest group was always an HD group. Only on one instance did an HG group finish faster than a C group on a given day.

M.5.2.1.3 4.6-km Hike Additional Insights

Based on the USMC standard of a 4.0-km/h pace over a 4.6-km route (which would result in a 69-minute 4.6-km completion time), the HD group was 1.83 minutes slower than that standard. In a battlefield situation, in which speed is essential, this delay is advantageous for the enemy. Marine Corps Doctrinal Publications (MCDPs) consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to "Being Faster" and states, "Physical speed, moving more miles per hour, is a powerful weapon in itself." MCDP-6 Command and Control also speaks to speed relative to the enemy and states, "The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results.

M.5.2.1.4 4.6-km Hike Subjective Comments

For subjective comments relating to this task, see the BP Open-MOS Appendix.

M.5.2.2 5.0-km Return Hike

M.5.2.2.1 5.0-km Return Hike Overview

This experimental task assessed a squad of 12 Marines moving 5.0 km while each Marine carried a 75-lb pack (resupply load) and an M-4. The route for this movement was on an unimproved surface and very hilly; the terrain was hard and rocky. This movement consisted primarily of elevation loss (downhill) as shown in Figure M-3. The recorded time for this task started when the squad departed the LTA and stopped when the squad arrived at the LBC. Each squad moved as fast as the slowest person and could take as many breaks as necessary.

Figure M-3. 5.0-km Return Hike Elevation Profile

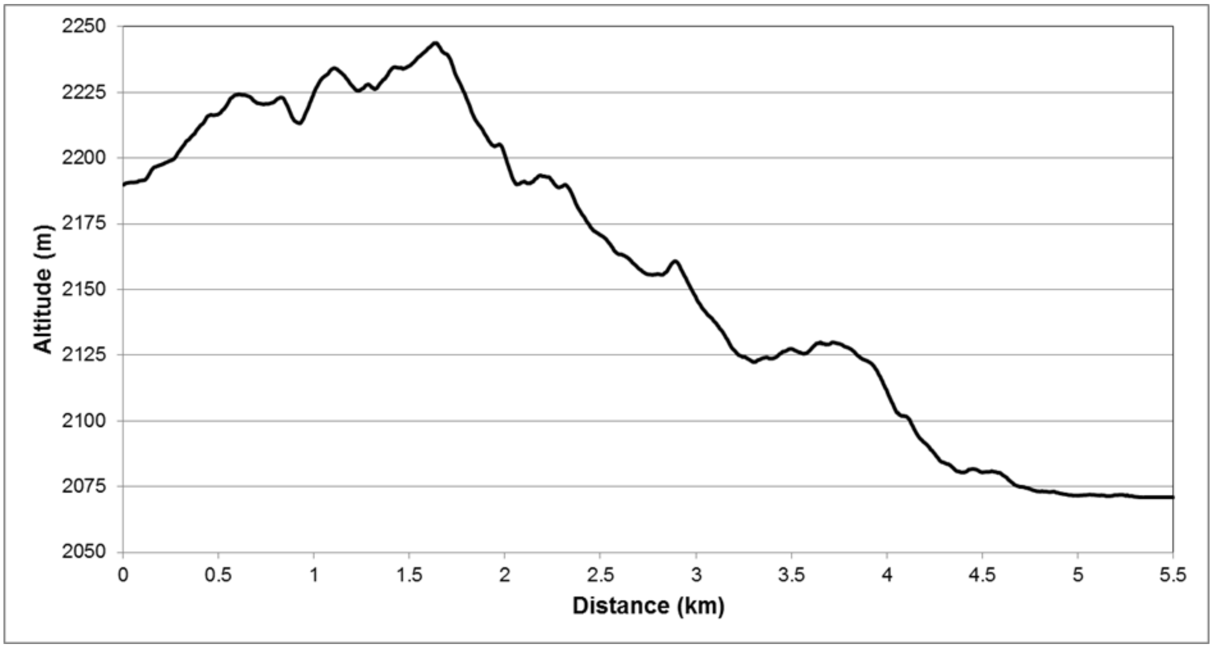
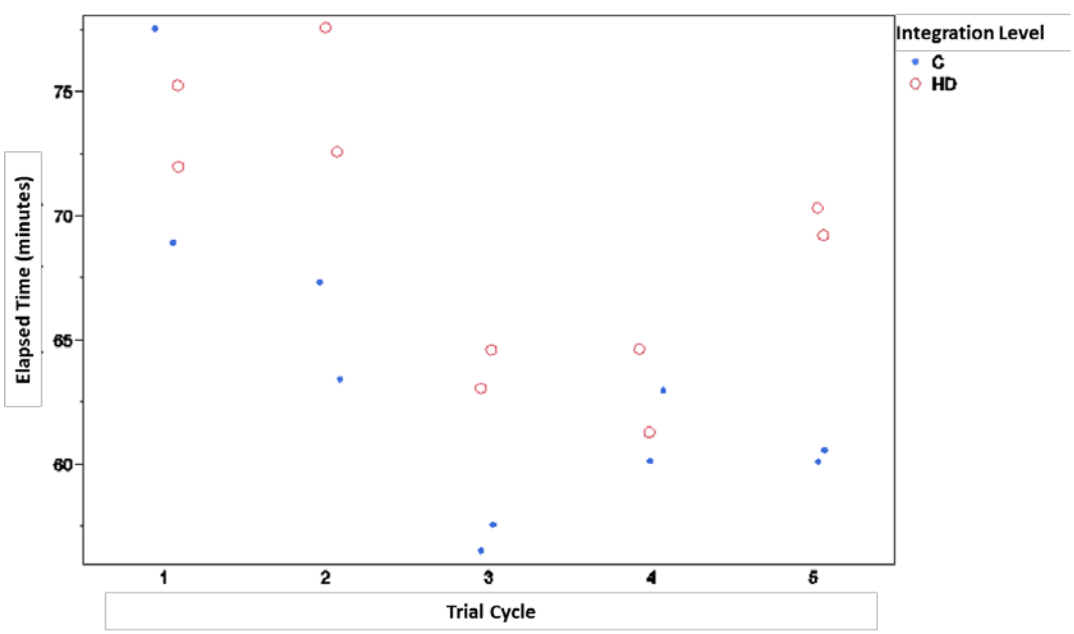


Figure M-4 displays all Bridgeport open MOS 5.0-km return hike data. All data on the scatter plot are valid for analysis.

Figure M-4. 5.0-km Hike



The C group had a mean time of 61.95 minutes. This time is statistically significantly faster than the HD mean time of 69.03 minutes. This difference results in an 11.43%, or 7.08-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 1.31-minute increase in SD (4.15 minutes for the C group and 5.46 minutes for the HD group). See Table M-2 and Table M-3 for detailed analytical results.

M.5.2.2.2 5.0-km Return Hike Contextual Comments

The difference in hike times is relevant to both the training and combat environment as it will take integrated squads more time to conduct foot marches. Per the tactical march standards noted in the 0311 Descriptive Statistics, the Marine Corps standard of hiking is 4.0 km/h. Both the C group and HD group met this standard. The average C group pace was 4.46 km/h, and the average HD group's pace was 4.00 km/h. To extrapolate this pace over a 20-km movement (an optimistic assumption that does not account for any further degradation of performance), the HD group would finish 30.9 minutes behind the C group. Furthermore, on any given day (under the same environmental/weather conditions), the fastest group was always a C group and, with the exception of one trial, the slowest group was always a HD group.

M.5.2.2.3 5.0-km Return Hike Additional Insights

MCDPs consistently emphasize the importance of speed. MCDP 1-3 Tactics devotes an entire chapter to "Being Faster" and states, "Physical speed, moving more miles per hour, is a powerful weapon in itself." MCDP-6 Command and Control also speaks to speed relative to the enemy and states, "The speed differential does not necessarily have to be a large one: a small advantage exploited repeatedly can quickly lead to decisive results."

M.5.2.2.4 5.0-km Hike Subjective Comments

For subjective comments relating to this task, see the BP Open-MOS Appendix.

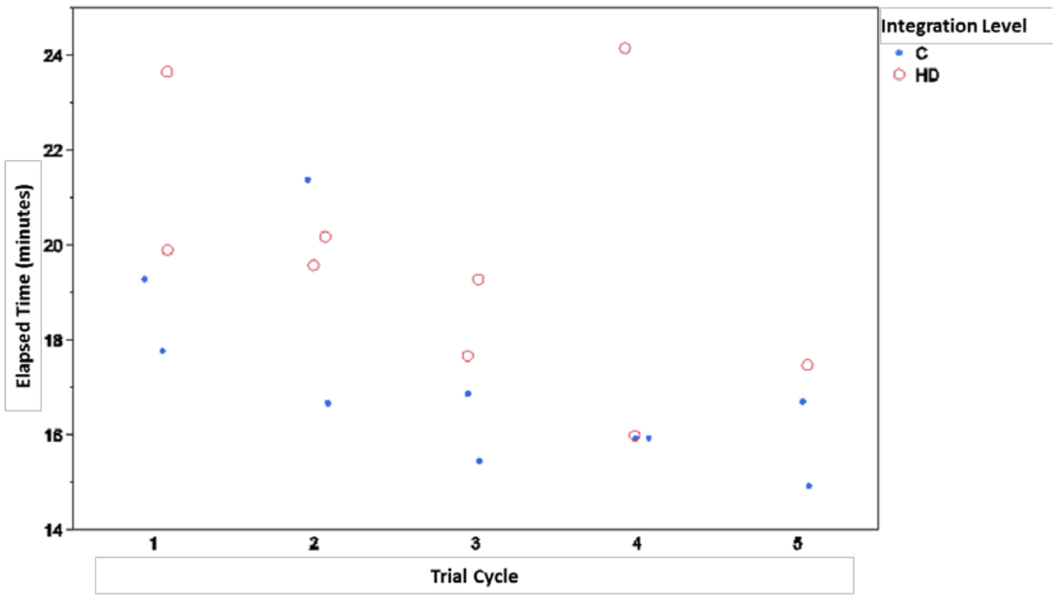
M.5.2.3 Gorge Crossing

M.5.2.3.1 Gorge Crossing Overview

This experimental task assessed a squad of 12 Marines moving across two 200-foot single-rope bridges. A mountaineering expert (a red hat) was the only one authorized to verify knots and hook up each Marine to the rope bridge. Crossing the gorge was an individual task, but the recorded time was for the squad as a whole. After each Marine reached the far side of the gorge, a squad leader reset each lane by pulling the tag-line back to the near side of the gorge for the next Marine. The recorded time for this task started when the first Marine of the squad touched the rope-bridge and stopped when last member of the squad was completely disconnected from the rope-bridge on the far side.

Figure M-5 displays all Bridgeport open MOS gorge crossing data. All data on the scatter plot are valid for analysis.

Figure M-5. Gorge Crossing



The C group had a mean time of 17.09 minutes. This time is statistically significantly faster than the HD mean time of 19.76 minutes. This difference results in a 15.62%, or 2.67-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 1.37-minute increase in SD (1.95 minutes for the C group and 2.71 minutes for the HD group). See Table M-2 and Table M-3 for detailed analytical results.

M.5.2.3.2 Gorge Crossing Contextual Comments

Sun Tzu stated, “Speed is the essence of war. Take advantage of the enemy’s unpreparedness; travel by unexpected routes and strike him where he has taken no precautions” (MCDP-1, p. 69). Gaining this advantage in a mountainous environment may involve crossing a gorge. Crossing a single-rope bridge by pulling oneself across is a physical strain on the upper body. While no purely objective standard can be set for crossing a gorge, any decrement in speed translates into increased exposure to enemy fires and greater risk for friendly casualties. On average, the HD group took 2.67 minutes longer than the C group. Furthermore, on any given day (under the same environmental conditions), the fastest group was always a C group, and, with the exception of one trial, the slowest group was always an HD group.

M.5.2.3.3 Gorge Crossing Additional Insights

A purely objective evaluation of 2.67 minutes is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated unit. Considering an enemy unit moving at 4.0 km/h on foot to defend a piece of key terrain, the enemy would be able to move an additional 178 meters as the HD group was still crossing the gorge. Far worse would be an enemy observer who could call for fire with mortars, artillery, or aviation assets against the integrated group given the time delay.

Additionally, one source of masking occurred during this task that was not anticipated during the design. The physically demanding aspect of this task involved pulling oneself across the gorge; however, the reset time was generally twice that of the execution time. For instance, many Marines were able to cross the gorge within 60 seconds. However, it consistently took approximately 2.5 minutes to unhook a Marine, pull the tag-line back to the near-side, and hook up the next Marine. Since six Marines went across each line, approximately 6 minutes were devoted to Marines crossing the gorge while approximately 12.5 minutes were devoted to the reset process. Strictly considering the performance of each Marine would magnify the results and make the gender-difference even greater.

M.5.2.3.4 Gorge Crossing Subjective Comments

For subjective comments relating to this task, see the BP Open-MOS Appendix.

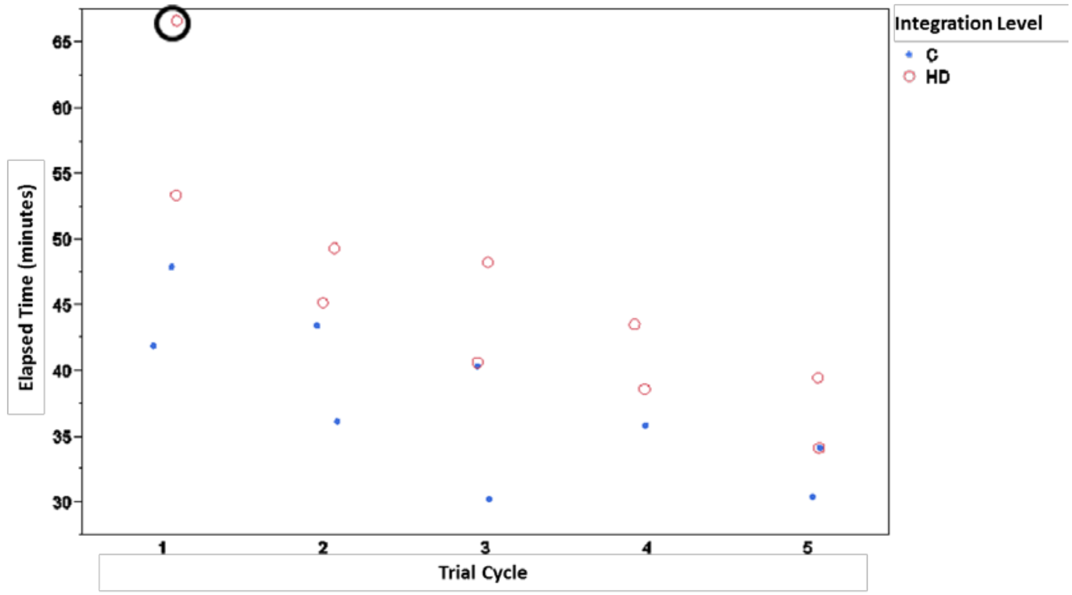
M.5.2.4 Cliff Ascent

M.5.2.4.1 Cliff Ascent Overview

This experimental task assessed a squad of 12 Marines climbing two 40-foot technical rock-climbing routes. A mountaineering expert (a red hat) was the only one authorized to verify knots and hook up each Marine to the belay line. Climbing the cliff was an individual task, but the recorded time was for the squad as a whole. After each Marine reached the top of the cliff, a fellow Marine on belay lowered the climber to the ground. The recorded time for this task started when the first Marine of the squad touched the rope and stopped when last member of the squad was completely disconnected from the rope.

Figure M-6 displays all Bridgeport open MOS cliff ascent data. There was one potential influential point—one of the HD points on trial cycle 1. Because the impact of these points is unknown, we perform all analysis with and without this point. All data on the scatter plot are valid for analysis with the potential influential point circled.

Figure M-6. Cliff Ascent with Potential Influential Point Circled



The inclusion of the potential influential points does not change the statistical significance between groups. It does, however, change the SD and percentage differences between the integration levels. Once we remove the potential influential points, the percentage differences between the C group and LD group, as well as the C group and HD group, decrease. The percentage difference between the LD and HD group increases. The SD for both the LD and HD groups decrease without the potential influential points. The following sections discuss results with and with the potential influential points.

M.5.2.4.1.1 Cliff Ascent Descriptive Statistics with Potential Influential Point

The C group had a mean time of 37.78 minutes. This time is statistically significantly faster than the HD mean time of 45.88 minutes. This difference results in a 21.45%, or 8.10-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 3.24-minute increase in SD (6.02 minutes for the C group and 9.26 minutes for the HD group). See Table M-2 and Table M-3 for detailed analytical results.

M.5.2.4.1.2 Cliff Ascent Descriptive Statistics without Potential Influential Point

The C group had a mean time of 37.78 minutes. This time is statistically significantly faster than the HD mean time of 43.57 minutes. This difference results in a 15.34%, or 5.79-minute, degradation in hike time between the groups. The HD group has greater variability, as shown by the 3.24-minute increase in SD (6.02 minutes for the C group and 9.26 minutes for the HD group). See Table M-2 and Table M-3 for detailed analytical results.

M.5.2.4.2 Cliff Ascent Contextual Comments

Sun Tzu stated, “Speed is the essence of war. Take advantage of the enemy’s unpreparedness; travel by unexpected routes and strike him where he has taken no precautions” (MCDP-1, p. 69). Gaining this advantage in a mountainous environment may involve climbing a cliff. Technical rock-climbing is a physical strain on the upper and lower body. While no purely objective standard can be set for climbing a cliff, any decrement in speed translates into an advantage to the enemy and greater risk for friendly casualties. On average, the HD group took 8.1 minutes longer than the C group. Furthermore, on any given day (under the same environmental conditions), the C group was always the fastest squad to finish, and, with the exception of one trial, the HD group was always the slowest squad to finish.

M.5.2.4.3 Cliff Ascent Additional Insights

A purely objective evaluation of 8.1 minutes is elusive but may possess some practical significance on the battlefield that would reduce the survivability of an integrated unit. Considering an enemy unit moving at 4.0 km/h on foot to defend a piece of key terrain, the enemy would be able to move an additional 540 meters as the HD group was still climbing the cliff. Far worse would be an enemy observer who could call for fire with mortars, artillery, or aviation assets against the integrated group given the time delay.

The steady “learning curve” that occurred can be explained in two ways: familiarity of the climbing route and the masking effect of belaying. First, there were two separate, but similar routes that Marines had to climb. Each subsequent opportunity to climb involved learned techniques/behaviors that resulting in a faster climb time. Second, a source of masking occurred during this task that was not anticipated during the design. The physically demanding aspect involved climbing the cliff; however, the belay/lowering time was a significant part of the overall time. The belay time also got drastically shorter as the assessment progressed because belay Marines became more comfortable with this skill. Strictly considering the performance of each Marine would have better informed the gender-difference involved in this task.

M.5.2.4.4 Cliff Ascent Subjective Comments

For subjective comments relating to this task, see the BP Open MOS Appendix.

Appendix to Annex M
Mountaineering Open MOSs Supplemental Information

This appendix provides supplemental information for the Mountaineering Open MOSs portion of the GCEITF experiment. It provides information regarding the GCEITF leadership subjective comments and additional descriptive and basic inferential statistics not described in Annex M.

Section 1. GCEITF Leadership Subjective Comments

The GCEITF leadership provided comments on their observations of the experiment throughout its execution. Table M A displays a summary of these comments broken down by task, integration level, gender, and type of comment.

Table M A – Summary of GCEITF Leadership Comments

Task and Metric Description	Gender	Falling behind/slowing movement			Requesting extra breaks			Requires extra assistance			Needs no assistance			Compensating for another Marine			Gear pass off			Other			No category			Total
		C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	C	HD	Total	
4.6-km Hike	M	7	0	7	4	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
	F	0	18	18	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5-km Return Hike	M	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	F	0	8	8	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	9
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gorge Crossing	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cliff Ascent	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	Unit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Section 2. Additional Task Basic and Inferential Statistics

Not all data collected during the experiment were used for MOS-specific task analysis and conclusions. The experiment included a number of tasks and associated metrics conducted to enhance operational realism and maintain consistent application of the experimental process. Where data were collected on tasks not used for analysis and conclusions, we simply report means, standard deviations, and percent differences. This section presents results for 10 additional Open MOS tasks. Annex M contains the descriptive statistics for the remainder of the Open MOS tasks. The words “metric” and “task” are used interchangeably throughout this Appendix; t. They both refer to the experimental task.

The table below displays the results for 10 additional Open MOS metrics. Table M B displays the metrics and integration levels with their respective sample sizes, means, standard deviations, and percent differences between integration levels.

Table M B – Bridgeport Open MOS Test Results (Descriptive Statistics)

Metric	Integration Level	Sample Size	Mean	SD	% Difference (HD-C)
4.6-km Hike; first km (minutes)	C	10	10.27	0.50	7.90%
	HD	10	11.08	0.37	
4.6-km Hike; second km (minutes)	C	10	11.09	0.53	17.37%
	HD	10	13.02	0.63	
4.6-km Hike; third km (minutes)	C	10	13.91	2.65	7.19%
	HD	10	14.91	1.54	
4.6-km Hike; fourth km (minutes)	C	10	18.06	3.77	12.22%
	HD	10	20.45	2.60	
4.6-km Hike; fifth km (minutes)	C	10	10.05	0.97	13.14%
	HD	10	11.37	0.73	
Return Hike; first km (minutes)	C	10	12.80	0.76	18.79%
	HD	10	15.20	1.00	
Return Hike; second km (minutes)	C	10	13.83	2.82	6.70%
	HD	10	14.76	1.09	
Return Hike; third km (minutes)	C	10	11.42	0.73	4.32%
	HD	10	11.91	1.31	
Return Hike; fourth km (minutes)	C	10	10.87	1.37	8.72%
	HD	10	11.81	1.26	
Return Hike; fourth km **[excluding potential influential point]** (minutes)	C	9	10.51	0.84	12.38%
	HD	10	11.81	1.26	
Return Hike; fifth km (minutes)	C	10	14.58	1.29	5.18%
	HD	10	15.34	1.68	

Additional Task Results:

4.6-km Hike by km. The general trend for the 4.6-km hike was that the difference between the HD and C groups increased over the course of the hike. These differences are largest during the points of the largest elevation change.

Return Hike by km. The general trend for the return hike was that the difference between the HD and C groups increased over the course of the hike. These differences are largest during the points of the largest elevation change.

Annex N.
Fatigue, Workload, and Cohesion Surveys

This annex details the results of the three self-report surveys (Fatigue, Workload, and Unit Cohesion) administered to all volunteers—in keeping with their respective MOS' Schemes of Maneuver—during the Ground Combat Element Integrated Task Force (GCEITF) experiment executed from 2 March – 18 May 2015. In the sections below, we outline Survey Design and Collection, Limitations, Data Set Description, and Detailed Survey Results.

N.1 Survey Design and Collection

N.1.1 Survey Design Overview

Three different surveys were administered—sometimes multiple times per trial—to GCEITF volunteers during the course of each run-cycle. All had been originally designed, or modified, to permit "in-stream" administration (to minimize interruption to each MOS' Scheme of Maneuver).

Two surveys—the individual Fatigue and Workload surveys—had been derived from the Crew Status Survey instrument, in order for Marines to self-report their perceived levels of fatigue (i.e., weariness or exhaustion from labor) and relative workload following some or all tasks within a trial.

For the Fatigue survey, volunteers were asked to pick the statement on a seven-point (ordinal) scale that best described how they felt at that moment, ranging from "1 – Fully Alert, Wide Awake; Extremely Peppy" to "7 – Completely Exhausted; Unable to Function Effectively; Ready to Drop".

For the workload survey—also on a seven-point (ordinal) scale—volunteers were asked to pick both the statement that best described the maximum workload they had experienced during the preceding work period, as well as their average workload during that period; responses could range from "1 – Nothing to do; No Task Demands" to "7 – Unmanageable; Potentially Dangerous; Unacceptable".

Unlike the original Crew Status Survey form—which can be found in the Experimental Assessment Plan (EAP), Annex D—we allowed the Fatigue and Workload surveys to be administered independently, and did not allow free-form comments.

The third survey—the Unit Cohesion survey—had been taken from the Group Environmental Questionnaire (GEQ), specifically its Group Integration-Task (GI-T) sub-section (see EAP, Annex M). Our survey had five statements (Q1 – Q5) designed to query an individual's beliefs about their team's relatively closeness, similarity, and bonding around a group's task. Responses fell along a nine-point Likert scale, with volunteers reporting what number best indicated their level of agreement with each statement, where "1" indicated strongly disagree and "9" indicated strongly agree.

N.1.2 Survey Administration and Collection

Our primary method for administering the surveys was on the same portable Toughbooks used by data collectors to collect observational data. At the appropriate break in each scenario (either before, during, or after), the data collection program would display the appropriate survey to be taken. The order each survey was to be taken was also fixed, based on billet. In the event that a volunteer was not available (either temporarily or because they'd been dropped from the trial or the study), the data collector could defer a particular survey or mark it not taken/invalid.

Marines responded to the survey by touching the appropriate multiple choice box—using either the Toughbook stylus or their finger tip. Volunteers would then hit "enter" to save their response—ensuring the anonymity of their responses. In the event of Toughbook failure, all data collectors carried paper backup copies and pens. Volunteers would fill out their response and return the paper to the data collector, who would add their Experimental ID code (EID), unit run order, and MOS in order to properly tie the paper backups to the correct trial.

Surveys were not administered to non-trial (i.e., loading units).

N.1.2.1 Pre-Trial Surveys

All volunteers participating in a trial run were given a pre-trial (i.e., baseline) Fatigue survey prior to the start of their trial. This was intended as a way to recognize when task or trial-specific fatigue results might have been influenced by initial fatigue levels.

N.1.2.2 Post-Trial Surveys

All volunteers participating in a trial were also given post-trial Fatigue and Workload surveys after completing their final task. Efforts were made to administer the surveys as close to the end of the trial as possible, although on occasion, range safety requirements necessitated that a survey be administered once the volunteers had come off the range and/or the range had gone cold.

In addition to the post-trial surveys, volunteers were given the Unit Cohesion survey following the last trial in their respective run-cycles.

N.1.2.3 Mid-Trial Surveys

In addition to the pre- and post-trial surveys mentioned above—which were given to all volunteers, regardless of MOS—some MOS also incorporated mid-trial surveys into their data collection plan. Where mid-trial Workload surveys were administered, data collectors instructed the volunteers to respond with their maximum and average workload levels over a work period defined as the period since their last survey. If no mid-trial surveys were administered, the work period was defined as the entire trial period. We discuss each mid-trial survey, when they occur, within each MOS section below.

N.2 Limitations

N.2.1 Limitations Overview

Unlike much of the quantitative data collected, analyzed, and presented elsewhere in this study, there are some additional limitations that should be taken into consideration with regards to survey results.

N.2.1.1 Respondent Biases

All surveys results represent volunteers' self-reports about their internal perceptions. As such, their responses are subject to potential (intentional and unintentional) biases. Some common types of bias include:

- Acquiescence – respondents answer based upon what they think the questioner wants to hear (or contrary to what they think they want to hear)
- Response fatigue – respondents answer in a uniform or inaccurate manner due to excessive survey requests or duration
- End aversion – respondents avoid selecting values at either end; also known as central tendency
- Yea- and nay-saying – respondents have a more or less global tendency answer positively or negatively
- Social desirability – respondents avoid responding with, or under-report, answers that might be viewed as socially unacceptable to the group

One reason we chose our primary method for collecting survey data (i.e., volunteers privately entered their responses directly into the Toughbook) was to attempt to mitigate against acquiescence and social desirability biases. We also limited, where possible, the number of times each survey was administered during a trial and streamlined the process as much as possible in order to stave off response fatigue biases. However, there is no way to completely protect against response bias, so we have tried to be cautious in the conclusions we draw below.

N.2.1.2 Original Survey Design

Our surveys also have certain limitations based upon their initial design. For example, the Fatigue and Workload surveys were taken from the Crew Status Survey instrument—which was designed to compare the fatigue and workload perceptions of each air crew member against their own previous responses over the course of their shift. For this purpose, the Crew Status survey designers had been able to verify the repeatability of each response level.

However, the Crew Status survey was not specifically designed to compare results between individuals—there being no way to know that different people will use the same internal criteria for selecting one response over another, beyond the wording of each potential response. As such, we caution against drawing too much from comparisons of fatigue and workload results between groups.

In addition, all three surveys (Fatigue, Workload, and Unit Cohesion) have subjectively-defined (i.e., ordinal) intervals, and as such we are not able to use most parametric analysis methods that make use of mathematical manipulation of the data. Statistical significance, where found, do not indicate a statistical difference between means—only that the two (or more) populations under inspection are sufficiently dissimilar.

N.2.1.3 Consistency of Administration

As mentioned above in Survey Administration and Collection, we scripted when and where each survey was taken into each MOS' Scheme of Maneuver. Every effort was made to ensure consistency in administration. However, changes to run order, unplanned interruptions, and range safety demands—to name a few—did, on occasion, interfere with survey administration and so it was not always possible to ensure consistency in the conditions under which each survey was conducted.

N.2.1.4 Inter-MOS variability

Because each MOS Scheme of Maneuver was independently designed and executed, with different definitions for unit size, task load, and integration level, we cannot directly compare survey results between different MOS groups—not even if, on the face of it, they have very similar Schemes of Maneuvers (e.g., 0331s and Provisional Machine Gunners).

N.3 Data Set Description

N.3.1 Data Set Overview

For the purposes of this annex, our data set includes all Fatigue, Workload, and Unit Cohesion surveys completed by members of a valid, trial unit.

Baseline fatigue results were included, regardless of whether that individual actually completed a trial. Mid- and Post-Trial survey results were kept as long as the FTMs and Data Managers for each site determined that the tasks conducted during the course of a trial were sufficiently consistent with the Scheme of Maneuver to maintain proper loading and division of labor (i.e., did not suffer from gross deviations in the execution of the trial).

We did not remove any outliers. We only culled individual responses if the response was not that of the Marine identified on the roster for that unit and billet (e.g., another Marine accidentally responded out of order). In some instances, Marines simply failed to complete all or some elements of a survey. The actual number of responses used for any particular analysis is detailed within the appropriate MOS-specific section below.

N.4 Detailed Survey Results

N.4.1 Analytical Approach Overview

As outlined in the GCEITF Experimental Assessment Plan (EAP), we focused primarily on the differences between male and female responses (i.e., gender differences) within the same MOS, and differences (in males' responses[1]) based upon integration level. This does not constitute the totality of analysis that we believe could be done with this data set and additional analysis—at some future date—may be desirable to gain additional insight potential drivers for the results collected.

Also, as previously mentioned above, survey responses are subjective (and ordinal) in nature. Responses can impart relative order but this does not imply even distribution along the scale[2]. As such, care should be taken to not draw conclusions based upon assumptions of equal intervals.

N.4.2 Infantry Rifleman (0311)

N.4.2.1 Fatigue Results

For the 0311 volunteers at Twentynine Palms, we have baseline, mid-trial, and final fatigue self-reports from 34 males and nine females. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.2.1.1 Fatigue Results by Gender

We first looked at looking at how fatigue self-reports compared between males and females, including their baseline (pre-trial), their final (post-trial, for both days in the run-cycle), and post-7km Hike (mid-trial, Defense trials) fatigue levels. Results are summarized in the table below (Table 1).

[1] By definition, females were only present in integrated groups.

[2] For example, an ordinal response of "1" is less than "2" and "2" is less than "3", but the difference between "1" and "2" may not be equal to the difference between "2" and "3".

Table 1 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) - 0311

MOS	Fatigue	Location / Day	Gender	N	Median / Mode	χ2 Statistic	χ2 p-Value
0311	Baseline (Pre-Trial)	29Palms/1,2	M	903	2 / 2	56.6	<0.01*
			F	82	3 / 3		
	Post-7km (Mid-Trial)	29Palms/2 "Defense"	M	450	3 / 3	36.1	<0.01*
			F	39	5 / 5		
	Final (Post-Trial)	29Palms/1 "Offense"	M	439	3 / 3	27.8	<0.01*
			F	42	4 / 4		
		29Palms/2 "Defense"	M	451	4 / 3	23.9	<0.01*
			F	40	4 / 4		

N.4.2.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We found a significant difference between female and male 0311 volunteers' pre-trial fatigue scores. Males tended to report being less fatigued, with responses typically clustered around "1-Fully Alert" (25.6%), "2-Very Lively" (36.8%), and "3-Okay" (24.5%), whereas females most often reported being "3-Okay" (61.0%) or "4-A Little Tired" (22.0%) before the start of a trial (Table 2).

However, we found very few of the higher fatigue scores for either gender (i.e., "5-Moderately Tired", "6-Extremely Tired", or "7-Exhausted")—only 2.4 percent of male responses (driven largely by one Marine), and 3.7 percent of female responses (driven entirely by two Marines)—and no obvious pattern of increasing scores over time, suggesting volunteers largely had sufficient recovery time between trials.

N.4.2.1.1.2 Final Fatigue Scores (by Gender)

In looking at the volunteers' post-trial responses, we found significant differences in the distribution of scores—by gender—for both days of the 0311 trial-cycle. While the majority of male responses were clustered within the "2-Very Lively" to "3-Okay" range for both days (56.3% for Offense trials, 52.9% for Defense trials), females' post-trial responses in this same range were far fewer—i.e., 28.6% and 30.0%, respectively.

Females more often reported being either "4-A Little Tired" or "5-Moderately Tired" (59.5% for "Offense" scenario, 67.5% for the "Defense" scenario). Similar to pre-trial results for both genders, we saw few examples of fatigue reports (4 reports from males, 10 reports from females) in the highest levels ("6-Extremely Tired", "7-Exhausted"); in all cases, these were associated with Offense trials.

N.4.2.1.1.3 Post-7km Hike Fatigue Scores (by Gender)

For the Defense scenario, 0311 volunteers also completed a fatigue survey following the completion of their 7km Hike under load; we found a significant difference in the distributions of male and female fatigue responses for this survey as well.

We found the single most post-7km Hike scores for males at “3-Okay” (29.6%), but with a relatively flat span (13% to 19%) of responses on either side—from “1-Fully Alert” to “5-Moderately Tired”.

A similar pattern was found for female responses, but shifted to the right with their peak (35.9%) at “5-Moderately Tired”, with other responses spanning from “3-Okay” (23.1%) to “6-Extremely Tired” (15.4%). As the female responses for the post-7km hike extend more to the right (i.e., more fatigued) than their post-trial responses for the same day, some recovery may have been occurring.

Table 2 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) - 0311

Distribution of 0311 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0311	29Palms/1&2 Pre-Trial	M	903	25.6%	36.8%	24.5%	10.7%	2.3%	0.0%	0.1%
		F	82	3.7%	9.8%	61.0%	22.0%	3.7%	0.0%	0.0%
	29Palms/1 Post-7km Hike	M	450	19.1%	17.6%	29.6%	18.0%	13.1%	2.4%	0.2%
		F	39	0.0%	2.6%	23.1%	20.5%	35.9%	15.4%	2.6%
	29Palms/1 Post-"Offense"	M	439	14.1%	26.0%	30.3%	16.6%	10.5%	2.3%	0.2%
		F	42	2.4%	11.9%	16.7%	21.4%	38.1%	9.5%	0.0%
	29Palms/2 Post-"Defense"	M	450	17.1%	20.0%	32.9%	19.6%	10.4%	0.0%	0.0%
		F	40	2.5%	7.5%	22.5%	37.5%	30.0%	0.0%	0.0%

N.4.2.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 0311s, we compared males’ responses in Control (i.e., all-male) and High-Density (i.e., 2 females) Rifle Squads. We have mid-trial and final fatigue self-reports from 34 males (33 in the Control group and 34 in the High-Density group); statistical results are summarized in the table below (Table 3).

Table 3 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) - 0311

MOS	Fatigue	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value
0311	Post-7km (Mid-Trial)	29Palms/2 "Defense"	C	246	3 / 3	3.6	0.06*
			HD	204	3 / 3		
	Final (Post-Trial)	29Palms/1 "Offense"	C	235	3 / 3	0.34	0.56
			HD	204	3 / 3		
		29Palms/2 "Defense"	C	247	3 / 3	6.0	0.01*
			HD	204	3 / 3		

N.4.2.1.2.1 Post-7km Hike Fatigue Scores (by Integration Level)
Following the 7km Hike under load (Defense trials), we found the distribution of males' fatigue responses in High-Density Rifle Squads to be significantly different (in the direction of less fatigued) with more responses (40.7%) in the "1-Fully Alert" to "2-Very Lively" range, compared to Control groups (33.3%), despite peaks for both at "3-Okay".

One potential reason for this shift could be the slower average pace typically seen in High-Density groups for the 7km Hike (see Annex A). However, this does not appear to be universal for all males because we still see a (non-trivial) number of male responses for both groups in the "4-A Little Tired" and "5-Moderately Tired" (32.1% in Control groups versus 29.9% in High-Density groups).

Regardless, with a majority of female responses (53.8%, representing five of the nine females) in integrated groups falling in the higher levels post-hike ("5-Moderately Tired", "6-Extremely Tired", and "7-Exhausted"), it does appear that a substantial number of female 0311 volunteers were "pushing it" to maintain the pace set by their unit (see Table 2 above).

N.4.2.1.2.2 Final Fatigue Scores (by Integration Level)
Post-trial fatigue results by Integration Level proved to be mixed—males' responses were significantly different between the two groups (Control and High-Density) for the Defense, but not for the Offense, scenario.

As with the distributions seen following the 7km Hike, for post-Defense trials we see a slight increase in the number of male responses in the "1-Fully Alert" to "2-Very Lively" range for High-Density groups (42.9%) compared to Control groups (32.4%), along with a corresponding (and slight) decrease in their responses at "3-Okay" and above.

However, if a slower pace could be the mechanism driving some males towards less fatigue post-7km hike, and post-trial on Defense trials, why is this not seen for Offense trials, which also found slower speeds (on average) for integrated groups on a number of tasks (see Annex A)?

It may be that it's not simply pace, but the period over which it's maintained (i.e., a cumulative effect)—possibly in combination with the higher pack-weight found in Defense trials—that drives a shift in fatigue for some males. Should relative pace and duration be the factor driving this difference in fatigue reports, we might expect a similar pattern to present—or strengthen—for tasks of greater duration and/or which drive a greater difference in pace (e.g., longer hikes). If pack-weight is a contributing factor, lighter pack-weights (or those scaled to a Marine's relative body-mass) over a similar 7km hike distance, might reduce the number of males in Control groups reporting "6-Extremely Tired" or "7-Exhausted", while maintaining the same pace.

Table 4 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) - 0311

Distribution of 0311 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0311	29Palms/1 Post-7km Hike	C	246	17.1%	16.3%	30.5%	17.5%	14.6%	3.7%	0.4%
		HD	204	21.6%	19.1%	28.4%	18.6%	11.3%	1.0%	0.0%
	29Palms/1 Post-"Offense"	C	235	11.9%	28.1%	29.8%	16.6%	10.6%	2.6%	0.4%
		HD	204	16.7%	23.5%	30.9%	16.7%	10.3%	2.0%	0.0%
	29Palms/2 Post-"Defense"	C	247	14.2%	18.2%	34.4%	20.6%	12.6%	0.0%	0.0%
		HD	203	20.7%	22.2%	31.0%	18.2%	7.9%	0.0%	0.0%

N.4.2.2 Workload Results

For the 0311 volunteers at Twentynine Palms, we have maximum workload self-reports from 34 males and nine females[3]. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.2.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females, for both days in their run-cycle. Results are summarized in the table below (Table 5).

[3] One male and two females only participated in two trials (each).

Table 5 - Summary of Maximum Workload Scores (by Gender) – 0311

MOS	Workload	Location / Day	Gender	N	Median / Mode	χ2 Statistic	χ2 p-Value
0311	Final (Max Workload for Trial)	29Palms/1 "Offense"	M	448	4 / 4	4.1	0.04*
			F	41	4 / 4		
		29Palms/2 "Defense"	M	452	4 / 4	4.5	0.03*
			F	40	4 / 4		

We found a significant difference between female and male 0311 volunteers' maximum workload scores for both days (Offense and Defense scenarios). In both cases, we found males slightly more likely (44.4%) to report lower maximum workload levels ("3-Active but Easy" and lower) than females (26.8%), although both groups had their peak at "4-Challenging" (Table 6).

We found few of the highest max workload scores for either group (i.e., "6-Overloaded" or "7-Unmanageable")—only 2.0 percent of male responses[4] and 2.4 percent of female responses (a single response). This suggests most Marines (regardless of gender) felt capable of accomplishing, if not always easily, the workload assigned them.

The difference seen between the two groups may be due to females having less time in the MOS; if so, we would expect the difference between the two groups to gradually disappear over time as they became more skilled—and hence more comfortable—in their responsibilities. Additionally, females' higher representation at "4-Challenging" and "5-Extremely Busy" may have been influenced by their slightly higher levels of fatigue, since even relatively low levels of fatigue can degrade concentration and performance of non-rote skills.

Table 6 – Maximum Workload Score Distributions (by Gender) - 0311

Distribution of 0311 Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0311	29Palms/1 "Offense"	M	448	2.7%	7.6%	34.2%	42.2%	11.4%	1.8%	0.2%
		F	41	0.0%	4.9%	22.0%	56.1%	14.6%	2.4%	0.0%
	29Palms/2 "Defense"	M	452	1.1%	7.5%	30.8%	44.0%	13.9%	2.4%	0.2%
		F	40	0.0%	2.5%	15.0%	65.0%	17.5%	0.0%	0.0%

[4] Eight 0311 males reported a max workload of "6-Overloaded" or "7-Unmanageable", though most did so only once or twice. One Marine reported a max workload in the range ten times—split between both scenarios.

N.4.2.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males' fatigue responses remained consistent at different integration levels. For 0311s, we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 2 females) Rifle Squads. We have maximum workload self-reports from 34 males (33 in the Control group and 34 in the High-Density group); statistical results are summarized in the table below (Table 7).

Table 7 - Summary of Males' Maximum Workload Scores (by Integration Level) – 0311

MOS	Workload	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value
0311	Final (Max Workload for Trial)	29Palms/1 "Offense"	C	239	4 / 4	3.3	0.07*
			HD	209	4 / 4		
		29Palms/2 "Defense"	C	248	4 / 4	8.5	<0.01*
			HD	204	4 / 4		

We saw a significant difference between males' responses in Control and High-Density groups. For both days, we saw a slight depression in their reported maximum workload levels (towards more manageable) in the High-Density group—with more reports at "3-Active but Easy" (Table 8)—than the number seen in the Control group.

It may be that the slightly slower times associated—on average—with High-Density squads (see Annex A) allowed some males to have an easier time addressing the workload demands of the scenario. However, without greater fidelity as to what tasks elicited their max scores, we can't say whether this is a viable theory.

Alternatively, integrated groups may have showed greater collaboration between members (i.e., spreading the "load") than Control groups. However, as most tasks were collected at the unit—and not at the individual—level, this would need to be explored through follow-on work.

Table 8 - Males' Maximum Workload Score Distributions (by Integration Level) - 0311

Distribution of 0311 Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0311	29Palms/1 "Offense"	C	239	2.5%	8.4%	29.3%	43.9%	13.8%	1.7%	0.4%
		HD	209	2.9%	6.7%	39.7%	40.2%	8.6%	1.9%	0.0%
	29Palms/2 "Defense"	C	248	0.8%	6.5%	26.2%	47.2%	15.7%	3.2%	0.4%
		HD	204	1.5%	8.8%	36.3%	40.2%	11.8%	1.5%	0.0%

N.4.2.3 Cohesion Results

For the 0311 volunteers at Twentynine Palms, we have Unit Cohesion survey results (with five statements per survey) from 34 males and nine females[5]. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.2.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 9).

Table 9 - Summary of Unit Cohesion Scores (by Gender) – 0311

MOS	Cohesion	Gender	N	Mean	SD	χ2 Statistic	χ2 p-Value
0311	Q1: Group was united in trying to reach goals	M	453	6.18	1.71	0.64	0.43
		F	39	6.06	1.96		
	Q2: We all take responsibility for task success of the group	M	453	6.15	1.82	1.06	0.30
		F	39	5.92	1.99		
	Q3: Group members have similar aspirations for success	M	453	5.92	1.90	0.12	0.72
		F	39	5.92	1.91		
	Q4: If members had problems with a task, all wanted to help	M	453	6.02	1.84	4.53	0.03*
		F	39	5.56	1.79		
	Q5: Members communicated freely about responsibilities	M	453	6.34	1.82	4.18	0.04*
		F	39	5.85	1.86		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	M	453	30.59	8.16	1.71	0.19
		F	39	29.31	9.04		

We found mixed results; males and females had significantly different responses on only two statements, "Q4-Help" (i.e., the group's willingness to help a struggling member) and "Q5-Comm" (i.e., the group communicating freely about responsibilities), with females reporting more pessimistic responses than males for both.

When we looked more closely at the response distributions of both statements, we found females responses for both tended to cluster more strongly between "4" and "6"

[5] Two females and one male only participated in two units (each).

(71.8% for “Q4-Help” and 61.5% for “Q5-Comm”)—straddling both sides of neutral. On the other hand, males’ responses to both statements tended to cluster between “5” to “8” (78.6% for “Q4-Help” and 80.6% for “Q5-Comm”)—spanning neutral to strong agreement.

One potential reason for the difference in perceptions between males and females for just these two components could have been differences in relative experience within the MOS. Females could have been seeking a greater degree of cooperation and communication to offset their inexperience than the more experienced males might have felt necessary.

We did not, however, find a significant difference in perceptions for the other three statements (“Q1-Unity”, “Q2-Responsibility”, and “Q3-Aspirations”), nor between their overall composite cohesion scores, suggesting a general agreement in the range of their groups’ respective cohesiveness in terms of goals and motivation.

We also graphed the means for each statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement (Figure 1). For the 0311 volunteers, we can see how both genders—while more positive than negative—appeared to be somewhat neutral in their perceptions of unit cohesion (with sharper divergence for “Q4-Help” and “Q5-Comm”).

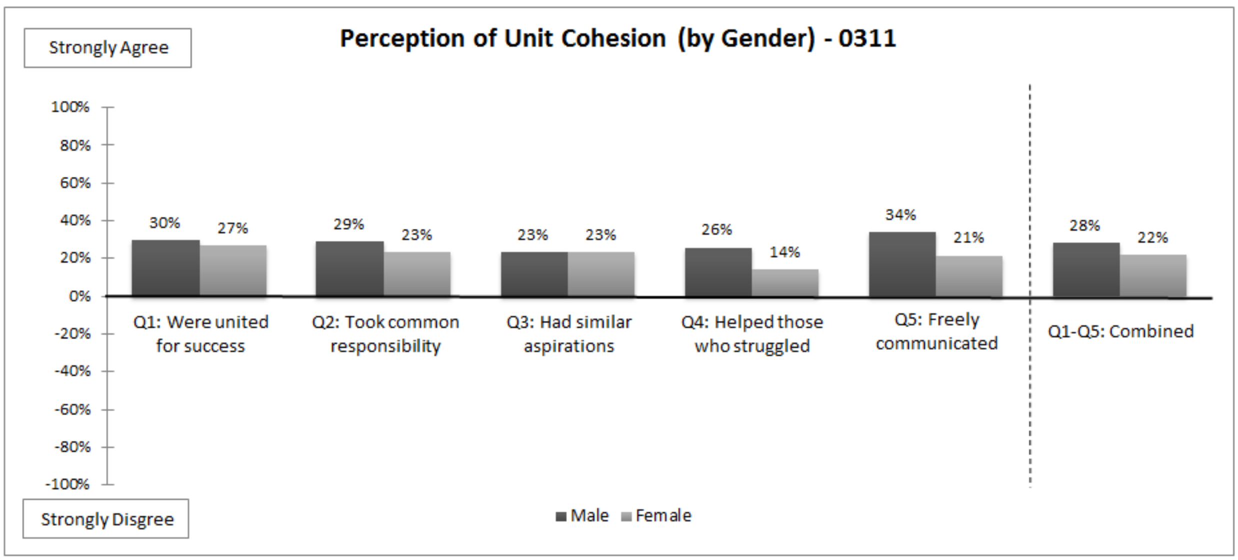


Figure 1 - Perception of Unit Cohesion (by Gender) – 0311

N.4.2.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males’ cohesion responses remained consistent at different integration levels. For 0311s, we compared males’ responses in Control (i.e., all-male) and High-Density (i.e., 2 females) Rifle Squads. We have Unit Cohesion survey results (five statements per survey) from 34 males (33 in the Control group and 34 in the High-Density group); statistical results are summarized in the table below (Table 10).

Table 10 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 0311

MOS	Cohesion	IL	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
0311	Q1: Group was united in trying to reach goals	C	249	6.31	1.69	4.86	0.03*
		HD	204	6.02	1.72		
	Q2: We all take responsibility for task success of the group	C	249	6.28	1.78	3.70	0.05*
		HD	204	5.98	1.86		
	Q3: Group members have similar aspirations for success	C	249	6.04	1.89	3.04	0.08*
		HD	204	5.77	1.92		
	Q4: If members had problems with a task, all wanted to help	C	249	6.15	1.85	4.60	0.03*
		HD	204	5.87	1.81		
	Q5: Members communicated freely about responsibilities	C	249	6.47	1.81	4.22	0.04*
		HD	204	6.18	1.82		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	249	31.24	8.15	4.82	0.03*
		HD	204	29.78	8.12		

We found a statistical difference—by integration level—between all five cohesion statements, as well as for the overall composite cohesion score. When we looked at the distributions for each statement, we saw a slight shift towards more agreement (responses ranging between “5” and “8”) in the Control group (79.5% to 87.2%) over those found in the High-Density group (74.4% to 78.9%) along with a corresponding dip in the Control group’s representation in the lower levels (below “5”).

We also graphed the means for each statement as a percentage of how strongly males agreed (positive) or disagreed (negative) with each question’s statement while in each group (Figure 2). While the figure is consistent with our statistical results—i.e., males in Control groups in slightly stronger agreement about unit cohesion than males in High-Density groups—we can see that, regardless of integration level, males were fairly neutral overall.

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group. Differences in relative experience may have been a driver—creating a perception of unequal contribution to the group effort. Incidentally, the average values for the integrated groups appear to align well with how females rated their perceptions of unit cohesiveness.

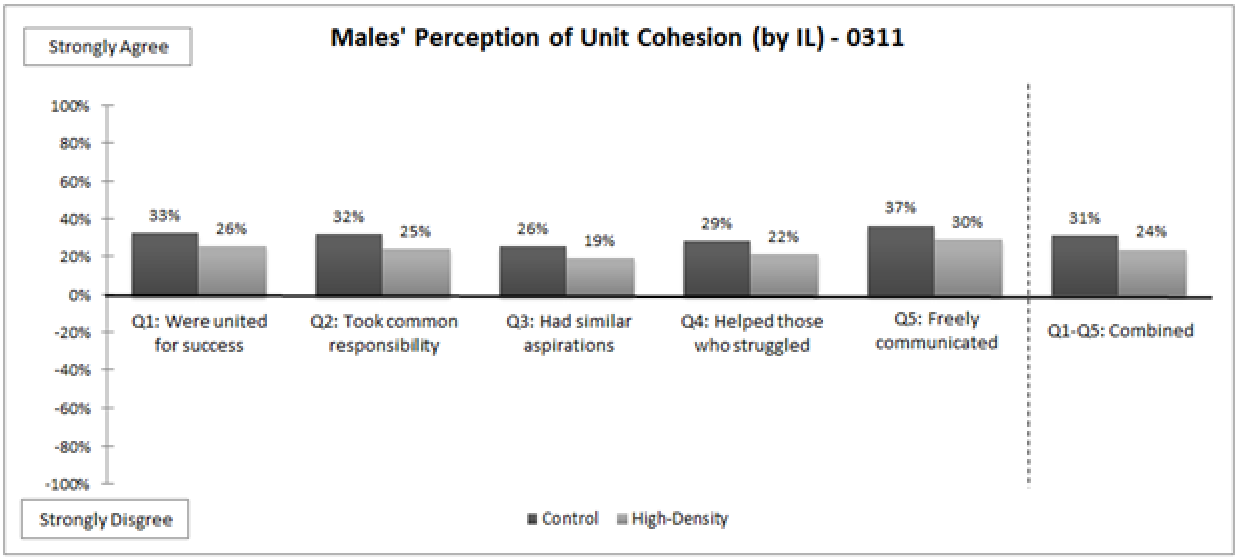


Figure 2 - Perception of Unit Cohesion (by Integration Level) – 0311

N.4.3 Infantry Machine Gunner (MOS 0331)

N.4.3.1 Fatigue Results

For the 0331 volunteers at Twentynine Palms, we have baseline, mid-trial, and final fatigue self-reports from six males and six females. Because Fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.3.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial), final (post-trial, for both days in the run-cycle), and post-7km Hike (mid-trial, Defense trials) fatigue levels. Results are summarized in the table below (Table 11).

Table 11 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) - 0331

MOS	Fatigue	Location / Day	Gender	N	Median / Mode	χ2 Statistic	χ2 p-Value
0331	Baseline (Pre-Trial)	29Palms/1,2	M	236	1 / 1	60.2	<0.01*
			F	118	2 / 2		
	Post-7km (Mid-Trial)	29Palms/2 “Defense”	M	117	3 / 3	1.4	0.24
			F	59	3 / 3		
	Final (Post-Trial)	29Palms/1 “Offense”	M	115	3 / 3	3.8	0.05*
			F	58	3.5 / 3		
		29Palms/2 “Defense”	M	118	3 / 3	2.5	0.11
			F	59	3 / 3		

N.4.3.1.1.1 Pre-Trial Fatigue Scores

We found a significant difference between female and male 0331 volunteers’ pre-trial fatigue scores. Males’ responses were heavily clustered around “1-Fully Alert” (53.8%) and “2-Very Lively” (41.5%), whereas females’ scores showed a more shallow distribution spanning “1-Fully Alert” (22.0%), “2-Very Lively” (38.1%), and “3-Okay” (25.4%) before the start of a trial (Table 12).

In addition, we found very few of the higher fatigue scores for either gender (i.e., “5-Moderately Tired”, “6-Extremely Tired”, or “7-Exhausted”)—only a single male response (0.4%), and six female responses (5.1%) of which all but one were driven by a single Marine. There was no obvious pattern of increasing scores over time, suggesting volunteers had sufficient recovery time between trials.

N.4.3.1.1.2 Final Fatigue Scores

In looking at the volunteers’ post-trial responses, we found a significant differences in the distribution of male and female scores for only one scenario—Offense trials. For these trials, both genders showed a relatively shallow distribution of scores in the low to medium range (Table 12). Males clustered between “2-Very Lively” and “4-A Little Tired” (peaking at “3-Okay” with 31.3% of responses), while and females clustered across a slightly broader range between “2-Very Lively” and “5-Moderately Tired” (but also peaking at “3-Okay” with 24.1% of their responses). The results for Defense trials were not significant.

For both days and genders, we did see some final fatigue scores in the higher “5-Moderately Tired” to “6-Extremely Tired” range. Unlike pre-trial results—where the handful of higher-than-average scores were driven primarily by two Marines (one male and one female)—these higher post-trial fatigue scores were distributed across all 12 volunteers over the course of the experiment, suggesting all Marines had been highly fatigued, on occasions. But, this appears to have been relatively rare, since we also see greater representation by both genders, and among most of the volunteers (six males and five females), in the lower fatigue levels (i.e., “3-Okay” and lower).

N.4.3.1.1.3 Post-7km Hike Fatigue Scores

For the Defense scenario, 0331 volunteers also completed a fatigue survey following the completion of their 7km Hike under load; we did not find a significant difference in male and female responses; both groups had peak responses at “3-Okay”.

Table 12 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) - 0331

Distribution of 0331 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	**Location/Day**	**IL**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0331	29Palms/1&2 Pre-Trial	M	236	53.8%	41.5%	3.4%	0.8%	0.4%	0.0%	0.0%
		F	118	22.0%	38.1%	25.4%	9.3%	3.4%	1.7%	0.0%
	29Palms/2 Post-7km Hike	M	117	2.6%	36.8%	37.6%	16.2%	5.1%	1.7%	0.0%
		F	59	13.6%	16.9%	35.6%	16.9%	13.6%	3.4%	0.0%
	29Palms/1 Post-"Offense"	M	115	5.2%	27.0%	31.3%	25.2%	9.6%	1.7%	0.0%
		F	58	6.9%	19.0%	24.1%	20.7%	22.4%	6.9%	0.0%
	29Palms/2 Post-"Defense"	M	118	8.5%	35.6%	39.0%	15.3%	1.7%	0.0%	0.0%
		F	59	8.5%	28.8%	30.5%	27.1%	3.4%	1.7%	0.0%

N.4.3.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 0331s, we compared males’ responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) Machine Gun Teams. We have mid-trial and final fatigue self-reports from six males (six in the Control and Low-Density groups, and five males for High-Density groups); statistical results are summarized in the table below (Table 13).

Table 13 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) - 0331

MOS	Fatigue	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0331	Post-7km (Mid-Trial)	29Palms/2 "Defense"	C	61	3 / 3	25.3	<0.01*	(LD-C) -3.6	<0.01*
			LD	35	2 / 2			(HD-C) -4.3	<0.01*
			HD	21	2 / 2			(HD-LD) -1.8	0.07
	Final (Post-Trial)	29Palms/1 "Offense"	C	59	3 / 3	12.3	<0.01*	(LD-C) -0.87	0.39
			LD	36	3 / 3			(HD-C) -3.6	<0.01*
			HD	20	2 / 2			(HD-LD) -2.4	0.02*
		29Palms/2 "Defense"	C	62	3 / 3	8.7	0.01*	(LD-C) -1.2	0.23
			LD	35	3 / 2,3			(HD-C) -2.8	<0.01*
			HD	21	2 / 2			(HD-LD) -2.2	0.03*

N.4.3.1.2.1 Post-7km Hike Fatigue Scores (by Integration Level)

Following the 7km Hike, we found the distribution of males' fatigue responses in Control teams to be significantly different from both Low-Density and High-Density MG Teams.

For both integrated groups, males' fatigue scores were shifted lower (towards less fatigue), with a much higher number of responses at "2-Very Lively" fatigue level (54.3% and 66.7%, for Low-Density and High-Density groups, respectively) compared to the Control group (16.4%). At the same time, we also saw a corresponding drop in their responses in integrated groups at the medium to high range levels (between "4-A Little Tired" and "6-Extremely Tired"), which suggests males were less fatigued when participating in an integrated group (Table 14).

One potential reason for the decrease in males' post-hike fatigue scores for integrated groups is their slower (on average) pace on the hike compared to Control groups (see Annex C). However, it is possible motivation or willingness to "push it" also differed between integrated and non-integrated groups.

As we saw in the fatigue results by gender, the majority of female responses following the 7km hike were "4-A Little Tired" or below (83.1%)—suggesting many female respondents might have been able to handle a more aggressive pace and/or were deliberately choosing a level of exertion that would allow for a greater energy reserve for follow-on tasks.

Conversely, all-male groups appear more willing to push a pace that results—on occasion—in some males (11.5%) reporting themselves to be “5-Moderately Tired” or “Extremely Tired” following the conclusion of the hike, compared to integrated groups (only 2.9% of responses in Low-Density groups; zero responses in High-Density groups).

N.4.3.1.2.2 Final Fatigue Scores (by Integration Level)

Post-trial fatigue results by Integration Level for 0331s proved to be mixed—males’ responses were significantly different between Control groups and High-Density groups, and between High-Density and Low-Density groups, regardless of day in the trial-cycle. However, we found no significant difference in males’ fatigue responses between Control groups and Low-Density groups for either day in the run-cycle.

For High-Density groups—similar to the post-7km hike results—we saw the distribution of males’ scores shifting left (i.e., towards less fatigued), with a sharp rise (65.0%) in the number of scores in the lowest levels (“1-Fully Alert”, “2-Very Lively”), compared to Control (22.0%) and Low-Density groups (30.6%).

As was hypothesized for the 7km hike results, it may be that this decrease in males’ fatigue responses in integrated groups is tied to the slower tempo typically seen for integrated units performing movement-related tasks (see Annex C).

However, as we also saw for the post-7km Hike results, female final fatigue scores in integrated units were also still largely distributed in the lower levels (“4-A Little Tired” and below)—for both Offense (70.7%) and Defense (94.9%)—with representation from all six female 0331 volunteers. This raises the potential that integrated units were setting tempos that neither gender found particularly taxing (i.e., not “pushing it”)—perhaps because they lacked sufficient motivation or were attempting to reserve energy for follow-on tasks.

Alternatively, it may be that all or some volunteers in integrated units were under-representing their physical fatigue levels on the surveys, and the trial tempos achieved properly reflect physical ability. Given the relatively small size of the 0331 volunteer pool (six males and six females), we should be cautious in extrapolating their results to the larger Marine Corps 0331 community.

Table 14 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) - 0311

Distribution of 0313 Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0331	29Palms/2 Post-7km Hike	C	61	1.6%	16.4%	47.5%	23.0%	8.2%	3.3%	0.0%
		LD	35	0.0%	54.3%	34.3%	8.6%	2.9%	0.0%	0.0%
		HD	21	9.5%	66.7%	14.3%	9.5%	0.0%	0.0%	0.0%
	29Palms/1 Post-"Offense"	C	59	0.0%	22.0%	33.9%	32.2%	8.5%	3.4%	0.0%
		LD	36	8.3%	22.2%	33.3%	19.4%	16.7%	0.0%	0.0%
		HD	20	15.0%	50.0%	20.0%	15.0%	0.0%	0.0%	0.0%
	29Palms/2 Post-"Defense"	C	62	11.3%	19.4%	48.4%	19.4%	1.6%	0.0%	0.0%
		LD	35	2.9%	42.9%	42.9%	8.6%	2.9%	0.0%	0.0%
		HD	21	9.5%	71.4%	4.8%	14.3%	0.0%	0.0%	0.0%

N.4.3.2 Workload Results

For the 0331 volunteers at Twentynine Palms, we have maximum workload self-reports from six males and six females. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.3.2.1 Workload Results by Gender

We first looked at looking at how maximum workload self-reports compared between males and females, for both days in their run-cycle. Statistical results are summarized in the table below (Table 15)

Table 15 - Summary of Maximum Workload Scores (by Gender) – 0331

MOS	Workload	Location / Day	Gender	N	Median / Mode	χ2 Statistic	χ2 p-Value
0331	Max Workload for Trial	29Palms/1 "Offense"	M	116	4 / 4	28.8	<0.01*
			F	58	4 / 4		
		29Palms/2 "Defense"	M	120	4 / 4	15.8	<0.01*
			F	60	4 / 5		

We found a significant difference between female and male 0331 volunteers' maximum workload scores for both days (Offense and Defense scenarios). In both cases, we found males more likely to report maximum scores between “3-Active but Easy” and “4-Challenging” (79.3% and 79.2%, for Offense and Defense, respectively) than females, who favored scores between “4-Challenging” and “5-Extremely Busy” for both days (84.5% and 78.3%, respectively) (Table 16).

For both genders, we also saw a very small number of reports of “6-Overloaded”—three from each group, though only one female reported in this range more than once (and only for a fraction of her total trials). This suggests most 0331 Marines—regardless of gender—felt capable of accomplishing (if not always easily) their assigned workload.

Table 16 - Maximum Workload Score Distributions (by Gender) - 0331

Distribution of 0331 Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0331	29Palms/1 "Offense"	M	116	0.0%	6.0%	37.9%	41.4%	12.9%	1.7%	0.0%
		F	58	0.0%	0.0%	10.3%	43.1%	41.4%	5.2%	0.0%
	29Palms/2 "Defense"	M	120	0.0%	0.8%	35.0%	44.2%	19.2%	0.8%	0.0%
		F	60	0.0%	0.0%	16.7%	35.0%	43.3%	5.0%	0.0%

The difference between the two groups may be due to females having less time in the MOS; if so, we would expect the difference between males and females to gradually disappear over time. It may have also been influenced by the slightly higher levels of fatigue reported by female 0331s, since even relatively low levels of fatigue can degrade concentration and performance on non-rote skills.

N.4.3.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males workload responses remained consistent at different integration levels. For 0331s, we compared males’ responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) Machine Gun Teams. We have maximum workload self-reports from six males (six in the Control and Low-Density groups, and five in the High-Density group); statistical results are summarized in the table below (Table 17).

Table 17 - Summary of Males' Maximum Workload Scores (by Integration Level) - 0331

MOS	Workload	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0331	Max Workload for Trial	29Palms/ 1 "Offense"	C	60	4 / 4	15.6	<0.01*	(LD-C) -2.2	0.03*
			LD	36	3.5 / 3			(HD-C) -3.8	<0.01*
			HD	20	3 / 3			(HD-LD) -1.9	0.06
		29Palms/ 2 "Defense"	C	63	4 / 4	28.3	<0.01*	(LD-C) -4.8	<0.01*
			LD	36	3 / 3			(HD-C) -3.7	<0.01*
			HD	21	3 / 3			(HD-LD) -0.47	0.64

We saw a significant difference in males' responses between Control and both Low- and High-Density groups (but not between integrated groups). In the integrated groups, we see a shift in males' responses toward lower maximum workload scores (i.e., towards more manageable), centered more strongly at "3-Active but Easy"—instead of "4-Challenging"—for both days (Table 18).

As we saw in the analysis by gender above, we found very few of the highest maximum workload scores at any Integration Level—only three male responses in total. But, among the next lower level (i.e., "5-Extremely Busy"), which is on the border for managing the workload, we see both the number of contributors—and the number of responses per contributor—decrease in the integrated groups (i.e., more males persistently found their workload to be more manageable in an integrated group).

As far as possible mechanisms, it may be that Control groups, with their slightly faster times (on average) were more likely to set tempos that every volunteer struggled with on occasion (see Annex C). However, perhaps the slower tempos of integrated MG Teams allowed more males to have an easier time addressing the workload demands of the scenario. However, without greater fidelity as to what tasks tended to elicit the Marines' maximum workload reports, it's difficult to say whether this is a viable theory.

Alternatively, it may be that integrated groups showed greater collaboration (i.e., spreading the "load") than Control groups. However, as most tasks were collected at the unit—and not at the individual—level, this would need to be explored through follow-on work, with a larger 0331 volunteer pool than what was used in this study.

Table 18 - Males' Maximum Workload Score Distributions (by Integration Level) - 0331

Distribution of 0331 Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0331	29Palms/1 "Offense"	C	60	0.0%	0.0%	30.0%	50.0%	16.7%	3.3%	0.0%
		LD	36	0.0%	8.3%	41.7%	38.9%	11.1%	0.0%	0.0%
		HD	20	0.0%	20.0%	55.0%	20.0%	5.0%	0.0%	0.0%
	29Palms/2 "Defense"	C	63	0.0%	0.0%	12.7%	57.1%	30.2%	0.0%	0.0%
		LD	36	0.0%	0.0%	58.3%	36.1%	5.6%	0.0%	0.0%
		HD	21	0.0%	4.8%	61.9%	19.0%	9.5%	4.8%	0.0%

N.4.3.3 Cohesion Results

For the 0331 volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) from six males and six females. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating “Strongly Disagree” and nine indicating “Strongly Agree”).

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.3.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 19).

Table 19 - Summary of Unit Cohesion Scores (by Gender) - 0331

MOS	Cohesion	Gender	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
0331	Q1: Group was united in trying to reach goals	M	119	8.08	1.73	8.08	<0.01*
		F	60	8.45	1.69		
	Q2: We all take responsibility for task success of the group	M	119	8.38	1.25	4.38	0.04*
		F	60	8.53	1.50		
	Q3: Group members have similar aspirations for success	M	119	7.92	1.80	7.01	0.01*
		F	60	8.47	1.56		
	Q4: If members had problems with a task, all wanted to help	M	119	8.39	1.37	0.02	0.88
		F	60	8.40	1.36		
	Q5: Members communicated freely about responsibilities	M	119	8.34	1.39	6.28	0.01*
		F	60	8.65	1.38		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	M	119	41.12	6.77	6.64	0.01*
		F	60	42.50	6.80		

We found significant differences between genders for four of the five cohesion statements (“Q1-Unity”, “Q2-Responsibility”, “Q3-Aspirations”, and “Q5-Comm”), as well as for the overall composite cohesion score. When we look more closely at the distribution of 0331 responses for each statements, we see females tending to be a bit more optimistic (i.e., higher scores) than males, though both groups tended towards rather strong agreement in their units’ cohesiveness.

For all of the statements (Q1-Q5), we see a majority of scores from both genders at level “9”—the strongest agreement with each statement, and a high indicator of unit cohesion. For the four questions where we saw a significant difference between the genders (“Q1-Unity”, “Q2-Responsibility”, “Q3-Aspirations”, and “Q5-Comm”), a small but visible percentage of responses were present in the “6” to “8” range for males (representing between 24.4% and 31.1% of their responses). By comparison, female scores in this same range were fewer, only representing between 8.3% and 16.7% of their responses, respectively.

For the one question where we did not see a significant difference (i.e., “Q4-Help”), both males and females showed the same pattern as seen among the males’ responses to other questions: the majority of responses (71.4% and 73.3%, for males and females, respectively) at the strongest level of agreement, “9”, with a small but visible percentage in that same “6” to “8” range (23.5% and 23.3%, respectively).The increase in these lower levels among females suggest they felt a bit more pessimistic—at least on occasion—about their units’ willingness to help those who struggled.

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement

(Figure 3). The average values (by gender) are reflective of our statistical results—with generally high perceptions of unit cohesion across the board, but with males slightly more pessimistic than females on the majority of survey statements.

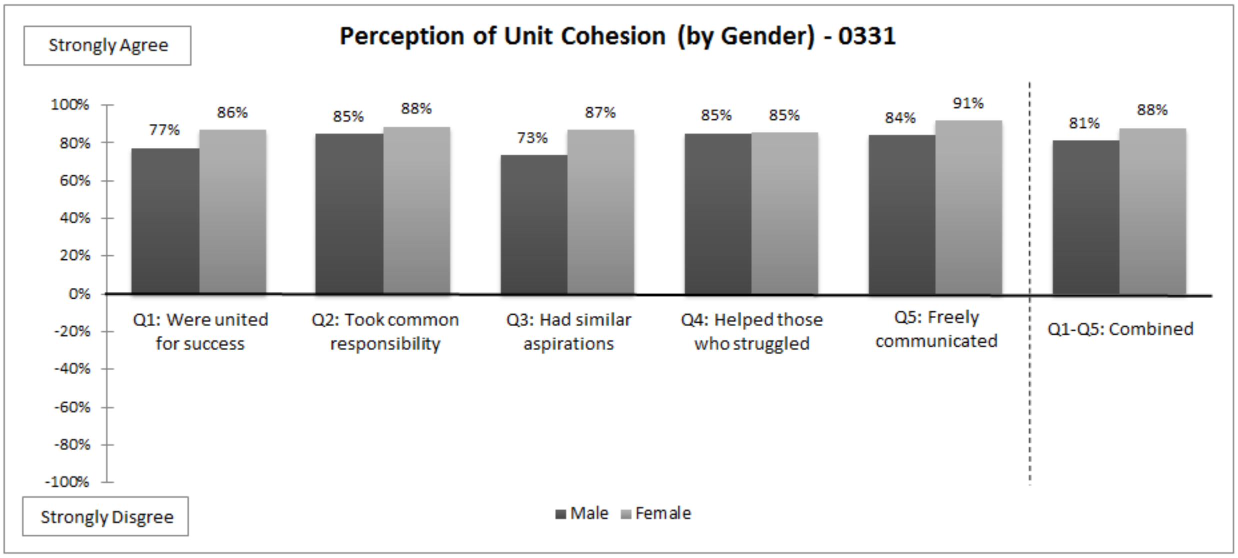


Figure 3 - Perception of Unit Cohesion (by Gender) – 0331

N.4.3.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For 0331s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) Machine Gun Teams. We have unit cohesion survey results (five statements per survey) from six males (six in the Control and Low-Density groups, and five in the High-Density group); statistical results are summarized in the table below (Table 20).

Table 20 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 0331

MOS	Cohesion	Gender	N	Mean	SD	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0331	Q1: Group was united in trying to reach goals	C	62	8.77	1.09	46.60	<0.01*	(LD-C) -6.04	<0.01*
		LD	36	7.58	1.70			(HD-C) -6.21	<0.01*
		HD	21	6.90	2.36			(HD-LD) -0.93	0.35
	Q2: We all take responsibility for task success of the group	C	62	8.94	0.40	42.96	<0.01*	(LD-C) -6.09	<0.01*
		LD	36	7.94	1.17			(HD-C) -5.98	<0.01*
		HD	21	7.48	2.04			(HD-LD) -0.49	0.63
	Q3: Group members have similar aspirations for success	C	62	8.69	1.21	45.98	<0.01*	(LD-C) -5.88	<0.01*
		LD	36	7.31	1.77			(HD-C) -6.11	<0.01*
		HD	21	6.71	2.26			(HD-LD) -0.91	0.36
	Q4: If members had problems with a task, all wanted to help	C	62	8.92	0.52	38.92	<0.01*	(LD-C) -6.07	<0.01*
		LD	36	7.94	1.33			(HD-C) -5.28	<0.01*
		HD	21	7.62	2.29			(HD-LD) 0.28	0.78
	Q5: Members communicated freely about responsibilities	C	62	8.94	0.31	39.30	<0.01*	(LD-C) -5.68	<0.01*
		LD	36	7.94	1.33			(HD-C) -5.74	<0.01*
		HD	21	7.29	2.33			(HD-LD) -0.70	0.48
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	62	44.26	3.35	47.95	<0.01*	(LD-C) -6.25	<0.01
		LD	36	38.72	6.10			(HD-C) -6.31	<0.01
		HD	21	36.00	10.12			(HD-LD) -0.78	0.44

We found significant differences between the Control and both Low- and High-Density groups for all five cohesion statements, as well as for their overall composite cohesion scores. However, we did not see a statistically significant difference between the Low- and High-Density groups.

When we look more closely at the distributions for all three groups, we see very sharp peaks at level "9" (i.e., Strongly Agree) across all statements for Control groups (from 90% to 97% of all responses). By comparison, the integrated groups have only a fraction of "9s" (ranging from 33% to 45% for Low-Density groups, and from 19% to

48% for High-Density groups), with more (though sporadic) representation between the “5” and “8” levels. In other words, males appear to have felt strongly cohesive when in an all-male unit, but their perceptions dropped to neutral to moderate levels when in an integrated group.

We also graphed the means for each statement as a percentage of how strongly males agreed (positive) or disagreed (negative) with each question’s statement while in each group (Figure 4).

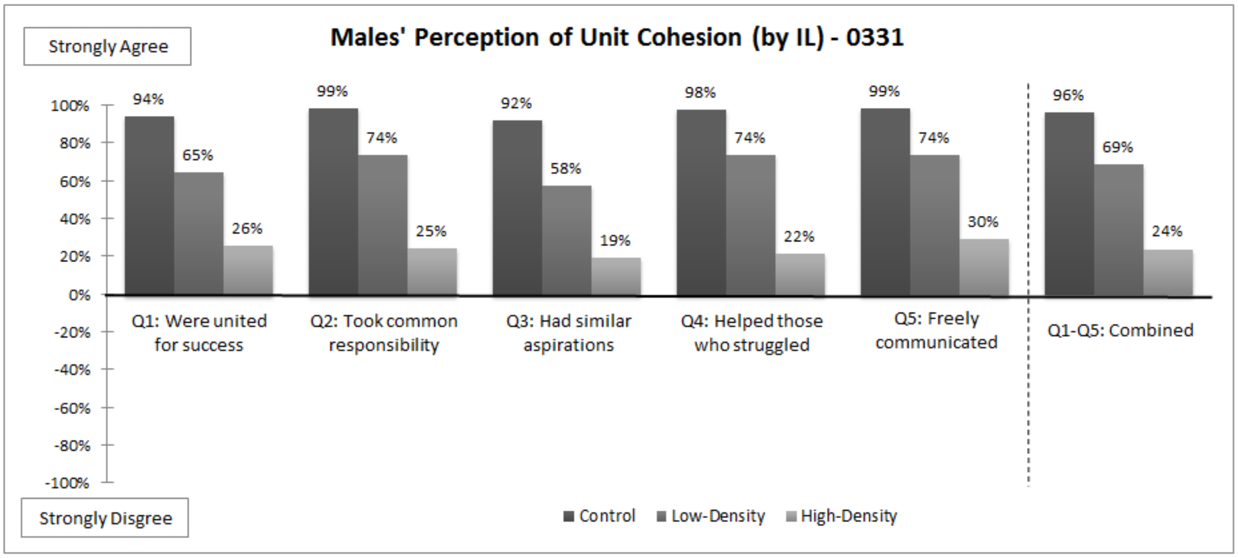


Figure 4 - Perception of Unit Cohesion (by Integration Level) - 0331

The average values (by integration level) are generally consistent[6] with what we saw with regards to males’ very high perceptions of unit cohesion within the Control group, along with the more pessimistic scores they reported while in Low-Density and High-Density groups.

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group—especially for a shift so large—but the experiential differential between males and females may have been a factor. We should also keep in mind that these are the result of only six Marines, who may or may not be representative for the larger 0331 community.

Incidentally, these are profoundly different perceptions—especially for the High-Density group—from the average values reported by females (see Figure 3). Given the very small unit size, and that females would have represented a significant (or even major)

[6] While there appears to be a large difference between the Low-Density and High-Density groups’ medians, we cannot say this is a significant difference using the Kruskal-Wallis (rank sum) test. We did not find a statistical difference between the two integrated groups’ responses.

percentage of the total unit, it may be some of the males' perceptions may have been adversely affected by a different (vice a truly non-cohesive) group dynamic.

N.4.4 Infantry Mortarman (MOS 0341)

N.4.4.1 Fatigue Results

For the 0341 volunteers at Twentynine Palms, we have baseline, mid-trial, and final fatigue self-reports from eight males[7] and four females. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \le 0.10$ for statistical significance (*).

N.4.4.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial), final (post-trial, for both days in the run-cycle), and post-7km Hike (mid-trial, Day 2) fatigue levels. Results are summarized in the table below (Table 21).

Table 21 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) – 0341

MOS	Fatigue	Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0341	Baseline (Pre-Trial)	1,2	M	256	3 / 1	1.9	0.17	(F-M) -1.4	0.17
			F	91	2 / 2				
	Post-7km (Mid-Trial)	2 "Defense"	M	128	3 / 3,4	0.02	0.88	(F-M) 0.15	0.88
			F	43	4 / 4				
	Final (Post-Trial)	1 "Offense"	M	130	4 / 4	8.08	<0.01*	(F-M) -2.84	<0.01*
			F	46	4 / 4				
		2 "Defense"	M	128	4 / 3	0.28	0.60	(F-M) -0.52	0.60
			F	45	3 / 3				

N.4.4.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We did not find a significant difference between female and male 0341 volunteers' pre-trial fatigue scores. However, we did observe a few peculiarities between the two populations (Table 22).

For example, where female pre-trial responses were tightly clustered at "2-Very Lively" and "3-Okay" (90.1%), males had a more shallow distribution spread between "1-Fully Alert" and "3-Okay" (64.8%) with a secondary, smaller peak at "7-Exhausted" (14.5%).

[7] One male only participated in two trials.

Looking more closely at the source of this secondary peak in males, we found that four males (out of eight) repeatedly reported baseline fatigue scores of “6-Extremely Tired” or “7-Exhausted” (18.8%) throughout the course of the study. In comparison, female reports in this same range represented only 1.1%--or a single response. This suggests that a substantial number of males in this MOS felt they were seriously fatigued (at times) before they even began a trial.

In a cursory exam of discrete trials where volunteers reported being “6-Extremely Tired” or “7-Exhausted” during the pre-trial survey, we found that two-thirds (66.7%) of the time they reported being only “5-Moderately Tired” or less by the final fatigue survey—i.e., they felt more energized after the trial. However, even the nearly[8] one-third (31.3%) of the time when they also reported being highly fatigued (i.e., “6” or “7”) after a trial, they had obviously completed it. This suggests these males’ high-fatigue scores may have had a major mental component, rather than being purely physical[9] in nature.

N.4.4.1.1.2 Final Fatigue Scores (by Gender)

In looking at the volunteers’ post-trial responses, we found mixed results. We found a significant difference between genders following the Offense scenario, but not for the Defense scenario. The majority of female post-trial responses for Offense trials were clustered between “3-Okay” and “4-A Little Tired” (76.1%), whereas the corresponding male responses were more loosely spread between “3-Okay” and “5-Moderately Tired” (68.8%). We also found a number of male responses (15.6%) at the highest range (“6-Extremely Tired” and “7-Exhausted”) that were not reported for females.

It is unclear why males may have been more likely to feel a bit more fatigued than females, but only for the Offense scenario. It may be that the circumstances that led to half of the males reporting very high pre-trial fatigue levels (“6-Extremely Tired”, “7-Exhausted”)—be they physical or mental—were compounded by the more rapid pace required by the Offense scenario.

For both males and females, we saw the majority of post-trial fatigue responses for Day 2/Defense largely at “3-Okay” or “4-A Little Tired”.

N.4.4.1.1.3 Post-7km Hike Fatigue Scores

For the Defense trial scenario, 0341 volunteers also completed a fatigue survey following the completion of their 7km Hike under load; we did not find a significant difference in the distributions of male and female responses for survey.

[8] One volunteer who reported a pre-trial fatigue of “7-Exhausted” did not complete a final survey.

[9] In bio-mechanical terms, “fatigue” typically refers a temporary (i.e., non-pathological) reduction in the ability of muscle fibers to generate force (i.e., contract), such as after vigorous exercise.

Table 22 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) - 0341

Distribution of 0341 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
0341	29Palms/1&2 Pre-Trial	M	256	25.4%	24.2%	15.2%	8.6%	7.8%	4.3%	14.5%
		F	91	3.3%	59.3%	30.8%	3.3%	2.2%	0.0%	1.1%
	29Palms/2 Post-7km Hike	M	128	6.3%	9.4%	24.2%	24.2%	20.3%	11.7%	3.9%
		F	43	0.0%	4.7%	30.2%	34.9%	23.3%	4.7%	2.3%
	29Palms/1 Post-"Offense"	M	130	3.1%	4.6%	25.4%	30.0%	20.0%	13.1%	3.8%
		F	46	4.3%	10.9%	26.1%	50.0%	8.7%	0.0%	0.0%
	29Palms/2 Post-"Defense"	M	128	4.7%	11.7%	32.8%	21.9%	21.1%	3.1%	4.7%
		F	45	0.0%	11.1%	42.2%	28.9%	13.3%	4.4%	0.0%

N.4.4.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 0341s, we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 2 females) Mortar Teams or Squads. We have mid-trial and final fatigue self-reports from eight males; statistical results are summarized in the table below (Table 23).

Table 23 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) - 0341

MOS	Fatigue	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
0341	Post-7km (Mid-Trial)	29Palms/2 "Defense"	C	83	4 / 4	8.2	<0.01*
			HD	45	3 / 3		
	Final (Post-Trial)	29Palms/1 "Offense"	C	84	4 / 4	0.54	0.46
			HD	46	4 / 4		
		29Palms/2 "Defense"	C	83	4 / 3	6.0	0.02*
			HD	45	3 / 4		

N.4.4.1.2.1 Post-7km Hike Fatigue Scores

Following the 7km Hike under load (Defense scenario trials), we found the distribution of males' fatigue responses in High-Density Mortar Squads to be significantly different from those in Control groups—towards less fatigued.

Instead of a fairly even spread in the "3-Okay" to "5-Moderately Tired" range seen in Control units, males in the integrated group were slightly more likely to respond with just "3-Okay" or "4-A Little Tired" (up to 51.1% from 47.0%), with an additional bump in responses in the lowest ranges (up to 24.4% from 10.8%). Males in the Control group were also much more likely to report being “6-Extremely Tired” or “7-Exhausted” (20.5%), than those in the integrated group (6.6%)

One potential reason for this difference might be the relative paces set by the two groups. High-Density groups typically took more time to complete the 7km hike (see Annex D)—perhaps setting a more manageable speed than some Control groups, for some of the males.

N.4.4.1.2.2 Final Fatigue Scores

Post-trial fatigue results by Integration Level for 0341s proved to be mixed—males’ responses were significantly different between the two groups (Control and High-Density) for the Defense, but not for the Offense, scenario.

As with the distributions seen following the 7km Hike, for post-Defense trials we see a generalized shift towards lower fatigue levels for males in integrated groups—more likely to respond with a fatigue score of “3-Okay” or lower (62.2%) than Control (42.2%).

Within the Defense scenario, the 7km Hike seems the mostly likely candidate driving the post-trial difference between Control and High-Density groups. However, if a slower pace could result in some males feeling less fatigued post-7km hike (and hence post-trial) within the Defense scenario, we might expect to see a similar effect for Offense trials, which also showed slower movements (on average) for integrated groups (see Annex D)—but we don’t.

It may be that the slower movements for Offensive trials were too short—as well as interspersed by non-movement tasks—to sufficiently influence males’ fatigue levels. If relative pace and duration was driving this difference, we might expect a similar pattern to present—or strengthen—for tasks of greater duration and/or which drive a greater difference in pace (e.g., longer hikes).

Pack-weight may also have been a contributing factor (which included the 81-mm mortar system split between the unit), so lighter pack-weights (or those scaled to a Marine’s relative body-mass) over a similar 7km hike distance, might reduce the number of males in Control groups reporting “6-Extremely Tired” or “7-Exhausted”.

Table 24 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) - 0341

Distribution of 0341 Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0341	29Palms/2 Post-7km Hike	C	83	3.6%	7.2%	20.5%	26.5%	21.7%	15.7%	4.8%
		HD	45	11.1%	13.3%	31.1%	20.0%	17.8%	4.4%	2.2%
	29Palms/1 Post-"Offense"	C	84	0.0%	4.8%	27.4%	28.6%	22.6%	13.1%	3.6%
		HD	46	8.7%	4.3%	21.7%	32.6%	15.2%	13.0%	4.3%
	29Palms/2 Post-"Defense"	C	83	2.4%	8.4%	31.3%	24.1%	25.3%	3.6%	4.8%
		HD	45	8.9%	17.8%	35.6%	17.8%	13.3%	2.2%	4.4%

N.4.4.2 Workload Results

For the 0341 volunteers at Twentynine Palms, we have max workload self-reports from eight males[10] and four females. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). If we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.4.2.1 Workload Results by Gender

We first looked at looking at how maximum workload self-reports compared between males and females, for both days in their run-cycle. Results are summarized in the table below (Table 25)

Table 25 - Summary of Maximum Workload Scores (by Gender) – 0341

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
0341	Max Workload for Trial	29Palms/1 "Offense"	M	130	4 / 4	1.5	0.22
			F	46	4 / 4		
		29Palms/2 "Defense"	M	129	4 / 4	0.79	0.37
			F	45	4 / 4		

We did not find a significant difference between female and male 0341 volunteers' maximum workload scores for either day. Both groups typically had maximum workload

[10] One male only participated in only two trials.

scores clustering between “3-Active by Easy” and “5-Extremely Tired” (77.5% to 83.8% and 88.9% to 93.5% for males and females, respectively Table 26).

For the highest maximum workload scores (i.e., “6-Overloaded”, “7-Unmanageable”), we also found a small, but not inconsequential number of reports (12.3% to 14.7% among males, and 6.5% to 11.1% among females) suggesting some Marines felt overwhelmed by their workload. When we looked more closely at the contributors to these scores, we see that three of the five male[11] contributors had quite a few reports (eight to 16 reports each), as did the sole female contributor (eight reports).

Given both the small size of this volunteer pool (only eight males and four females), and the signs that a substantial number may have struggled—at times—to accomplish the necessary workload, we should be cautious in extrapolating the 0341 study results to the larger Marine Corps community.

Table 26 - Maximum Workload Score Distributions (by Gender) - 0341

Distribution of 0341 Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0341	29Palms/1 "Offense"	M	130	0.0%	3.8%	19.2%	40.8%	23.8%	6.9%	5.4%
		F	46	0.0%	0.0%	4.3%	56.5%	32.6%	6.5%	0.0%
	29Palms/2 "Defense"	M	129	3.9%	3.9%	19.4%	30.2%	27.9%	10.9%	3.9%
		F	45	0.0%	0.0%	4.4%	57.8%	26.7%	11.1%	0.0%

N.4.4.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males’ responses remained consistent at different integration levels. For 0341s, we compared males’ responses in Control (i.e., all-male) and High-Density (i.e., 2 females) Mortar Teams and Squads. We have maximum workload self-reports from eight males; statistical results are summarized in the table below (Table 27).

[11] One of the male contributors who only reported a single max workload score in the highest range only participated in a single run-cycle.

Table 27 - Summary of Males' Max Workload Scores (by Integration Level) - 0341

MOS	Workload	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value
0341	Max Workload for Trial	29Palms/1 "Offense"	C	84	4 / 4	0.55	0.46
			HD	46	4 / 4		
		29Palms/2 "Defense"	C	83	5 / 5	10.5	<0.01*
			HD	46	4 / 3		

We found mixed results when we looked at males' maximum workload responses between Control and High-Density groups—statistically significant for Defense trials but not for Offense trials.

For the Defense trials, males in a High-Density group were more likely to have maximum workload responses clustered at the slightly lower "3-Active but Easy" and "4-Challenging" range (54.3%), than males in Control groups, who more responded in the "4-Challenging" and "5-Extremely Busy" range (66.3%) (Table 28).

As for possible mechanisms, it is unlikely that a slower tempo is a factor, since the more workload-intensive portion of the Defense scenario (i.e., 81-mm Mortar Engagement) did not reveal any difference in times between Control and integrated groups (Annex D).

However, the difference in reported maximum workloads trials for Defense trials is consistent with the higher fatigue results reported by males in Control groups—for Defense (and not Offense) trials (see Table 24)—given that even relatively low increases in fatigue levels can degrade concentration and performance.

Table 28 - Males' Max Workload Score Distributions (by Integration Level) - 0341

Distribution of 0341 Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0341	29Palms/1 "Offense"	C	84	0.0%	0.0%	21.4%	41.7%	23.8%	7.1%	6.0%
		HD	46	0.0%	10.9%	15.2%	39.1%	23.9%	6.5%	4.3%
	29Palms/2 "Defense"	C	83	1.2%	1.2%	14.5%	32.5%	33.7%	13.3%	3.6%
		HD	46	8.7%	8.7%	28.3%	26.1%	17.4%	6.5%	4.3%

N.4.4.3 Cohesion Results

For the 0341 volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) from eight males and four females. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.4.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 29).

Table 29 - Summary of Unit Cohesion Scores (by Gender) - 0341

MOS	Cohesion	Gender	N	Mean	SD	χ2 Statistic	χ2 p-Value
0341	Q1: Group was united in trying to reach goals	M	130	7.98	1.87	7.34	<0.01*
		F	45	7.62	1.76		
	Q2: We all take responsibility for task success of the group	M	130	8.00	1.86	3.25	0.07*
		F	45	8.02	1.22		
	Q3: Group members have similar aspirations for success	M	130	7.92	2.02	5.54	0.02*
		F	45	7.69	1.58		
	Q4: If members had problems with a task, all wanted to help	M	130	8.01	1.90	4.39	0.04*
		F	45	7.76	1.65		
	Q5: Members communicated freely about responsibilities	M	130	8.01	1.93	3.56	0.06*
		F	45	8.02	1.14		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	M	130	39.92	9.39	4.03	0.04*
		F	45	39.11	6.18		

We found a significant difference between 0341 males and females for all five cohesion statements ("Q1-Unity", "Q2-Responsibility", "Q3-Aspirations", "Q4-Help" and "Q5-Comm"), as well as for the overall composite cohesion score. When we look more closely at the distribution of 0341 responses for each statement, we see males tending to be a bit more optimistic (i.e., higher scores) than females, though both groups have the majority of their responses in the top three levels (towards strongly agree).

For males, their highest contributions occur at the highest level (ranging from 66.9% to 68.5%) and tapering off quickly over the next few lower levels, but with non-zero responses (1% to 2%) all the way down to “1” (i.e., Strongly Disagree). Females show far fewer contributions at level “9” (ranging from 35.6% to 46.7%) for any question—with much more shallow tapers at “7” and “8”—but fewer responses below “5”.

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement (Figure 5).

The average values (by gender) shown reflect the generally high perception of unit cohesion we found across the board for both genders, with males a bit more optimistic than females for most statements. Where the means for males appear to be on par with females’ responses, we can see the influence of the relatively few (but profoundly) negative (i.e., towards Strongly Disagree) cohesion scores that males’ reported.

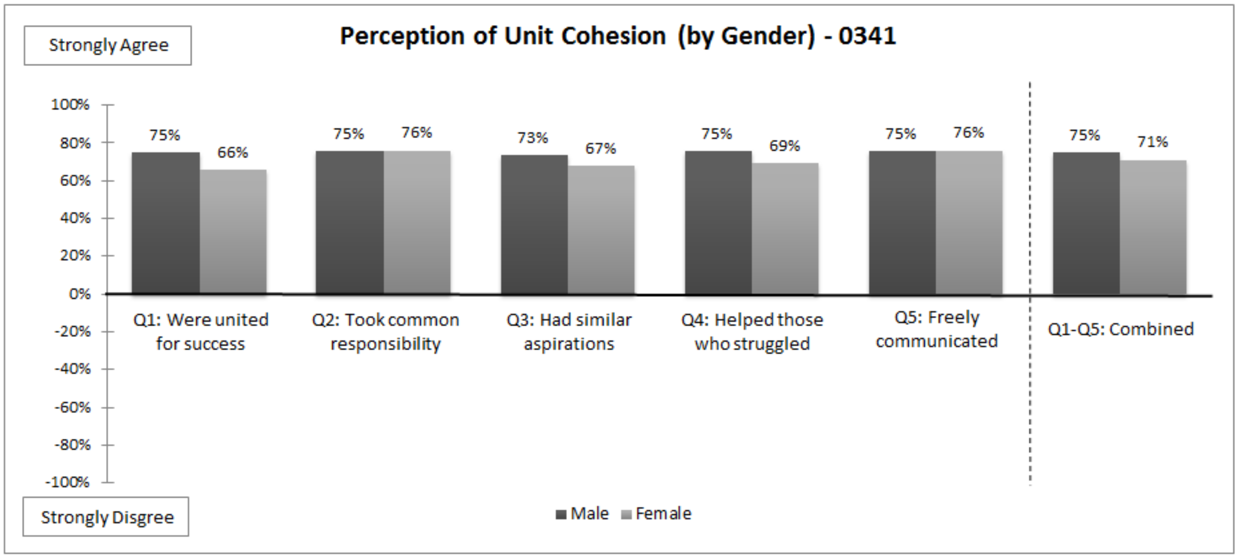


Figure 5 - Perception of Unit Cohesion (by Gender) – 0341

N.4.4.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males’ cohesion responses remained consistent at different integration levels. For 0341s, we compared males’ responses in Control (i.e., all-male) and High-Density (i.e., 2 females) Mortar Teams and Squads. We have unit cohesion survey results (five statements per survey) from eight males; statistical results are summarized in the table below (Table 30).

Table 30 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 0341

MOS	Cohesion	IL	N	Mean	SD	χ2 Statistic	χ2 p-Value
0341	Q1: Group was united in trying to reach goals	C	84	8.51	1.11	14.22	<0.01*
		HD	44	7.11	2.38		
	Q2: We all take responsibility for task success of the group	C	84	8.51	1.12	13.13	<0.01*
		HD	44	7.18	2.37		
	Q3: Group members have similar aspirations for success	C	84	8.51	1.12	12.43	<0.01*
		HD	44	6.95	2.66		
	Q4: If members had problems with a task, all wanted to help	C	84	8.50	1.15	10.45	<0.01*
		HD	44	7.22	2.46		
	Q5: Members communicated freely about responsibilities	C	84	8.54	1.10	11.85	<0.01*
		HD	44	7.16	2.54		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	84	42.57	5.58	13.91	<0.01*
		HD	44	35.64	11.97		

We found a significant difference between the Control and High-Density groups for all five cohesion statements, as well as for the overall composite cohesion score. When we look more closely at the distributions, we see a single sharp peak at level "9" when in the Control group (76%, for all statements), and much shorter peaks (52% to 57%), though still largely singular, in the High-Density group. For the integrated group, we also see small, but steady responses (2% to 18%) all the way down to "1" not seen in Control groups.

We also graphed the means for each statement as a percentage of how strongly males agreed (positive) or disagreed (negative) with each question's statement while in each group (Figure 6).

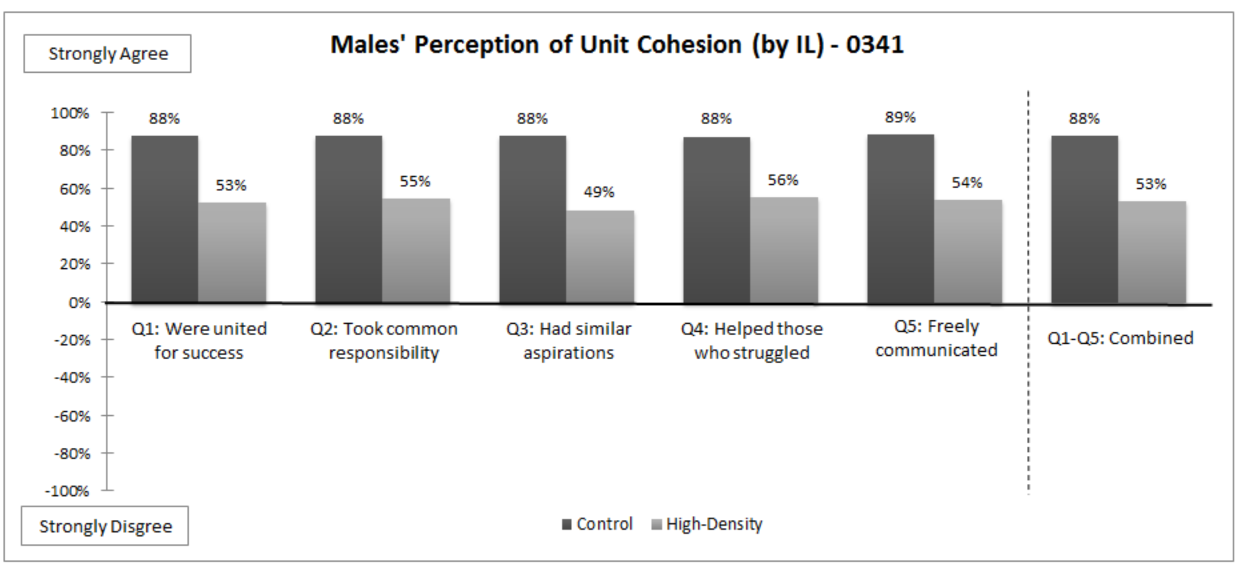


Figure 6 - Males' Perception of Unit Cohesion (by Integration Level) - 0341

The average values shown are consistent with what we saw with regards to males' very high perceptions of unit cohesion within the Control group, along with the sharp decline in their perception of cohesion when in the High-Density group.

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group, but differences in relative experience between males and females may have been a factor. We should also keep in mind that these are the result of only eight Marines, who may or may not be representative for the larger 0341 community.

Incidentally, these results differ quite a bit from how females rated their perception of unit cohesiveness in integrated groups; they were quite a bit more optimistic (see Figure 5). With the small unit size—and the fact that males would have had equal numbers with the females—some of their perceptions may have been colored by a different (vice a truly non-cohesive) group dynamic.

N.4.5 Infantry Assaultman (0351) / Antitank Missileman (0352)

N.4.5.1 Fatigue Results

For the 035X volunteers at Twentynine Palms, we have baseline, mid-trial, and final fatigue self-reports from six males and six females[12]. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.5.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial), final (post-trial, for both days in the run-cycle), and post-7km Hike (mid-trial, Day 2) fatigue levels. Statistical results are summarized in the table below (Table 31).

[12] One female only participated in two trials.

Table 31 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) – 035X

MOS	Fatigue	Day	Gender	N	Median / Mode	χ2 Statistic	χ2 p-Value
035X	Baseline (Pre-Trial)	1,2	M	189	2 / 2	65.8	<0.01*
			F	149	3 / 2		
	Post-7km (Mid-Trial)	2 “Defense”	M	94	3 / 3	38.5	<0.01*
			F	70	4 / 4		
	Final (Post-Trial)	1 “Offense”	M	96	3 / 2	68.1	<0.01*
			F	73	4 / 4		
		2 “Defense”	M	95	2 / 2	49.5	<0.01*
			F	71	3 / 3		

N.4.5.1.1.1 Pre-Trial Fatigue Scores (by Gender)
We found a significant difference between female and male 035X volunteers’ pre-trial fatigue scores, with males more likely to report being slightly less fatigued. Males’ responses typically clustered around “1-Fully Alert” (35.4%) and “2-Very Lively” (43.9%), whereas females were more likely to report being “2-Very Lively” (38.9%) or “3-Okay” (36.2%) before the start of a trial (Table 32).

We found very few of the higher fatigue scores (i.e., “5-Moderately Tired”, “6-Extremely Tired”, or “7-Exhausted”) for either gender—only 3.2 percent of male responses (driven largely by one Marine), and 2.0 percent of female responses (one response each by two Marines)—and no obvious pattern of increasing scores over time, suggesting volunteers largely had sufficient recovery time between trials.

N.4.5.1.1.2 Final Fatigue Scores (by Gender)
In looking at the volunteers’ post-trial responses, we found significant differences in the distribution of scores for both days of the 035X run-cycle (i.e., Offense and Defense), in favor of generally lower fatigue scores for males. Most male responses were clustered within the “2-Very Lively” to “3-Okay” range for Offense trials (84.4%) and “1-Fully Alert” to “3-Okay” for Defense trials (95.8%).

On the other hand, most females’ responses were shifted slightly higher, with post-Offense scores strongly clustered from “3-Okay” to “5-A Little Tired” (91.8%) and post-Defense scores from “2-Very Tired” to “A Little Tired” (84.5%). Female also had greater (though not a lot of) representation in the highest ranges (i.e., “6-Extremely Tired” and “7-Exhausted”) on both days—with 6.8% post-Offense, 4.2% post-Defense—than males (2.1% and zero, respectively).

N.4.5.1.1.3 Post-7km Hike Fatigue Scores (by Gender)
For the Defense trial scenario, 035X volunteers also completed a fatigue survey following the completion of their 7km Hike under load. As with post-trial results, males

were more likely to report lower fatigue levels post-hike, clustered between "2-Very Lively" and "4-A Little Tired" (85.1%); females had a much smaller majority in this same range (64.3%), with a sizeable percentage at "5-Moderately Tired" and above (35.7%).

Both genders' responses post-Defense appear shifted to the left (towards lower fatigue) compared to their post-7km Hike surveys for that same day. This suggests many Marines felt more fatigued by the 7km Hike than by the follow-on Defensive tasks (e.g., TOW live-fire engagement and mounting/dismounting TOW).

Table 32 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) – 035X

Distribution of 035X Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
035X	29Palms/1&2 Pre-Trial	M	189	35.4%	43.9%	11.1%	6.3%	3.2%	0.0%	0.0%
		F	149	2.7%	38.9%	36.2%	20.1%	1.3%	0.7%	0.0%
	29Palms/2 Post-"7km Hike"	M	94	3.2%	24.5%	40.4%	20.2%	8.5%	2.1%	1.1%
		F	70	0.0%	1.4%	18.6%	44.3%	24.3%	10.0%	1.4%
	29Palms/1 Post-"Offense"	M	96	3.1%	45.8%	38.5%	8.3%	2.1%	2.1%	0.0%
		F	73	0.0%	1.4%	30.1%	31.5%	30.1%	6.8%	0.0%
	29Palms/2 Post-"Defense"	M	95	24.2%	46.3%	25.3%	2.1%	2.1%	0.0%	0.0%
		F	71	1.4%	19.7%	40.8%	23.9%	9.9%	4.2%	0.0%

N.4.5.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 035Xs, we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 2 or 3 females) Assault or Anti-Armor Squads. We have mid-trial and final fatigue self-reports from six males; statistical results are summarized in the table below (Table 33).

Table 33 – Summary of Mid-Trial and Final Fatigue Scores (by Integration Level) – 035X

MOS	Fatigue	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value
035X	Post-7km (Mid-Trial)	29Palms/2 "Defense"	C	82	3 / 3	7.8	<0.01*
			HD	12	2 / 2		
	Final (Post-Trial)	29Palms/1 "Assault"	C	83	2 / 2	1.1	0.31
			HD	13	3 / 3		
		29Palms/2 "Defense"	C	83	2 / 2	10.1	<0.01*
			HD	13	1 / 2		

N.4.5.1.2.1 Post-7km Hike Fatigue Scores (by Integration Level)

Following the 7km Hike under load, we found males' fatigue responses in High-Density squads to be significantly different (towards less fatigued)—with a majority of responses in the "1-Fully Alert" to "2-Very Lively" range (58.3%) compared to the Control group (23.2%) (Table 34).

Both groups had a sizeable percentage in the mid-range levels ("3-Okay" to "5-Moderately Tired"), but this was more pronounced in Control responses (73.2%) than in High-Density responses (41.7%). The Control groups also had a small number of responses (3.7%) in the highest levels ("6-Extremely Tired", "7-Exhausted"), suggesting some males in Control groups were significantly fatigued (at times).

One potential reason for the shift in males' responses in integrated groups could be their (typically) slower hike times (see Annex E). Combined with the presence of more higher-end fatigue levels ("5-Moderately Tired" to "7-Exhausted) in the Control group (13.4%), this would suggest some 035X males were not comfortable with the faster pace almost always seen among all-male groups.

From the gender results above—where a substantial number of female responses (35.7%) fell in the higher-end levels ("5-Moderately Tired" and above), we should note that most of these responses (19 out of 25) occurred in "high-density" units composed entirely of females.

When we re-evaluate the gender results above—but looking only at High-Density squads with at least one male (i.e., same trials used in Table 33)—then we find that the number of females responding with higher-end scores drops (to 25% of total responses), with most responses at "3-Okay" and "4-A Little Tired". In other words, females in High-Density squad with at least one male (i.e., not all-female) have fewer reports of very high fatigue levels. This suggests that males in integrated groups may have been providing some type of compensation that also reduced females' fatigue levels (to a degree).

N.4.5.1.2.2 Final Fatigue Scores (by Integration Level)

Post-trial fatigue results by Integration Level proved to be mixed—males' responses were significantly different between the two groups (i.e., Control and High-Density) for the Defense scenario trials, but not for the Offense scenario. Similar to the post-7km Hike results—but stronger—males executing Defense trials in High Density groups reported more "1-Fully Alert" responses (66.7%) than Control groups (18.1%).

A comparison of each volunteer's responses for both the post-7km Hike and post-trial surveys reveal that males were overwhelmingly less fatigued by the end of the trial than after the hike (77.7%), with the remaining (22.3%) indicating they were as tired (i.e., no more tired) by the end of the trial, regardless of which integration level of their group.

This suggests that the 7km Hike was the primary fatiguing event for the Defense scenario for males, and as such, the significantly lower post-trial fatigue levels found for High-Density groups could well be an extension of the 7km Hike driver—i.e., perhaps slower pace—rather than the result of an additional (or alternate) mechanism.

Table 34 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) – 035X

Distribution of 035X Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
035X	29Palms/2 Post-7km Hike	C	82	1.2%	22.0%	41.5%	22.0%	9.8%	2.4%	1.2%
		HD	12	16.7%	41.7%	33.3%	8.3%	0.0%	0.0%	0.0%
	29Palms/1 Post-"Offense"	C	83	3.6%	48.2%	34.9%	8.4%	2.4%	2.4%	0.0%
		HD	13	0.0%	30.8%	61.5%	7.7%	0.0%	0.0%	0.0%
	29Palms/2 Post-"Defense"	C	83	18.1%	49.4%	27.7%	2.4%	2.4%	0.0%	0.0%
		HD	12	66.7%	25.0%	8.3%	0.0%	0.0%	0.0%	0.0%

N.4.5.2 Workload Results

For the 035X volunteers at Twentynine Palms, we have maximum workload self-reports from six males and six females[13]. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

[13] One female only participated in only two trials.

N.4.5.2.1 Workload Results by Gender

We first looked at looking at how maximum workload self-reports compared between males and females, for both days in their run-cycle. Statistical results are summarized in the table below (Table 35)

Table 35 - Summary of Maximum Workload Scores (by Gender) – 035X

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
035X	Max Workload for Trial	29Palms/1 “Offense”	M	93	3 / 3	55.7	<0.01*
			F	72	4 / 4		
		29Palms/2 “Defense”	M	94	4 / 3	44.0	<0.01*
			F	73	4 / 4		

We found a significant difference between female and male 035X volunteers’ maximum workload scores. For both days, males heavily favored scores between “3-Active but Easy” and “4-Challenging” (91.4% and 87.2% for Offense and Defense, respectively), whereas females most often reported in a slightly higher workload range, between “4-Challenging” and “5-Extremely Busy” (73.6% and 76.7%).

Females were also more likely to report peak scores in the highest range (i.e., “6-Overloaded” and “7-Unmanageable”), with five out of six Marines contributing at least once, and a few multiple times); by comparison, only one male Marine reported in this high-end range. However, all instances of female reports[14] of “6-Overloaded” and above—indeed most female responses overall—were in all-female “High-Density” units. As such, it may well be that lack of experience within their unit was a major influence in females’ perception of workload demands, and in the gender difference seen here.

Table 36 - Max Workload Score Distributions (by Gender) – 035X

Distribution of 035X Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
035X	29Palms/1 "Offense"	M	93	0.0%	5.4%	64.5%	26.9%	3.2%	0.0%	0.0%
		F	72	0.0%	0.0%	15.3%	51.4%	22.2%	8.3%	2.8%
	29Palms/2 "Defense"	M	94	0.0%	3.2%	45.7%	41.5%	6.4%	1.1%	2.1%
		F	73	0.0%	0.0%	6.8%	47.9%	28.8%	12.3%	4.1%

[14] Incidentally, the one male that reported max workload scores of “6-Overloaded” or higher did so while in a Control (i.e., all-male), unit.

N.4.5.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males' responses remained consistent at different integration levels. For the 035Xs, we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 2 - 3 females) Assault and Anti-Armor Squads. We have maximum workload self-reports from six males; statistical results are summarized in the table below (Table 37).

Table 37 - Summary of Maximum Workload Scores (by Integration Level) – 035X

MOS	Workload	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value
035X	Max Workload for Trial	29Palms/1 "Offense"	C	82	3 / 3	1.5	0.23
			HD	11	3 / 3		
		29Palms/2 "Defense"	C	80	4 / 3	4.4	0.04*
			HD	13	3 / 3		

We found mixed results when we looked at males' max workload responses between Control and High-Density groups—statistically significant for Defense trials but not for Offense trials. When we examine the distribution of workload scores for the two groups during Defense trials (Table 38), we see a slight increase in representation at "2-Little to Do", and a lack of any representation at the higher "5-Extremely Busy" and higher, for High-Density groups (i.e., towards more manageable) not seen for Control groups.

Table 38 - Males' Maximum Workload Score Distributions (by Integration Level) – 035X

Distribution of 035X Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
035X	29Palms/1 "Offense"	C	82	0.0%	6.1%	65.9%	24.4%	3.7%	0.0%	0.0%
		HD	11	0.0%	0.0%	54.5%	45.5%	0.0%	0.0%	0.0%
	29Palms/2 "Defense"	C	81	0.0%	1.2%	44.4%	43.2%	7.4%	1.2%	2.5%
		HD	13	0.0%	15.4%	53.8%	30.8%	0.0%	0.0%	0.0%

As for possible mechanisms, It's is unlikely that a slower tempo is a factor—as has been proposed for some units showing reduced workload perceptions in integrated groups. From our quantitative results (see Annex E), High Density groups consistently showed slower times for both days. If tempo were the primary factor in shifting males' maximum workload reports between Control and High-Density groups, then we'd also expect to see a difference in reports for Offense trails—which we didn't see.

However, males' maximum workload scores are consistent with differences seen between integration levels with fatigue responses (see Table 33 and Table 34). Males' showed significantly lower fatigue levels for Defense trials—but not Offense trials—when in a High-Density group, compared with Control responses. Since even relatively low levels of fatigue can degrade concentration and performance, the lower fatigue levels for High-Density groups could shift their perception of their ability to manage their workload towards lower scores.

However, we should also note that these results are based on both a small volunteer pool (six males), and on a very small number of High-Density results (11-13 total) compared to the number available for Control groups (82-81 total). As such, we should be cautious in extrapolating these results to the larger Marine Corps.

N.4.5.3 Cohesion Results

For the 035X volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) from six males and six females[15]. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.5.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 39).

[15] One female only participated in only one run-cycle.

Table 39 - Summary of Unit Cohesion Scores (by Gender) - 035X

MOS	Cohesion	Gender	N	Mean	SD	χ2 Statistic	χ2 p-Value
035X	Q1: Group was united in trying to reach goals	M	95	7.91	1.22	15.80	<0.01*
		F	71	8.48	1.01		
	Q2: We all take responsibility for task success of the group	M	95	7.76	1.42	22.25	<0.01*
		F	71	8.54	0.92		
	Q3: Group members have similar aspirations for success	M	95	7.76	1.39	13.32	<0.01*
		F	71	8.44	0.97		
	Q4: If members had problems with a task, all wanted to help	M	95	7.76	1.50	16.81	<0.01*
		F	71	8.49	0.97		
	Q5: Members communicated freely about responsibilities	M	95	7.80	1.48	18.31	<0.01*
		F	71	8.48	1.26		
	Composite Score (sum of Q1-Q5)	M	95	39.98	6.69	20.28	<0.01*
		F	71	42.42	4.52		

We found a significant difference between 035X males and females for all five cohesion statements (“Q1-Unity”, “Q2-Responsibility”, “Q3-Aspirations”, “Q4-Help” and “Q5-Comm”), as well as for the overall composite cohesion score. When we look more closely at the distribution of responses for each statement, we see that females tend to be a lot more optimistic (i.e., higher scores) than males, though both groups have the majority of their responses in the top three levels (towards strongly agree).

The shapes of the distributions, however, are quite different between the two groups. Females have a heavy majority of responses in the highest level, “9”, for all five questions (64.8% to 74.7%), compared to the number of responses at that same level for males (33.7% to 40.0%). Males, in addition to a more shallow clustering of responses between “7” to “9”, are also more likely to have scores (in each statement) in the more neutral “5” to “6” range (12% to 13%) compared to females (1% to 4%).

We also graphed the means for each statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each question’s statement (Figure 7).

The average values—by gender—reflect the pronounced differences we saw in unit cohesion perceptions of males versus females. It’s unclear why male 035X’s might have such a reduced perception of cohesion compared to females.

One potential reason for the difference may be the influence of small group size and relative experience. Females—who were on par with one another with regards to experience—were more likely to be paired with another female (especially at the two-man team level that many tasks were conducted). This might have fostered an increased sense of cooperation—leading to the many “9” scores reported.

If equality of experience can act as a positive influence on cohesion, a difference in experience—such as what males might have felt in integrated groups—might have a corresponding dampening effect, which might also explain males' more pessimistic results.

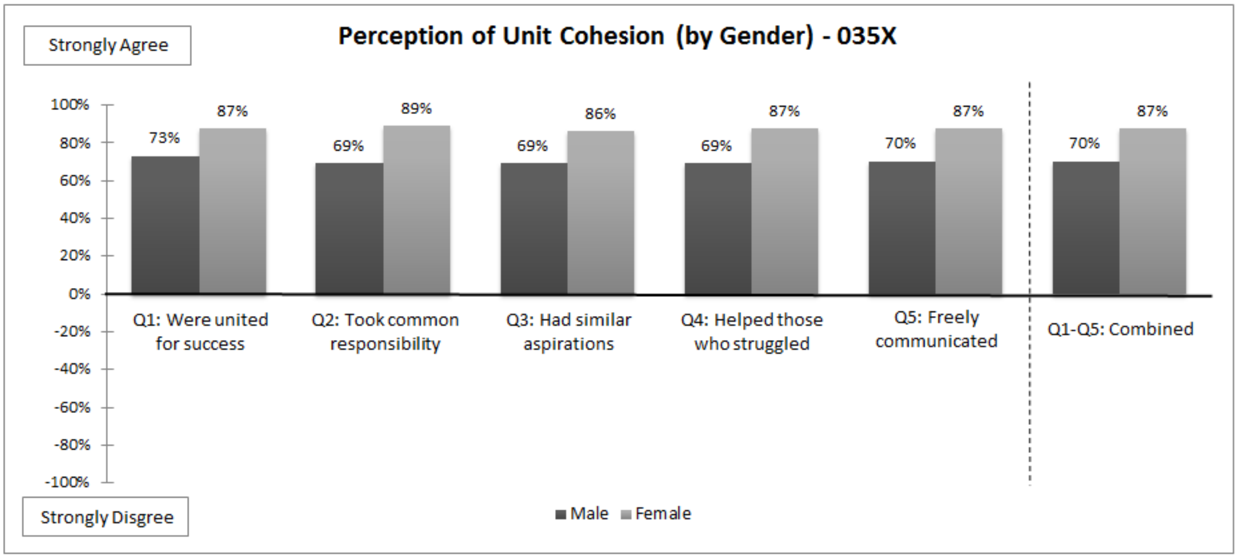


Figure 7 - Perception of Unit Cohesion (by Gender) - 035X

N.4.5.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For 035Xs, we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 2 - 3 females) Assault and Anti-Armor Squads. We have unit cohesion survey results (five statements per survey) from six males; statistical results are summarized in the table below (Table 40).

Table 40 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 0341

MOS	Cohesion	IL	N	Mean	SD	χ2 Statistic	χ2 p-Value
035X	Q1: Group was united in trying to reach goals	C	82	7.99	1.24	6.07	0.01*
		HD	13	7.38	0.96		
	Q2: We all take responsibility for task success of the group	C	82	7.88	1.43	8.30	<0.01*
		HD	13	7.00	1.15		
	Q3: Group members have similar aspirations for success	C	82	7.89	1.39	8.97	<0.01*
		HD	13	6.92	1.12		
	Q4: If members had problems with a task, all wanted to help	C	82	7.88	1.54	10.28	<0.01*
		HD	13	7.00	0.91		
	Q5: Members communicated freely about responsibilities	C	82	7.99	1.42	13.86	<0.01*
		HD	13	6.62	1.32		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	82	39.62	6.77	8.30	<0.01*
		HD	13	34.92	4.55		

We found a significant difference between the Control and High-Density groups for all five cohesion statements, as well as for the overall composite cohesion score. When we look more closely at the distributions, we see males' responses clustering largely around the two highest levels, "8" and "9", across all questions when in the Control group (ranging between 72% and 79%), suggesting high perceptions of cohesion among all-male groups.

In the High-Density group, perceptions are not so universal and we see two variations in their response pattern. For the "Q1-Unity" and "Q4-Help" questions, we actually see a sharper (85%) or comparable (77%) peak—for Q1 and Q4, respectively—within this same range, but with slower tapers (i.e., slightly greater representation) spanning "7" and "6".

For the remaining questions ("Q2-Responsibility", "Q3-Aspirations", and "Q5-Comm"), we see slightly less representation in the highest two levels (62% to 69%), but with similar slow tapers to "6". This suggests that males in integrated groups—while feeling relatively strong with regards to unity and helping members that struggle (though not as strong as in Control groups)—are less convinced with regards to taking equal responsibility, similar aspirations, or open communication.

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement (Figure 8). The average values (by integration level) shown are consistent with what we saw with regards to males' higher perceptions of unit cohesion within the Control group, along with the decline in their perception of cohesion when in the High-Density group.

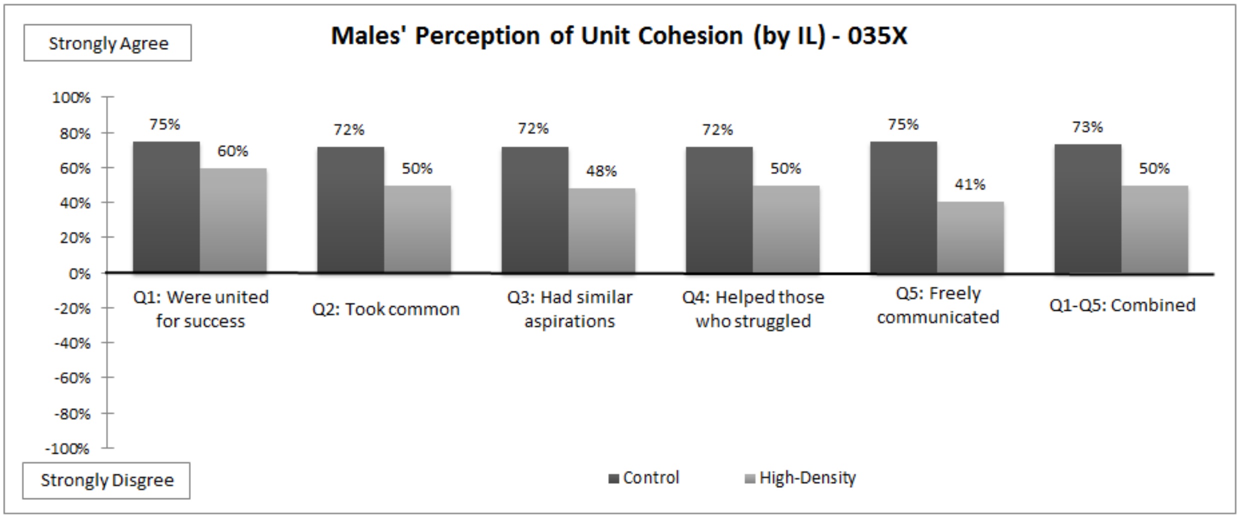


Figure 8 - Males' Perception of Unit Cohesion (by Integration Level) - 035X

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group—especially for a shift so pronounced—but differences in experience levels may have been a factor. We should also keep in mind that these are the result of only six Marines, with a very small number of reports from integrated groups (i.e., only 13, compared to the 82 reports from the Control group). As such, these results may shift with the addition of a more balanced data set, or with a different mix of Marines from the larger 035X community.

Incidentally, these results are quite different from how females rated their perception of unit cohesiveness within these same integrated groups (though, as discussed above, females' responses may have been influenced by being paired with another female of comparable experience). Also, with the small unit size—and the fact that males may often have been outnumbered by females, some of their perceptions may have been colored by a different (vice a non-cohesive) group dynamic.

N.4.6 Provisional Infantry (PI)

N.4.6.1 Fatigue Results

For the Provisional Infantry (PI) volunteers at Twentynine Palms, we have baseline, mid-trial, and final fatigue self-reports from 32 males and eight females. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p < 0.033$ to adjust for multiple comparisons).

N.4.6.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial), final (post-trial, for both days in the run-cycle), and post-7km Hike (mid-trial, Defense) fatigue levels. Statistical results are summarized in the table below (Table 41).

Table 41 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) - PI

MOS	Fatigue	Day	Gender	N	Median / Mode	χ^2 Statistic	χ^2 p-Value
PI	Baseline (Pre-Trial)	1,2	M	1135	2 / 2	123.7	<0.01*
			F	226	3 / 3		
	Post-7km (Mid-Trial)	1 "Defense"	M	564	3 / 2	36.0	<0.01*
			F	109	4 / 3,5		
	Final (Post-Trial)	2 "Offense"	M	565	3 / 4	33.4	<0.01*
			F	110	4 / 4		
		1 "Defense"	M	559	3 / 2	38.4	<0.01*
			F	108	4 / 5		

N.4.6.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We found a significant difference between female and male PI volunteers' pre-trial fatigue scores, with males more often reporting lower fatigue levels than females. Males' responses typically clustered around "1-Fully Alert" (31.1%) and "2-Very Lively" (41.4%), whereas females most often reported with the slightly higher "3-Okay" (44.7%) or "2-Very Lively" (21.2%) before the start of a trial (Table 42).

We found very few of the higher fatigue scores for either sex (i.e., "5-Moderately Tired", "6-Extremely Tired", or "7-Exhausted")—only 0.5 percent of male responses (six responses by six different Marines), and 7.1 percent of female responses (driven largely by three Marines)—and no obvious pattern of increasing scores over time, suggesting volunteers largely had sufficient recovery time between trials.

N.4.6.1.1.2 Final Fatigue Scores (by Gender)

For post-trial responses, we found significant differences for both days of the PI run-cycle (i.e., Defense, Offense), with males generally reporting lower fatigue levels than females (Table 42). Both genders tended toward relatively consistent, shallow distributions—males mostly between "2-Very Lively" and "4-A Little Tired" for both days (79.3% and 79.2%, for Defense and Offense, respectively), and females slightly higher, between "3-Okay" and "5-Moderately Tired" for both days (85.2%, 81.5%).

N.4.6.1.1.3 Post-7km Hike Fatigue Scores

For the Defense scenario, PI volunteers also completed a fatigue survey following the completion of their 7km Hike under load; we found a significant difference between male and female responses for this as well, with the same pattern and general clustering found post-trial (i.e., the distribution of males' responses shifted a bit left of females, centered largely between "2-Very Lively" and "4-A Little Tired"; females' responses centered largely between "3-Okay" and "5-Moderately Tired").

Table 42 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) - PI

Distribution of PI Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
PI	29Palms/1&2 Pre-Trial	M	1135	31.1%	41.4%	20.0%	7.0%	0.5%	0.0%	0.0%
		F	226	10.6%	21.2%	44.7%	16.4%	5.8%	0.4%	0.9%
	29Palms/1 Post-7km Hike	M	564	13.8%	27.5%	25.9%	18.6%	9.6%	4.3%	0.4%
		F	109	0.0%	16.5%	28.4%	19.3%	28.4%	4.6%	2.8%
	29Palms/2 Post-"Offense"	M	559	7.9%	28.8%	28.3%	22.2%	10.7%	2.1%	0.0%
		F	108	1.9%	12.0%	25.0%	24.1%	32.4%	4.6%	0.0%
	29Palms/1 Post-"Defense"	M	560	8.0%	26.6%	26.3%	26.4%	11.4%	1.1%	0.2%
		F	108	3.7%	6.5%	26.9%	34.3%	24.1%	1.9%	2.8%

N.4.6.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For Provisional Infantry Rifle Squads, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 2 females), and High-Density (i.e., 4-5 females) units. We have mid-trial and final fatigue self-reports from 32 males; statistical results are summarized in the table below (Table 43).

Table 43 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) - PI

MOS	Fatigue	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
PI	Post-7km (Mid-Trial)	29Palms/1 "Defense"	C	248	3 / 2	4.9	0.08*	(LD-C) -0.46	0.64
			LD	178	3 / 3			(HD-C) -2.2	0.03*
			HD	138	3 / 3			(HD-LD) -1.6	0.10
	Final (Post-Trial)	29Palms/2 "Offense"	C	246	3 / 2	3.3	0.20		
			LD	175	3 / 3				
			HD	138	3 / 2				
		29Palms/1 "Defense"	C	249	3 / 4	4.9	0.09*	(LD-C) -0.07	0.94
			LD	180	3 / 3			(HD-C) -2.1	0.04
			HD	136	3 / 2			(HD-LD) -1.8	0.07

N.4.6.1.2.1 Mid-Trial Fatigue Scores (by Integration Level)

Following the 7km Hike, we found males' fatigue responses in High-Density Rifle Squads to be significantly different (towards less fatigue) from both Low-Density and Control squads, but Low-Density squads were not significantly different from Control. When we look at the three groups respective distributions of post-hike scores (Table 44), we see that males' responses in High-Density units were more likely to fall within the lowest three fatigue levels (76.1%) than either Control (63.3%) or Low-Density (65.7%) units, along with fewer responses in any of the higher levels.

One potential reason for this shift in males' responses in integrated units (towards lower fatigue reporting) following the 7km Hike has been the slower average pace typically seen in integrated groups (see Annex F). However, this does not really explain our results; both the Low- and High-Density groups were statistically slower than Control groups on the 7km Hike (and times between the two integrated groups were not), but only High-Density groups showed the lower shift in fatigue levels here.

When we looked at how the range of responses varied among all male PIs—depending on Integration Level—we found that 12 (38%) showed no change in the range of values they'd reported over the course of the Twentynine Palms phase and six (19%) showed a shift up (towards a higher upper-bound) when in integrated units. In other words, even if some Marines might report a lower post-hike fatigue score for a particular, integrated trial, 56 percent of them had also reported values as high as (or higher) than highest value they'd reported in a Control group.

This suggests—at least for PI volunteers—that the slower pace typically seen with integrated groups does not necessarily confer reduced fatigue levels for males. Indeed, only 14 male PI volunteers (44%) showed any indication of consistency of this sort, with a shift down (towards a lower lower-bound and/or a lower upper-bound) when in an integrated group. As such, it's likely that a number of factors beyond pace (e.g., intra-group dynamics, motivation, individual conditioning) might be influencing males' fatigue responses, post-hike, leading to these seemingly contradictory results.

N.4.6.1.2.2 Final Fatigue Scores (by Integration Level)

Post-trial fatigue results by Integration Level proved to be mixed. On Offense scenario trials, we only saw a significant difference in post-trial fatigue levels in males between High-Density and Low-Density squads—with a slight shift towards lower fatigue levels ("1-Fully Alert", "2-Very Lively")—but not between Control and either integration group

On Defense trials, post-trial responses for males were similar to what we found for post-7km Hike results: responses in High-Density Rifle Squads were significantly different (towards less fatigue) from both Control and Low-Density groups, with no difference found between Low-Density and Control groups.

Since the majority of individuals reported being as tired or less tired (69.5%) at the end of the Defense trial as they had after the 7km hike, it's likely whatever factors influenced post-hike fatigue levels—resulting in any differences (or lack of differences) between groups—were also at work for post-trial reporting.

Table 44 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) - PI

Distribution of PI Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	**Location/Day**	**IL**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
PI	29Palms/1 Post-7km Hike	C	248	12.1%	28.2%	23.0%	19.4%	11.7%	5.2%	0.4%
		LD	178	14.6%	25.3%	25.8%	19.1%	10.1%	4.5%	0.6%
		HD	138	15.9%	29.0%	31.2%	16.7%	5.1%	2.2%	0.0%
	29Palms/2 Post-"Offense"	C	246	8.1%	29.3%	26.0%	25.2%	8.5%	2.8%	0.0%
		LD	175	6.3%	25.1%	32.0%	21.1%	13.1%	2.3%	0.0%
		HD	138	9.4%	32.6%	27.5%	18.1%	11.6%	0.7%	0.0%
	29Palms/1 Post-"Defense"	C	249	6.4%	26.9%	23.3%	31.3%	11.6%	0.4%	0.0%
		LD	175	9.1%	22.9%	29.7%	20.6%	14.9%	2.3%	0.6%
		HD	136	9.6%	30.9%	27.2%	25.0%	6.6%	0.7%	0.0%

N.4.6.2 Workload Results

For the Provisional Infantry (PI) volunteers at Twentynine Palms, we have maximum workload self-reports from 32 males and eight females. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.6.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females, for both days in their run-cycle. Statistical results are summarized in the table below (Table 45)

Table 45 - Summary of Maximum Workload Scores (by Gender) – PI

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
PI	Max Workload for Trial	29Palms/2 “Offense”	M	566	4 / 4	62.4	<0.01*
			F	109	5 / 5		
		29Palms/1 “Defense”	M	568	4 / 4	32.6	<0.01*
			F	111	4 / 4		

We found a significant difference between female and male Provisional Infantry (PI) volunteers’ maximum workload scores for both days (Defense and Offense). Males most often reported scores between “3-Active but Easy” and “4-Challenging” (70.7% and 68.0% for Defense and Offense, respectively), whereas females most often reported scores of “4-Challenging” and “5-Extremely Busy” (84.4% and 78.4%) (Table 46).

Table 46 - Maximum Workload Score Distributions (by Gender) - PI

Distribution of PI Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
PI	29Palms/2 "Offense"	M	566	0.4%	8.3%	31.8%	38.9%	13.8%	6.2%	0.7%
		F	109	0.0%	0.0%	6.4%	39.4%	45.0%	9.2%	0.0%
	29Palms/1 "Defense"	M	568	0.4%	8.1%	24.6%	43.3%	14.8%	7.7%	1.1%
		F	111	0.0%	0.0%	8.1%	48.6%	29.7%	10.8%	2.7%

For both groups, we see a small, but visible, number of maximum workload scores at the highest range (i.e., "6-Overloaded" and "7-Unmanageable") indicating that a sub-set of both genders felt unable to keep up with their workload demands (five of eight females, and 14 out of 32 males)—at least on occasion. Indeed, eight Marines reported in this range at least four times (and several in the double-digits). This suggests a sizeable number of all PI Marines felt (at times) overwhelmed operating in a Rifle Squad.

The difference we see between the two groups is unlikely to be due to differences in experience, as none of the volunteers in the PI pool were 0311s. This may, however, be linked to the higher levels of fatigue reported by female PIs (nearly one-quarter to one-third of all female post-trial fatigue responses were "5-Moderately Tired" or higher—see Table 42), since even relatively low levels of fatigue can degrade concentration and performance of non-rote skills.

N.4.6.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males' responses remained consistent at different integration levels. For PI, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 2 females), and High-Density (i.e., 4-5 females) Rifle Squads. We have maximum workload self-reports from 32 males; statistical results are summarized in the table below (Table 47).

Table 47 - Summary of Males' Maximum Workload Scores (by Integration Level) - PI

MOS	Workload	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
PI	Max Workload for Trial	29Palms/2 "Offense"	C	248	4 / 4	1.3	0.52		
			LD	179	4 / 4				
			HD	139	4 / 4				
		29Palms/1 "Defense"	C	251	4 / 4	7.7	0.02*	(LD-C) 0.33	0.74
			LD	179	4 / 4			(HD-C) -2.4	0.02*
			HD	138	4 / 4			(HD-LD) -2.6	0.01*

We found mixed results when we looked at males' max workload responses by Integration Level. For Defense trials, we saw a significant difference between High-Density and Control or Low-Density groups. We did not see any difference for Offense trials.

From a visual inspection of the scores reported by the different groups (Table 48), we see that males in High-Density groups (on Defense days) have slightly fewer reports

(15.2%) at the higher end (i.e., "5-Extremely Busy" and higher) than Control and Low-Density groups (26.3%), along with a corresponding (slight) increase in scores at the lower end (39.1% for HD groups, but only 32.3% and 29.6% for Control and Low-Density groups, respectively). In other words, males in High-Density groups appear to be slightly more optimistic in their ability to manage their workload.

Table 48 - Males' Max Workload Score Distributions (by Integration Level) - PI

Distribution of PI Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	**Location/Day**	**IL**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
PI	29Palms/2 "Offense"	C	248	0.0%	7.7%	34.7%	35.9%	15.7%	4.8%	1.2%
		LD	179	1.1%	7.8%	27.9%	41.3%	12.3%	9.5%	0.0%
		HD	139	0.0%	10.1%	31.7%	41.0%	12.2%	4.3%	0.7%
	29Palms/1 "Defense"	C	251	0.4%	6.4%	25.5%	41.4%	16.7%	8.4%	1.2%
		LD	179	0.0%	7.8%	21.8%	44.1%	15.6%	8.9%	1.7%
		HD	138	0.7%	11.6%	26.8%	45.7%	10.1%	5.1%	0.0%

As for possible mechanisms, these results are unlikely to be driven by relative experience given that all volunteers—male and female—are not operating in their primary MOS. However, these mixed results are consistent with differences seen by Integration Level with fatigue above (see Table 43 and Table 44).

Males' showed significantly lower fatigue levels for Defense trials—but not Offense trials—when in a High-Density group, compared with Control and Low-Density responses. Since even relatively low levels of fatigue can degrade concentration and performance, the lower fatigue levels for High-Density groups on Defense trials could shift their perception of their ability to manage their workload towards lower scores.

N.4.6.3 Cohesion Results

For the Provisional Infantry (PI) volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) from 32 males and eight females. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—

then we conducted Wilcoxon pairwise comparisons (using p<0.033 to adjust for multiple comparisons).

N.4.6.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 49).

Table 49 - Summary of Unit Cohesion Scores (by Gender) - PI

MOS	Location	Cohesion	Gender	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
PI	29Palms	Q1: Group was united in trying to reach goals	M	556	8.14	1.16	0.27	0.60
			F	111	8.12	1.26		
		Q2: We all take responsibility for task success of the group	M	556	8.15	1.18	0.01	0.77
			F	111	8.04	1.31		
		Q3: Group members have similar aspirations for success	M	556	8.09	1.37	0.00	0.98
			F	111	8.08	1.29		
		Q4: If members had problems with a task, all wanted to help	M	556	8.17	1.22	2.28	0.13
			F	111	7.70	1.74		
		Q5: Members communicated freely about responsibilities	M	556	8.21	1.22	0.95	0.33
			F	111	8.07	1.26		
		Composite Score (sum of Q1-Q5)	M	556	40.77	1.22	1.18	0.28
			F	111	40.01	1.26		

We did not find a significant difference between gender for any question (Q1-Q5), nor for the composite score. Only "Q4-Help" is on the border for significance, likely due to a mini-peak among female responses at "4" (towards Significantly Disagree)—driven largely by a single female—but which is not seen for any other questions, nor among males.

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement (Figure 9). The average values—by gender—reflect the common perception (by statement and by composite score) of males and females regarding their units' relatively high cohesion. The one borderline question ("Q4-Help")—driven to a lower value due to a series of low responses by a single female Marine (i.e., ten "4" scores)—appears to be an outlier.

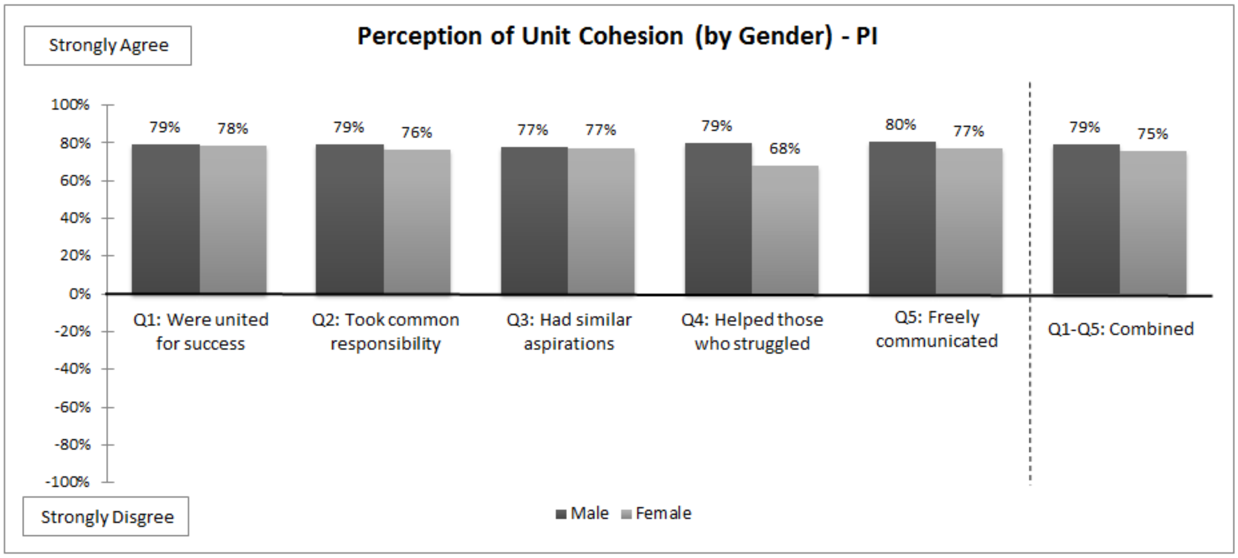


Figure 9 - Perception of Unit Cohesion (by Gender) – PI

N.4.6.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For Provisional Infantry (PI), we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 2 females), and High-Density (i.e., 4-5 females) Rifle Squads. We have unit cohesion survey results (five statements per survey) from 32 males; statistical results are summarized in the table below (Table 50).

Table 50 - Summary of Males' Unit Cohesion Scores (by Integration Level) - PI

MOS	Cohesion	IL	N	Mean	SD	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
PI	Q1: Group was united in trying to reach goals	C	248	8.33	0.88	8.67	0.01*	(LD-C) -1.96	0.05*
		LD	179	8.09	1.18			(HD-C) -2.80	<0.01*
		HD	139	7.88	1.47			(HD-LD) -0.94	0.35
	Q2: We all take responsibility for task success of the group	C	248	8.33	0.90	7.04	0.03*	(LD-C) -1.50	0.13
		LD	179	8.10	1.25			(HD-C) -2.61	0.01*
		HD	139	7.89	1.46			(HD-LD) -1.13	0.26
	Q3: Group members have similar aspirations for success	C	248	8.35	0.98	12.20	<0.01*	(LD-C) -2.68	<0.01*
		LD	179	7.92	1.60			(HD-C) -3.15	<0.01*
		HD	139	7.85	1.57			(HD-LD) -0.62	0.54
	Q4: If members had problems with a task, all wanted to help	C	248	8.39	0.86	11.66	<0.01*	(LD-C) -2.19	0.03*
		LD	179	8.09	1.28			(HD-C) -3.29	<0.01*
		HD	139	7.88	1.56			(HD-LD) -1.19	0.23
	Q5: Members communicated freely about responsibilities	C	248	8.40	0.91	9.20	0.01*	(LD-C) -2.52	0.01*
		LD	179	8.11	1.34			(HD-C) -2.58	0.01*
		HD	139	8.01	1.47			(HD-LD) -0.28	0.78
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	248	41.81	3.97	11.17	<0.01*	(LD-C) -2.12	0.03*
		LD	179	40.32	5.91			(HD-C) -3.24	<0.01*
		HD	139	39.50	7.08			(HD-LD) -1.15	0.25

We found a statistical difference between the Control group and both integration levels (i.e., High-Density, and Low-Density groups) for all five cohesion statements, as well as for the overall composite cohesion score. However, we did not see a difference between the two integration levels.

When we looked at the distributions of each question, we see very similar patterns—the majority of responses are heavily skewed to the far right (i.e., “9”, at Significantly Agree), and taper off relatively quickly to the left. Where the differences occur is in the sharpness of the main peak at “9” and the span of the taper. For Control groups, we see the sharpest peak (53% to 63% across all five statements) that largely tapers off by “7”, with virtually no responses beyond that.

For the two integrated groups, the peaks are a bit more shallow (42% to 50%), and the taper spans a wider (and more negative) set of responses; for the Low-Density group the taper ends around “6”, while it extends out (though at rather low values) to “5” for the High-Density group.

We also graphed the means for each statement as a percentage of how strongly males agreed (positive) or disagreed (negative) with each question’s statement while in each group (Figure 10). The average values are consistent with what we saw with regards to males’ relatively high perceptions of unit cohesion within the Control group, along with the easy decline in their perception of cohesion when in the Low- and High-Density groups.

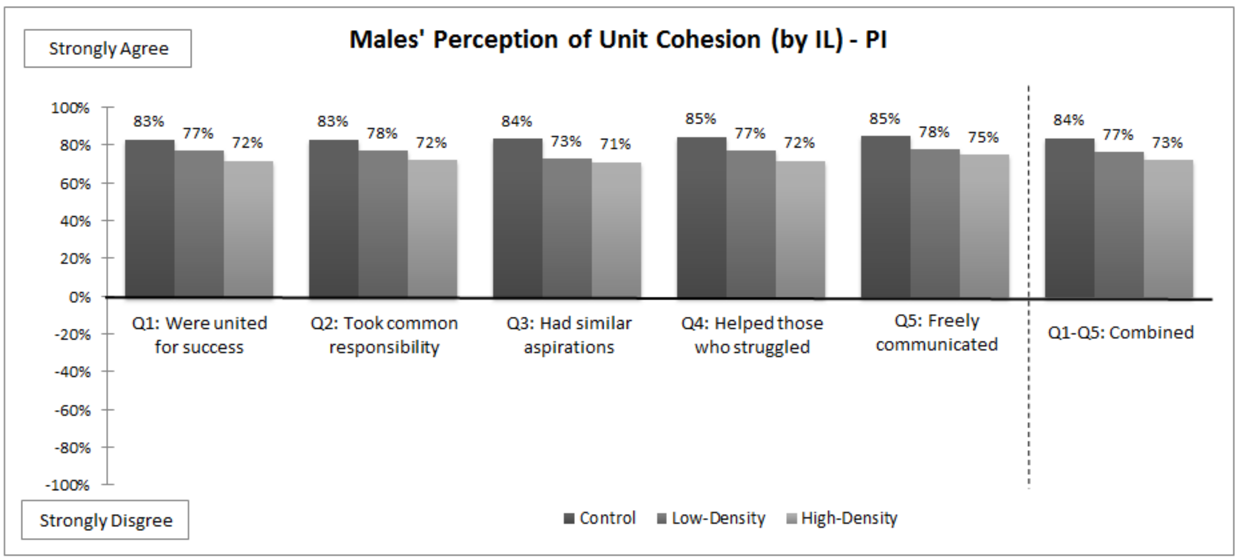


Figure 10 - Males' Perception of Unit Cohesion (by Integration Level) - PI

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group. With the larger population pool (compared to some of the other 03XX MOS analyzed), we can likely have more faith that these trends might well persist.

Differences in experience also appears to be a less likely driver, as all volunteers are operating outside of their PMOS, although we can’t rule it out completely as females have also operated under additional assignment limitations that might have minimized their exposure to the field environment. Incidentally, the average values for the integrated groups appear to align well with how females rated their perceptions of unit cohesiveness.

N.4.7 Provisional Infantry Machine Gunner (PIMG)

N.4.7.1 Fatigue Results

For the Provisional Infantry Machine Gunner (PIMG) volunteers at Twentynine Palms, we have baseline, mid-trial, and final fatigue self-reports from four males[16] and four females. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of p ≤ 0.10 for statistical significance (*).

N.4.7.1.1 Fatigue Results by Gender

We first looked at looking at how fatigue self-reports compared between males and females, including their baseline (pre-trial), their final (post-trial, for both days in the run-cycle), and post-7km Hike (mid-trial, Day 1) fatigue levels. Statistical results are summarized in the table below (Table 51).

Table 51 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) - PIMG

MOS	Fatigue	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
PIMG	Baseline (Pre-Trial)	29Palms/1,2	M	116	3 / 3	0.72	0.40
			F	124	3 / 3		
	Post-7km (Mid-Trial)	29Palms/1 “Defense”	M	55	3 / 3	13.4	<0.01*
			F	60	4 / 4		
	Final (Post-Trial)	29Palms/2 “Offense”	M	56	3 / 3	12.1	<0.01*
			F	59	4 / 4		
		29Palms/1 “Defense”	M	55	3 / 3	9.3	<0.01*
			F	59	4 / 4		

N.4.7.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We did not find a significant difference between female and male PIMG volunteers’ pre-trial fatigue scores. Both males and females were most likely to report fatigue levels in the “2-Very Lively” to “3-Okay” range (84.5% and 72.6%, respectively) prior to the start of a trial (Table 52).

We found very few of the higher fatigue scores for either sex (i.e., “5-Moderately Tired”, “6-Extremely Tired”, or “7-Exhausted”)—only 1.7 percent of male responses (two responses by different Marines), and 5.6 percent of female responses (driven almost entirely by a single Marine)—and no obvious pattern of increasing scores over time, suggesting volunteers largely had sufficient recovery time between trials.

[16] One male participated in only one run-cycle, and did not respond to the mid-trial or final surveys for Day 1.

N.4.7.1.1.2 Final Fatigue Scores (by Gender)
In looking at the volunteers' post-trial responses, we found significant differences in their responses—by gender—for both days of the PIMG run-cycle (i.e., Defense and Offense). For both days, males' responses clustered largely between "3-Okay" and "4-A Little Tired" (76.4% and 76.8%, for Defense and Offense days, respectively), whereas females' post-trial responses tended towards a more shallow and wider distribution between "2-Very Lively" and "5-Moderately Tired" (86.4% for both days).

N.4.7.1.1.3 Post-7km Hike Fatigue Scores (by Gender)
For the Defense scenario, PIMG volunteers also completed a fatigue survey following the completion of their 7km Hike under load. As with post-trial responses, we found a significant difference in male and female responses, with males most likely to report "3-Okay" (43.6%) and females most likely to report "4-A Little Tired" (41.7%).

However, given the "non-majority" of both of these peaks and the rather small number of contributing Marines, we also looked as the range of values each Marine reported—post-hike—over the course of the study. For the males, two of the three volunteers reported values ranging between "3-Okay" and "5-Moderately Tired"; the third had a wider range between "1-Fully Alert" and "5-Moderately Tired".

For the females, two reported between "2-Very Lively" and "6-Extremely Tired" while the other two reported between "3-Okay" and "Extremely Tired". Given the wide ranges within Marines for the same task and the small number of contributors, we should be extremely cautious in drawing conclusions about propensity toward fatigue for this MOS from these results.

Table 52 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) - PIMG

Distribution of PIMG Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
PIMG	29Palms/1&2 Pre-Trial	M	116	6.0%	32.8%	51.7%	7.8%	1.7%	0.0%	0.0%
		F	124	5.6%	35.5%	37.1%	16.1%	3.2%	1.6%	0.8%
	29Palms/1 Post-7km Hike	M	55	1.8%	10.9%	43.6%	20.0%	23.6%	0.0%	0.0%
		F	60	0.0%	5.0%	13.3%	41.7%	30.0%	10.0%	0.0%
	29Palms/2 Post-"Offense"	M	56	0.0%	10.7%	46.4%	30.4%	8.9%	3.6%	0.0%
		F	59	0.0%	5.1%	22.0%	37.3%	27.1%	8.5%	0.0%
	29Palms/1 Post-"Defense"	M	55	1.8%	12.7%	60.0%	16.4%	9.1%	0.0%	0.0%
		F	59	0.0%	13.6%	25.4%	39.0%	22.0%	0.0%	0.0%

N.4.7.1.2 Fatigue Results by Integration Level

For previous MOS, we looked at whether males' fatigue responses remained consistent at different integration levels. However, for PIMG, there was only a single valid run-cycle (with one unit) that had an integrated unit (i.e., one male and two females). All other run-cycles were either all-male or all-female.

N.4.7.2 Workload Results

For the PI Machine Gunner (PIMG) volunteers at Twentynine Palms, we have maximum workload self-reports from four males[17] and four females. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.7.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females, for both days in their run-cycle. Statistical results are summarized in the table below (Table 53).

Table 53 - Summary of Maximum Workload Scores (by Gender) – PIMG

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
PIMG	Max Workload for Trial	29Palms/2 "Offense"	M	57	4 / 4	23.1	<0.01*
			F	62	4 / 4		
		29Palms/1 "Defense"	M	56	4 / 4	16.7	<0.01*
			F	62	4.5 / 4		

We found a significant difference between female and male Provisional Infantry Machine Gunner (PIMG) volunteers' maximum workload scores for both days (Defense and Offense). Males most often reported scores between "3-Active but Easy" and "4-Challenging" (80.7% and 73.2% for Defense and Offense, respectively), whereas females most often reported scores in the slightly higher range of "4-Challenging" and "5-Extremely Busy" (87.1% and 80.6%) (Table 54).

We also saw an increased likelihood of maximum workload scores of "6-Overloaded" or "7-Unmanageable" among females; all four female volunteers reported in this range at least once, but total numbers in this range were driven by one Marine. However, even among the males, two of the four male volunteers had also reported in this range at least once.

[17] One male only participated in two trials.

While relative experience is unlikely to be the driver in the difference we found between genders in maximum workload responses, there may be a tie to fatigue—which also tended to be higher, post-trial, among female PIMG volunteers.

However, given the small size of this volunteer pool (four males—of which one only participated for one run-cycle—and four females) and the signs that most Marines may have struggled—on occasion—to accomplish the necessary workload of a Machine Gunner, we should be cautious in extrapolating the PIMG study results to the larger Marine Corps community. At the very least, the extent to which non-0331 Marines might be able to function at the same workload demands as an 0331 Machine Gunner warrants further study.

Table 54 - Maximum Workload Score Distributions (by Gender) - PIMG

Distribution of PIMG Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
PIMG	29Palms/2 "Offense"	M	57	0.0%	8.8%	22.8%	57.9%	8.8%	1.8%	0.0%
		F	62	0.0%	0.0%	4.8%	51.6%	35.5%	8.1%	0.0%
	29Palms/1 "Defense"	M	56	1.8%	5.4%	25.0%	48.2%	16.1%	3.6%	0.0%
		F	62	0.0%	0.0%	8.1%	41.9%	38.7%	9.7%	1.6%

N.4.7.2.2 Workload Results by Integration Level

For previous MOS, we looked at whether males' fatigue responses remained consistent at different integration levels. However, for PIMG, there was only a single valid run-cycle (with one unit) that had an integrated unit (i.e., one male and two females). All other run-cycles were either all-male or all-female.

N.4.7.3 Cohesion Results

For the Provisional Infantry Machine Gunner (PIMG) volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) from four males[18] and four females. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions assumption for parametric data, we restricted our statistical analysis to the

[18] One male only participated in one run-cycle.

non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of p ≤ 0.10 for statistical significance (*).

N.4.7.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five statements, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 55).

Table 55 - Summary of Unit Cohesion Scores (by Gender) - PIMG

MOS	Cohesion	Gender	N	Mean	SD	χ2 Statistic	χ2 p-Value
PIMG	Q1: Group was united in trying to reach goals	M	57	7.77	0.91	7.21	<0.01*
		F	61	8.18	0.90		
	Q2: We all take responsibility for task success of the group	M	57	7.81	0.77	6.95	<0.01*
		F	61	8.16	0.88		
	Q3: Group members have similar aspirations for success	M	57	7.65	1.09	9.58	<0.01*
		F	61	8.20	0.93		
	Q4: If members had problems with a task, all wanted to help	M	57	7.91	0.79	3.51	0.06*
		F	61	8.13	0.99		
	Q5: Members communicated freely about responsibilities	M	57	7.81	0.90	5.91	0.02*
		F	61	8.15	1.03		
	Composite Score (sum of Q1-Q5)	M	57	38.95	3.78	7.75	<0.01*
		F	61	4.82	4.53		

We found a significant difference between PIMG males and females for all five cohesion statements ("Q1-Unity", "Q2-Responsibility", "Q3-Aspirations", "Q4-Help" and "Q5-Comm"), as well as for the overall composite cohesion score. When we look more closely at the distribution of responses for each statement, we see that females tended to be a bit more optimistic (i.e., higher scores) than males, though both groups have the majority of their responses in the top three levels (towards strongly agree).

For females, we see much more representation at the highest level, "9" (41.0% to 45.9%), than we do among male responses (15.8% to 21.1%). However, it seems that this difference is relatively minor; when we look at the total contributions within the top three levels (i.e., "7" to "9")—both males and females have more than 86% of all responses in this range with almost no responses below "5" for either group. This suggests that both males and females perceived their groups to possess fairly high cohesiveness.

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement

(Figure 11). The average values—by gender—reflect the differences we saw in the distribution of unit cohesion scores—females tended towards more optimistic scores than males, but both are relatively solid towards positive cohesion.

It is unclear why males might have lower perceptions of unit cohesion than females. This MOS did not have integrated groups per se—there being only a single "integrated" unit (i.e., one male and two females); all other units were either all-male (i.e., Control) or all-female. The slightly higher scores in the all-female groups may be a function of a common level of experience or a greater degree of cooperation.

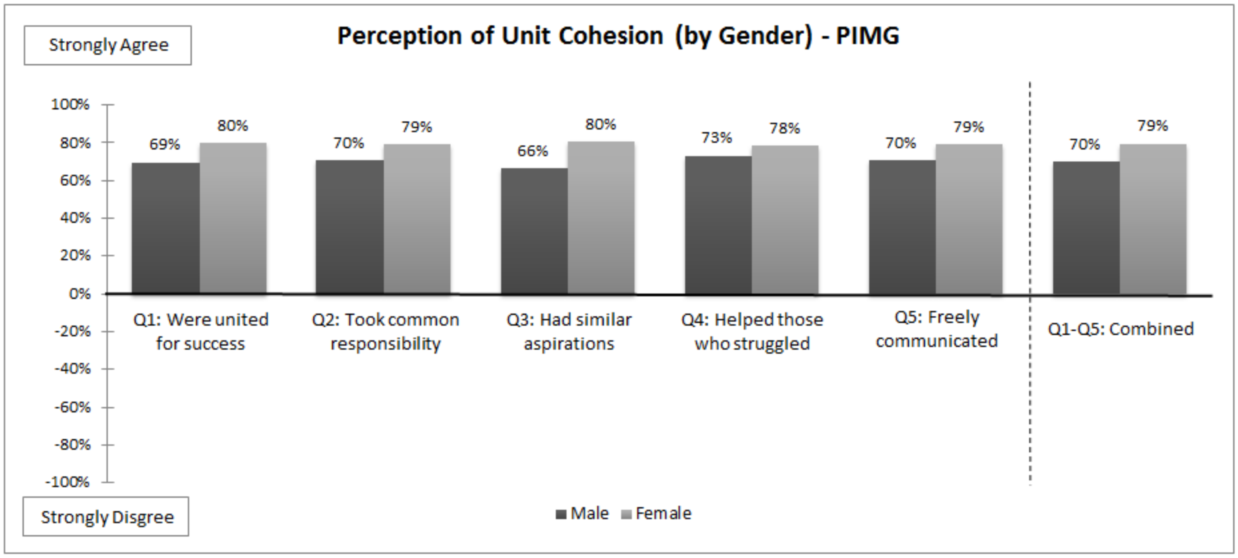


Figure 11 - Perception of Unit Cohesion (by Gender) - PIMG

Of course, we can only guess at the relative level of experience among the all-male groups. That some struggled[19] (on occasion) suggests that there may have been some differences. However, some of the females struggled, too, so experience can't be the entire answer—or at least it mattered less among females. But, given the small size of this volunteer pool (four males—of which one only participated for one run-cycle—and four females) we should be cautious in extrapolating the PIMG study results to the larger Marine Corps community.

N.4.7.3.2 Cohesion Results by Integration Level

For previous MOS, we looked at whether males' cohesion responses remained consistent at different integration levels. However, for PIMG, there was only a single valid run-cycle (with one unit) that had an integrated unit (i.e., one male and two females). All other run-cycles were either all-male or all-female.

[19] Based upon two of the four males having reported—on occasion—max workload scores of "6-Overloaded".

N.4.8 Combat Engineer (MOS 1371)

N.4.8.1 Fatigue Results

For the 1371 volunteers at Twentynine Palms, we have baseline, mid-trial, and final fatigue self-reports from 18 males and eight females[20]. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p < 0.033$ to adjust for multiple comparisons).

N.4.8.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial), final (post-trial, for both days in the run-cycle), and post-7km Hike (mid-trial, Defense) fatigue levels. Statistical results are summarized in the table below (Table 56).

Table 56 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) - 1371

MOS	Fatigue	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
1371	Baseline (Pre-Trial)	29Palms/1,2	M	576	2 / 3	0.53	0.47
			F	150	2 / 2		
	Post-7km (Mid-Trial)	29Palms/2 “Defense”	M	285	3 / 3	7.0	<0.01*
			F	73	4 / 4		
	Final (Post-Trial)	29Palms/1 “Offense”	M	286	3 / 2	0.05	0.82
			F	73	3 / 2		
		29Palms/2 “Defense”	M	286	3 / 2	0.03	0.86
			F	75	3 / 3		

N.4.8.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We did not find a significant difference between female and male 1371 volunteers’ pre-trial fatigue scores. Both groups showed the highest clustering of pre-trial fatigue responses at “2-Very Lively” and “3-Okay” (Table 57).

We found very few of the higher fatigue scores for either gender (i.e., “5-Moderately Tired”, “6-Extremely Tired”, or “7-Exhausted”)—only 7.3 percent of male responses (driven overwhelmingly by two Marines), and 2.0 percent of female responses (three

[20] One female only participated in a single run-cycle.

reports by two Marines)—and no obvious pattern of increasing scores over time, suggesting volunteers largely had sufficient recovery time between trials.

N.4.8.1.1.2 Final Fatigue Scores (by Gender)

In looking at the volunteers' post-trial responses, we did not find a significant difference for either day. Similar to pre-trial distributions, both males and females were most likely to report fatigue levels of either “2-Very Lively” or “3-Okay”, although we did see a respectable representation for both groups at “4-A Little Tired (21% for females, and 12-16% for males) not see in pre-trial scores.

For the small number of post-trial responses we saw in the higher ranges (“5-Moderately Tired” or higher), most contributions by individual were sparse (i.e., one to four responses over the course of the study). However, the two male Marines who contributed substantially more fatigue scores in this range (seven and 21 reports) had also been the major contributors to pre-trial fatigue levels in this range.

N.4.8.1.1.3 Post-7km Hike Fatigue Scores (by Gender)

The one area where we did see significant difference between the genders for 1371s was the post-7km Hike in the Defense scenario. Following the hike under load, females showed a slight shift in fatigue levels (towards more fatigue)—more often clustering around “3-Okay” and “4-A Little Tired” (60.3%)—than males, whose responses more often fell around “2-Very Lively” and “3-Okay” (57.9%). In both groups, however, we did see relatively shallow distributions covering the “2-Very Lively” to “5-Moderately Tired” span (89.8% and 91.8%, for males and females, respectively).

When we looked at the responses within individual Marines, we saw that most Marines' ranges (84%)—for both males and females—fell across all or most of this span (i.e., three or more levels), suggesting a great degree of intra-subject variability within the 1371 volunteer pool.

Table 57 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) - 1371

Distribution of 1371 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
1371	29Palms/1&2 Pre-Trial	M	576	17.2%	33.0%	34.0%	8.5%	5.2%	1.7%	0.3%
		F	150	3.3%	56.0%	33.3%	5.3%	2.0%	0.0%	0.0%
	29Palms/2 Post-"7km Hike"	M	285	7.7%	24.6%	33.3%	21.1%	10.9%	1.1%	1.4%
		F	73	2.7%	17.8%	28.8%	31.5%	13.7%	4.1%	1.4%
	29Palms/1 Post-"Live Fire"	M	286	8.0%	37.4%	29.4%	15.7%	5.6%	2.4%	1.4%
		F	73	8.2%	35.6%	32.9%	20.5%	2.7%	0.0%	0.0%
	29Palms/3 Post-"Non-Live Fire"	M	286	8.4%	35.0%	33.9%	12.2%	8.4%	1.7%	0.3%
		F	75	8.0%	32.0%	36.0%	21.3%	2.7%	0.0%	0.0%

N.4.8.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 1371s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 2 females), and High-Density (i.e., 4 females) Engineer Squads. We have mid-trial and final fatigue self-reports from 18 males for the Control and Low-Density units, and from 15 males in the High-Density units; statistical results are summarized in the table below (Table 58).

Table 58 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) - 1371

MOS	Fatigue	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
1371	Post-7km (Mid-Trial)	29Palms/2 "Defense"	C	154	3 / 3	8.6	0.01*	(LD-C) -0.21	0.84
			LD	83	3 / 3			(HD-C) -2.9	<0.01*
			HD	48	3 / 3			(HD-LD) -2.4	0.02*
	Final (Post-Trial)	29Palms/1 "Offense"	C	110	2 / 2	1.3	0.52		
			LD	84	2 / 2				
			HD	44	2 / 1				
		29Palms/2 "Defense"	C	114	2 / 2	1.0	0.59		
			LD	84	2 / 2				
			HD	46	2 / 2				

N.4.8.1.2.1 Post-7km Hike Fatigue Scores (by Integration Level)
Following the 7km Hike under load (Defense scenario), we found the distribution of males' fatigue responses in High-Density squads to be significantly different (towards less fatigue) from both Control and Low-Density squads. We did not see a statistical difference between Control and Low-Density groups.

Looking more closely at the distributions of the various groups (Table 59), we saw a shift to the left (i.e., less fatigue) for males in the High-Density groups—clustering more strongly around "2-Very Lively" and "3-Okay" (72.9%) than those same levels for Control (56.5%) and Low-Density (51.8%) squads, which tended towards broader (and more shallow) distributions spanning "2-Very Lively" to "4-A Little Tired". Control and Low-Density Groups also had higher (though still relatively small) representations in the higher range ("5-Moderately Tired" to "7-Exhausted")—with 15.6% and 14.5%, respectively—compared to the High-Density group (4.2%).

One potential reason for this relatively minor shift towards less fatigue in High-Density groups could be their (typically) slower hike times (see Annex K). However, given that we also saw a relatively wide degree of intra-subject variability across all Marines' fatigue responses for the 7km Hike, we also looked to see if this variability changed by integration level.

For the 15 males that contributed fatigue scores for a High-Density squad, most (73%) showed both a narrowing and depression in the range of fatigue values they reported compared with their ranges while in a Control or even Low-Density squad (with some indications of this pattern emerging in Low-Density groups as well, for some Marines). This suggests that a number of males were persistently—though maybe only slightly—less fatigued while in a High-Density squad, potentially because the slower pace was less physically demanding.

Another potential influence might be group dynamics. Males in all-male groups are more likely to report scores at the highest levels (i.e., "6-Extremely Tired", "7-Exhausted"), female responses in this range (see Table 57) are from Low-Density groups only. Well not definitive, we might also be seeing some group dynamic in terms of how hard all-male versus integrated units will press. All-male groups may be more willing to push a faster pace (even if it heavily fatigues some members) than an integrated group might.

N.4.8.1.2.2 Final Fatigue Scores (by Integration Level)
For 1371s, we did not see any significant difference in post-trial fatigue results based on Integration Level for either of the two trial scenarios (i.e., Day 1/Offense and Day 2/Defense). Across all integration levels and locations, we see males' responses largely clustering around "2-Very Lively" and "3-Okay" (ranging between 59.0% and 77.1%), with smaller percentages tapering off to either side (Table 59).

Table 59 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) - 1371

Distribution of 1371 Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
1371	29Palms/1 Post-7km Hike	C	154	5.8%	23.4%	33.1%	22.1%	12.3%	1.9%	1.3%
		LD	83	9.6%	20.5%	31.3%	24.1%	12.0%	0.0%	2.4%
		HD	48	10.4%	35.4%	37.5%	12.5%	4.2%	0.0%	0.0%
	29Palms/1 Post-"Offense"	C	160	6.3%	36.9%	29.4%	18.8%	5.0%	2.5%	1.3%
		LD	78	10.3%	35.9%	28.2%	11.5%	9.0%	3.8%	1.3%
		HD	48	10.4%	41.7%	31.3%	12.5%	2.1%	0.0%	2.1%
	29Palms/2 Post-"Defense"	C	155	6.5%	36.8%	34.8%	12.3%	6.5%	2.6%	0.6%
		LD	83	10.8%	26.5%	32.5%	14.5%	14.5%	1.2%	0.0%
		HD	48	10.4%	43.8%	33.3%	8.3%	4.2%	0.0%	0.0%

N.4.8.2 Workload Results

For the 1371 volunteers at Twentynine Palms, we have max workload self-reports from 18 males and eight females[21]. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.8.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females, for both days in their run-cycle. Statistical results are summarized in the table below (Table 60).

[21] One female only participated in a single run-cycle.

Table 60 - Summary of Maximum Workload Scores (by Gender) – 1371

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
1371	Max Workload for Trial	29Palms/1 "Offense"	M	275	3 / 3	5.9	0.02*
			F	72	4 / 4		
		29Palms/2 "Defense"	M	287	4 / 4	10.5	<0.01*
			F	75	4 / 4		

We found a significant difference between female and male 1371 volunteers' maximum workload scores for both days (Offense and Defense), with males showing slightly higher representation at the lower workload levels—i.e., "3-Active but Easy" and below (Table 61).

Table 61 - Maximum Workload Score Distributions (by Gender) - 1371

Distribution of 1371 Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
1371	29Palms/1 "Offense"	M	275	1.8%	16.7%	34.9%	29.8%	12.7%	4.0%	0.0%
		F	72	1.4%	2.8%	30.6%	51.4%	13.9%	0.0%	0.0%
	29Palms/2 "Defense"	M	287	1.4%	11.1%	32.4%	35.2%	15.0%	3.8%	1.0%
		F	75	0.0%	0.0%	28.0%	40.0%	26.7%	5.3%	0.0%

However, we also noticed the presence of some "1-No Demands" scores, for both genders, as the maximum workload a Marine experienced during a trial. This seems highly unlikely, given the design of the 1371 Scheme of Maneuver. When we looked more closely at the contributors of these scores, we see that they represent one female response and three male contributors—one reporting in this range seven times (for all of the last-half of his trials). This suggests the possibility of that some type of "survey fatigue" or reporting bias may have been introduced, though exclusion of these trials does not change the overall statistical result by gender—i.e., a statistical difference exists between males and females for max workload reports.

As for this difference between the two genders, it's unlikely that this can be explained through differences in fatigue levels (which we didn't find post-trial). Plus, since all 1371 volunteers were operating in their primary MOS, familiarity with the tasks performed shouldn't differ.

However, it is possible that the historical assignment restrictions on female 1371s could have introduced an experience bias in our population, especially in field-related tasks.

Unfortunately, without knowing what tasks elicited each peak workload report, we cannot estimate how likely the latter theory might be.

N.4.8.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males' responses remained consistent at different integration levels. For 1371s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 2 females), and High-Density (i.e., 4 females) Engineer Squads. We have max workload self-reports from 18 males (18 in the Control and Low-Density groups, and 15 in the High-Density group); statistical results are summarized in the table below (Table 62).

Table 62 - Summary of Males' Max Workload Scores (by Integration Level) – 1371

MOS	Workload	Location / Day	IL	N (scores)	Median / Mode	χ^2 Statistic	χ^2 p-Value	Z-Test Statistic	Z-Test p-Value
1371	Max Workload for Trial	29Palms/1 "Offense"	C	149	3 / 3	2.9	0.23		
			LD	78	3 / 3				
			HD	48	3 / 3				
		29Palms/2 "Defense"	C	156	4 / 4	5.2	0.08*	(LD-C) -0.31	0.59
			LD	83	4 / 4			(HD-C) -1.8	0.02*
			HD	48	3 / 3			(HD-LD) -1.6	0.10

We found mixed results when we looked at males' maximum workload responses by Integration Level. For Defense trials, we saw a significant difference between High-Density and Control groups, but not between any other pairing—although it's borderline between High- and Low-Density groups. We also did not see any differences for Offense trials.

When we look at the scores reported by the different groups (Table 63), we see that males in High-Density groups (on Defense days) have fewer reports (10.4%) at the higher end (i.e., "5-Extremely Busy" and higher) than Control groups (22.4%), along with a corresponding increase in scores (58.3%) at the lower end (i.e., "3-Active but Easy" and lower) compared to Control (41.0%). Low-Density scores fall somewhere in between the other two groups—hence the lack of significance.

Table 63 - Males' Maximum Workload Score Distributions (by Integration Level) - 1371

Distribution of 1371 Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
1371	29Palms/1 "Offense"	C	149	1.3%	14.1%	34.9%	32.9%	14.1%	2.7%	0.0%
		LD	78	1.3%	20.5%	30.8%	26.9%	15.4%	5.1%	0.0%
		HD	48	4.2%	18.8%	41.7%	25.0%	4.2%	6.3%	0.0%
	29Palms/2 "Defense"	C	156	0.6%	10.3%	30.1%	36.5%	17.9%	3.8%	0.6%
		LD	83	1.2%	12.0%	31.3%	34.9%	14.5%	3.6%	2.4%
		HD	48	4.2%	12.5%	41.7%	31.3%	6.3%	4.2%	0.0%

As for possible mechanisms, these results have some factors in common with what we saw with fatigue surveys above, following the 7km Hike on Day Defense—except post-hike fatigues scores differed between both integration groups and Control, and not just between High-Density and Control groups. Plus, the differences in fatigue seen post-hike did not appear to extend to post-trial fatigue scores.

Without knowing what tasks in the Defense scenario might have been driving males' maximum workload scores, it's not possible to know if the decrease in fatigue from the 7km Hike could really have influenced males' to report lower maximum workload scores when in High Density groups. Fatigue is a plausible driver, since even relatively low levels of fatigue can degrade concentration and performance, but results are too mixed—and we have too little fidelity on why males chose the scores they did.

Alternatively, it may be that the High-Density group encouraged greater collaboration (i.e., spreading the "load") than the Control group, though this doesn't explain why we see an effect for Defense, and not for Offense, trials.

N.4.8.3 Cohesion Results

For the 1371 volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) from 18 males and eight females[22]. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-

[22] One female only participated in a single run-cycle.

parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p < 0.033$ to adjust for multiple comparisons).

N.4.8.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 64).

Table 64 - Summary of Unit Cohesion Scores (by Gender) - 1371

MOS	Cohesion	Gender	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
1371	Q1: Group was united in trying to reach goals	M	287	8.21	1.33	2.25	0.13
		F	75	7.80	2.01		
	Q2: We all take responsibility for task success of the group	M	287	8.16	1.43	4.72	0.03*
		F	75	7.79	1.82		
	Q3: Group members have similar aspirations for success	M	287	8.05	1.54	1.12	0.29
		F	75	7.72	2.05		
	Q4: If members had problems with a task, all wanted to help	M	287	8.08	1.48	1.58	0.21
		F	75	7.60	2.22		
	Q5: Members communicated freely about responsibilities	M	287	8.18	1.43	0.45	0.50
		F	75	7.88	1.92		
	Composite Score (sum of Q1-Q5)	M	287	40.68	6.83	1.12	0.29
		F	75	38.79	9.63		

We found significant differences between genders for only one the five cohesion statements ("Q2-Responsibility"), but not for the others or the overall composite cohesion score. When we look more closely at the distributions of each statement, most (for both genders) display a relatively sharp peak at "9" that tapers off over "7" and "8", with only a couple (two to four) responses below "5".

For "Q2-Responsibility", however, we see a very slight variation—the peak is split almost equally between "8" and "9". Since the non-parametric Kruskal-Wallis looks at rank order, this is likely sufficient to drive the statistical difference we found. However, it's hard to see how this represents a practical difference in overall perception of unit cohesion, nor what group dynamic might lead females to report a slightly lower score for just one statement (and this one in particular) out of five.

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement

(Figure 12). While the average values appear to show females with slightly lower unit cohesion scores than males—this would be overstating our results. Indeed, based upon the lack of significance for most questions—and the similarity of distributions we saw between the one statement that did show a significant difference ("Q2-Responsibility")—we might expect male and female averages to converge with the addition of more trials.

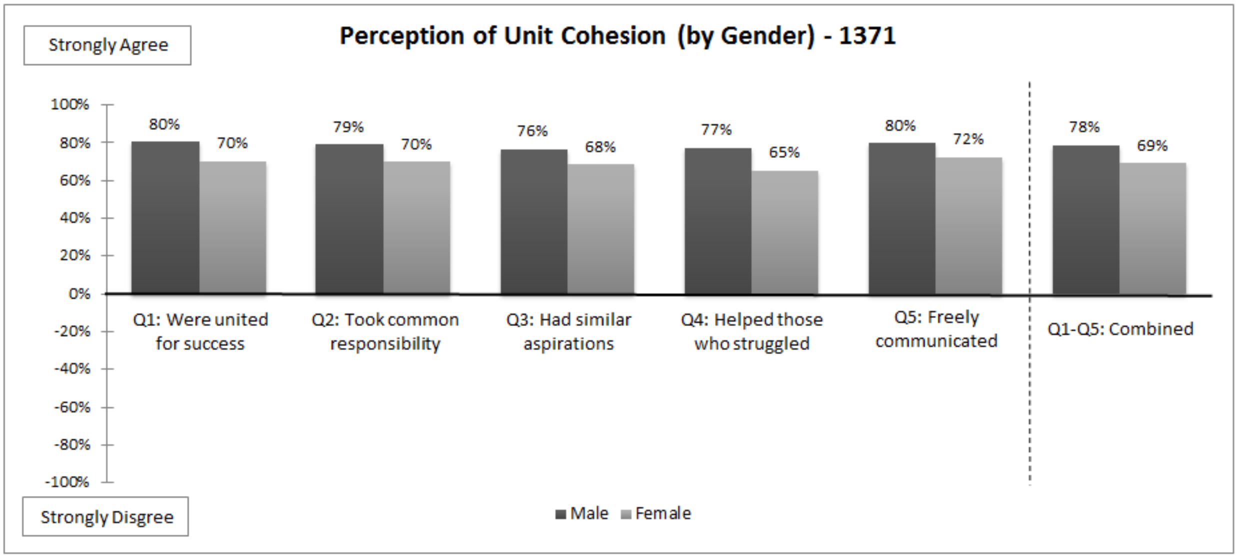


Figure 12 - Perception of Unit Cohesion (by Gender) – 1371

N.4.8.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For 1371s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 2 females), and High-Density (i.e., 4 females) Engineer Squads. We have unit cohesion survey results (five statements per survey) from 18 males (18 in the Control and Low-Density groups, and 15 in the High-Density group); statistical results are summarized in the table below (Table 65).

Table 65 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 1371

MOS	Cohesion	IL	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
1371	Q1: Group was united in trying to reach goals	C	158	8.53	0.89	16.55	<0.01*	(LD-C) -3.24	<0.01*
		LD	82	7.90	1.45			(HD-C) -3.41	<0.01*
		HD	47	7.68	1.96			(HD-LD) -0.41	0.68
	Q2: We all take responsibility for task success of the group	C	158	8.47	0.94	11.26	<0.01*	(LD-C) -2.73	<0.01*
		LD	82	7.93	1.43			(HD-C) -2.73	<0.01*
		HD	47	7.53	2.32			(HD-LD) -0.33	0.74
	Q3: Group members have similar aspirations for success	C	158	8.39	1.05	14.08	<0.01*	(LD-C) -3.24	<0.01*
		LD	82	7.73	1.60			(HD-C) -2.85	<0.01*
		HD	47	7.47	2.37			(HD-LD) -.10	0.92
	Q4: If members had problems with a task, all wanted to help	C	158	8.42	0.98	15.35	<0.01*	(LD-C) -3.00	<0.01*
		LD	82	7.80	1.50			(HD-C) -3.34	<0.01*
		HD	47	7.40	2.31			(HD-LD) -0.71	0.48
	Q5: Members communicated freely about responsibilities	C	158	8.48	0.97	11.82	<0.01*	(LD-C) -2.74	<0.01*
		LD	82	7.96	1.34			(HD-C) -2.85	<0.01*
		HD	47	7.53	2.37			(HD-LD) -0.43	0.67
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	158	42.30	4.51	14.01	<0.01*	(LD-C) -3.29	<0.01*
		LD	82	39.33	6.62			(HD-C) -2.81	<0.01*
		HD	47	37.62	11.06			(HD-LD) -0.08	0.94

We found a statistical difference between the Control group and both integration levels (i.e., High-Density, and Low-Density groups) for all five cohesion statements, as well as for the overall composite cohesion score. However, we did not see a difference between the two integration levels.

When we looked at the distributions of each question, we see very similar patterns—the majority of responses are heavily skewed to the far right (i.e., "9", at Significantly Agree), and taper off relatively quickly to the left. Where the differences occur is in the sharpness of the main peak at "9" and the span of the taper. For Control groups, we see the sharpest peak (59% to 65% across all five questions) that tapers off, predominantly, by “7” with virtually no responses beyond that.

For the two integrated groups, the peaks are more shallow (39% to 51%), and the taper spans a wider (and more negative) set of responses; for the Low-Density group the taper ends around “6”, while it extends out (though at rather low values) to “4” for the High-Density group (along with a few sporadic reports at “2” and “1”.

We also graphed the means for each statement as a percentage of how strongly males agreed (positive) or disagreed (negative) with each question’s statement while in each group (Figure 13). The average values (by integration level) shown are consistent with what we saw with regards to males’ relatively high perceptions of unit cohesion within the Control group, along with the easy decline in their perception of cohesion when in the Low- and High-Density groups.

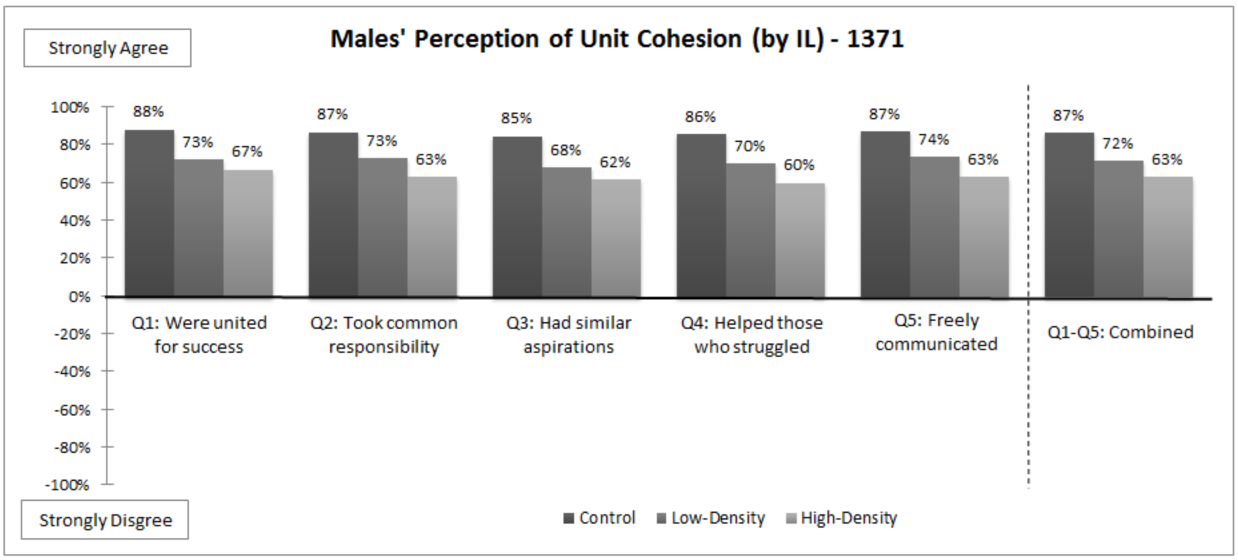


Figure 13 - Males' Perception of Unit Cohesion (by Integration Level) - 1371

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group. Differences in experience also appear to be a less likely, as all volunteers were operating within their PMOS, but we can’t rule it out completely as females 1371s have also operated under additional assignment limitations that might have minimized their exposure to the field environment. Incidentally, males’ average values for the integrated groups appear to align well with how females rated their perceptions of unit cohesiveness.

N.4.9 Mountaineering assessment (closed MOSs only)

N.4.9.1 Fatigue Results

For the Closed MOS (03XX) volunteers at Bridgeport, we have baseline, mid-trial, and final fatigue self-reports from 43 males and 19 females. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.9.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial), their post-5km Hike (mid-trial), and final (post-trial) fatigue levels. Results are summarized in the table below (Table 66).

Table 66 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) – Closed MOS

MOS	Fatigue	Location / Day	Gender	N	Median / Mode	χ2 Statistic	χ2 p-Value
03XX	Baseline (Pre-Trial)	Bridgeport/1 "Mountain"	M	222	2 / 2	18.3	<0.01*
			F	78	3 / 3		
	Post-5km (Mid-Trial)	Bridgeport/1 "Mountain"	M	214	2 / 3	13.4	<0.01*
			F	72	4 / 3		
	Final (Post-Trial)	Bridgeport/1 "Mountain"	M	220	2.5 / 2	15.5	<0.01*
			F	78	3 / 3		

N.4.9.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We found a significant difference between female and male 03XX volunteers' pre-trial fatigue scores at Bridgeport. Males' responses typically clustered around "1-Fully Alert" and "2-Very Lively" (75.7%), whereas females had smaller numbers—though still a majority—in this range (56.4%), with a more substantial number at "3-Okay" (37.2%) compared to males (17.6%).

However, we found only three reports in the higher fatigue scores (i.e., "5-Moderately Tired", "6-Extremely Tired", or "7-Exhausted") for males (1.4%) and none among the females' responses. Given the short duration of this phase of the study, and the relatively small number of trials per individual, it's not possible to determine whether there might be a pattern over time. However, with most pre-trial scores falling at "4-A Little Tired" or below, we believe most volunteers had sufficient recovery time between trials.

N.4.9.1.1.2 Final Fatigue Scores (by Gender)

In looking at the volunteers' post-trial responses at Bridgeport, we found significant differences in the distribution of scores by gender. Males showed a majority—and nearly flat distribution—across the first three levels (75.2%) before tapering sharply from "4-A Little Tired" to "7-Exhausted" (8.4%) (Table 67). Females showed a very similar pattern, but shifted to the right (towards slightly higher fatigue) by one level (i.e., "2-Very Lively" to "4-A Little Tired", accounting for 69.4%), with a sharp (and shorter) taper from "5-Moderately Tired" to "6-Extremely Tired" (22.2%).

In looking more closely at the higher range scores for both genders (i.e., "5-Moderately Tired", "6-Extremely Tired", and "7-Exhausted"), we see that these scores represent only a relatively small sub-set of each group (i.e., seven of 19 females and eight of 43 males), with most contributions to this range being only one or two reports per Marine; only two females and four males had more than two reports each.

However, as each Marine—male or female—only participated in a very few number of trials at Bridgeport (an average of four trials for females, and five for males), and only small sub-populations ever reported a post-trial fatigue greater than "4-A Little Tired", we should be cautious in trusting the robustness of this difference between males and females.

N.4.9.1.1.3 Post-5km Hike Fatigue Scores (by Gender)

The 03XX volunteers also completed a fatigue survey following the completion of their 5km Hike under load at Bridgeport; we found a significant difference in the distributions of male and female responses for this as well. Most post-5km Hike scores for males were at "3-Okay" and below (81.8%), whereas females had a slightly flatter profile that spanned a wider range, from "2-Very Lively" to "5-Moderately Tired" (88.9%).

As with the post-trial results, when we looked more closely at who contributed to the scores for both groups at "5-Moderately Tired" and above, only five of 19 females contributed (but only two—the same two who had reported multiple scores in this range after the 5km Hike—did so more than once).

For the males, we saw 17 out of 43 volunteers report at least once in this higher range, but only three had done so more than once. As such—with so few trials and a handful of Marines that reported in this range more than once—it's difficult to say whether the difference between males and females in this range is persistent and meaningful.

Table 67 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) – 03XX

Distribution of 03XX Fatigue Results (by Gender) Bridgeport				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
03XX	Bridgeport/1	M	222	40.1%	35.6%	17.6%	5.4%	0.9%	0.0%	0.5%
		F	78	11.5%	44.9%	37.2%	6.4%	0.0%	0.0%	0.0%
	Bridgeport/1 Post-"5km Hike"	M	214	22.9%	29.0%	29.9%	7.9%	7.5%	1.9%	0.9%
		F	72	9.7%	23.6%	23.6%	27.8%	11.1%	2.8%	1.4%
	Bridgeport/1 Post-Trial	M	214	22.9%	26.2%	26.2%	16.4%	6.5%	1.4%	0.5%
		F	72	8.3%	20.8%	23.6%	25.0%	19.4%	2.8%	0.0%

N.4.9.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For the 03XXs at Bridgeport, we compared males' responses in Control (i.e., all-male) versus High-Density (i.e., 6 females) Rifle Squads. We have mid-trial and final fatigue self-reports from 43 males in total—42 in the Control group and 36 in the High-Density group; statistical results are summarized in the table below (Table 68).

Table 68 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) – 03XX

MOS	Fatigue	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value
03XX	Post-5km (Mid-Trial)	Bridgeport/1 "Mountain"	C	143	3 / 3	7.3	<0.01*
			HD	71	2 / 1		
	Final (Post-Trial)	Bridgeport/1 "Mountain"	C	142	3 / 3	12.7	<0.01*
			HD	72	2 / 1		

N.4.9.1.2.1 Post-5km Hike Fatigue Scores (by Integration Level)

Following the 5km Hike, we found the distribution of males' fatigue responses in High-Density Squads to be significantly different from those in Control Squads (towards less fatigue). In both groups, we saw relatively shallow distributions, predominantly clustered in the lower range ("3-Okay" and below). However, for the High-Density group, we saw a heavier representation in the first two levels (62.0%) compared to the Control group (46.9%), suggesting males were more likely to be less tired, post-hike, in the integrated group.

For the higher range ("5-Moderately Tired", "6-Extremely Tired", and "7-Exhausted"), we also see more scores—though still relatively low numbers—in the Control group

(14.7%) than in the High-Density group (1.4%). In the Control group, this percentage represents mostly one-off responses by a significant number of the male volunteers (17 out of 42); only three actually responded more than once in this range, while the High-Density results are the result of a single response.

Given the small numbers of both contributors and trials in this phase of the study, we cannot say at this time whether the higher-end scores represent a meaningful pattern between integration levels, but it does seem clear that there's a consistent (though relatively slight) depression in the fatigue responses by males in the High-Density group from that seen in all-male groups.

At the very least, 12 of the 13 Marines who had been in both groups and who had reported at least once (while in the Control group) a fatigue level of "5-Moderately Tired" or higher did not report any scores in this range when in the High-Density group. This is hardly definitive, since the number of trials any individual male might have participated in ranged from only one to five, but it is suggestive that this is a pretty consistent pattern.

As was proposed in a number of the Infantry results above for the 7km Hike under load—which frequently showed a similar reduction in fatigue for males in certain integrated groups—one potential reason for this shift after the 5km Hike could be the slower pace typically seen by High-Density groups (see Annex A).

N.4.9.1.2.2 Final Fatigue Scores (by Integration Level)

Post-trial fatigue results by Integration Level are very similar to those seen in the post-5km Hike above—the two groups are significantly different, with High-Density participants showing a higher representation (63.9%) in lowest fatigue levels (i.e., "1-Fully Alert" or "2-Very Lively") compared to males in the Control group (41.5%), with a corresponding decrease in the three higher-end scores (4.2% and 10.6%, for High-Density and Control groups, respectively).

When we looked more closely at the contributors to these higher scores ("5-Moderately Tired" to "7-Exhausted") in the Control group, we see a smaller sub-population (eight out of 42) compared to the 5km hike results. Indeed, seven of these eight had also contributed higher-end scores following the hike, which suggests—given the constraints of a relatively small number of trials—that this is not a universal trait (i.e., not representative of the majority of volunteers).

Conversely, of the 36 males who participated in both High-Density and Control groups, only two failed to show a general lowering of their post-trial fatigue range while in an integrated group (or were not already consistently reporting scores of "2-Very Lively" or less). Of course, we still suffer from a small number of trials for the Bridgeport phase), but this does suggest that the shift (though minor) towards less fatigue in the High-Density group is persistent.

As discussed in the post- hike section, one potential mechanism for this shift may be the generally slower pace of integrated squads, compared to Control squads. Of the thirteen integrated trials that were executed at Bridgeport, eleven had at least one female reporting a score of "5-Moderately Tired" or higher, although the majority of females (12 out of 19) never reported a fatigue score in this range.

If pace is the mechanism driving the difference between Control and integrated groups—and assuming that the Marines who are most tired (i.e., "5-Moderately Tired" or higher) drive the slower pace—then we might see different results if we didn't include the same subjects multiple times. In other words, because we had a sub-population that repeatedly reported high fatigue levels, including them over and over within a small set of trials may have over-inflated their contribution compared to the larger Marine Corps population.

It may also be that the slower pace in integrated groups could be influenced by group dynamics. The increased representation of the higher-end scores ("5-Moderately Tired" or higher) in all-male groups could mean a greater willingness to push a pace despite it being fatiguing to some members.

Table 69 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) – 03XX

Distribution of 03XX Fatigue Results (by Integration Level) Bridgeport				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
03XX	Bridgeport/1 Post-"5km Hike"	C	143	18.2%	28.7%	32.2%	6.3%	10.5%	2.8%	1.4%
		HD	71	32.4%	29.6%	25.4%	11.3%	1.4%	0.0%	0.0%
	Bridgeport/1 Post-Trial	C	142	16.9%	24.6%	28.2%	19.7%	7.7%	2.1%	0.7%
		HD	72	34.7%	29.2%	22.2%	9.7%	4.2%	0.0%	0.0%

N.4.9.2 Workload Results

For the Closed MOS (i.e., 03XX) volunteers at Bridgeport, we have maximum workload self-reports from 43 males and 19 females[23]. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-

[23] Two Marines—one male and one female—only participated in two trials (each).

Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.9.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females. Statistical results are summarized in the table below (Table 70).

Table 70 - Summary of Max Workload Scores (by Gender) – Closed MOS (03XX)

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ^2 Statistic	χ^2 p-Value
03XX	Max Workload for Trial	Bridgeport “Mountain”	M	212	4 / 4	14.1	<0.01*
			F	72	4 / 4		

We found a significant difference—by gender—among Closed MOS (i.e., 03XX) volunteers at Bridgeport. Males’ responses were typically clustered around “3-Active but Easy” or “4-Challenging” (75.0%), whereas females’ responses appear more broadly spread between “3-Active but Easy”, “4-Challenging”, and “5-Extremely Busy” (88.9%).

Table 71 - Max Workload Score Distributions (by Gender) – 03XX

Distribution of 03XX Workload Results (by Gender) Bridgeport				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
03XX	Bridgeport/1 "Mountain"	M	212	4.2%	9.9%	34.4%	40.6%	8.0%	1.9%	0.9%
		F	72	1.4%	2.8%	25.0%	43.1%	20.8%	6.9%	0.0%

For this scenario, we saw a rather wide range of responses for both genders (spanning from six to seven—out of seven possible—levels), yet nothing about the scenario could be said to be “MOS-specific” in nature and very little opportunity for compensation mechanisms—making it a bit difficult to ascertain what “workload” might mean in this context.

As such, we can’t rule out that the differences we saw in maximum workload scores between male and female 03XXs may have been closely tied to fatigue, which we had previously found to be significantly different between males and female. Even relatively low levels of fatigue can degrade concentration and performance of non-rote skills.

Although we don’t have great fidelity as to what tasks or elements elicited a particular max workload report, we do know that all “max” scores were reported in the final survey (as opposed to the post-5km Hike survey)—which suggests that peak “loading” occurred during either the gorge crossing, cliff ascent, or return hike. If females were—in general—a bit more tired following the 5km hike, then it would be understandable that (potentially) unfamiliar techniques associated with crossing a gorge and/or rock climbing might be more challenging, even if the actual work "load" remained constant from trial to trial.

N.4.9.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males' responses remained consistent at different integration levels. For the 03XXs, we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 6 females) squads. We have maximum workload self-reports from 43 males (42 in the Control group, and 36 in the High-Density group); statistical results are summarized in the table below (Table 72).

Table 72 - Summary of Males' Maximum Workload Scores (by Integration Level) - 03XX

MOS	Workload	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value
03XX	Max Workload for Trial	Bridgeport "Mountain"	C	140	4 / 4	15.3	<0.01*
			HD	72	3 / 3		

We saw a significant difference between males' responses in Control and High-Density groups at Bridgeport, likely the result of the pronounced shift towards lower maximum workload scores in the High-Density group (mostly "3-Active but Easy") compared to Control (mostly "4-Challenging"), as illustrated by the differences in their respective medians and modes.

Table 73 - Males' Max Workload Score Distributions (by Integration Level) – Closed MOS

Distribution of 03XX Workload Results (by Integration Level) Bridgeport				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
03XX	Bridgeport/1 "Mountain"	C	140	4.3%	7.1%	27.9%	45.0%	12.1%	2.1%	1.4%
		HD	72	4.2%	15.3%	47.2%	31.9%	0.0%	1.4%	0.0%

This might be driven by a couple of factors, including the lower fatigue scores often seen (by males) in High Density groups or a shift towards greater collaboration (i.e., a "spreading of the load") in integrated groups. Given the lack of much "MOS-specific" tasks at Bridgeport—and that the non-hiking tasks (i.e., the Gorge Crossing and Cliff Ascent tasks) aren't really conducive to collaboration between individuals, the influence of reduced fatigue (perhaps do to the slower pace typically seen in High Density groups—see Annex L) seems like the most plausible theory.

N.4.9.3 Cohesion Results

For the Close MOS (03XX) volunteers at Bridgeport, we have Unit Cohesion survey results (five statements per survey) from 43 males and 19 females[24]. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.9.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 74).

Table 74 - Summary of Unit Cohesion Scores (by Gender) - Closed MOS (03XX)

MOS	Cohesion	Gender	N	Mean	SD	χ2 Statistic	χ2 p-Value
03XX	Q1: Group was united in trying to reach goals	M	213	7.13	2.32	7.32	<0.01*
		F	72	6.28	2.69		
	Q2: We all take responsibility for task success of the group	M	213	7.12	2.30	5.73	0.02*
		F	72	6.38	2.58		
	Q3: Group members have similar aspirations for success	M	213	7.09	2.45	7.53	<0.01*
		F	72	6.29	2.59		
	Q4: If members had problems with a task, all wanted to help	M	213	7.17	2.30	15.09	<0.01*
		F	72	5.82	2.70		
	Q5: Members communicated freely about responsibilities	M	213	7.22	2.28	8.81	<0.01*
		F	72	6.22	2.64		
	Composite Score (sum of Q1-Q5)	M	213	35.73	11.35	9.17	<0.01*
		F	72	30.99	12.62		

We found a significant difference between 03XX males and females for all five cohesion statements ("Q1-Unity", "Q2-Responsibility", "Q3-Aspirations", "Q4-Help" and "Q5-Comm"), as well as for the overall composite cohesion score. When we look more

[24] Two Marines—one male and one female—only participated in two run-cycles (each).

closely at the distribution of responses for each statement, we see that females tended to be quite a bit more pessimistic (i.e., lower scores) than males.

Where males tended to have responses that cluster between “7” and “9” for all five questions (ranging from 69.5% to 72.8% of all responses) with a scattering of responses below “6” (towards Significantly Disagree), females had fewer contributions in the top three levels (45.8% to 58.3%), with a corresponding increase in representation below “6”—including roughly 10% at the lowest level, “1”.

We also graphed the means for each statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each question’s statement (Figure 14).

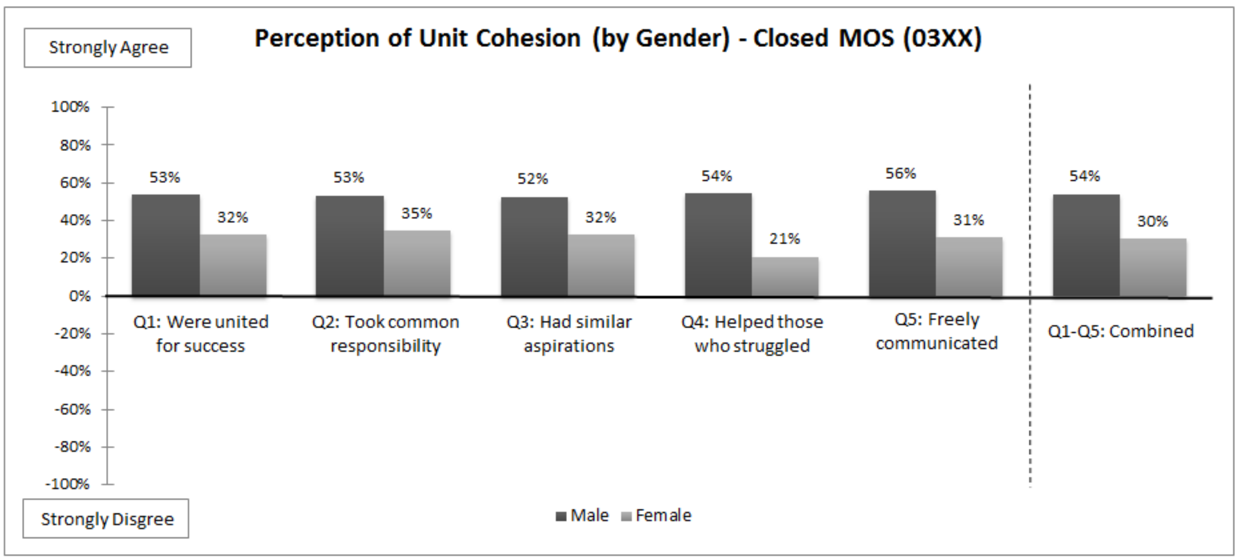


Figure 14 - Perception of Unit Cohesion (by Gender) - Closed MOS (03XX)

In looking at the average values for the 03XX volunteers, we can see how both genders—while more positive than negative—appeared to be somewhat neutral in their perceptions of unit cohesion, with females consistently showing a more pessimistic outlook on unit cohesion.

As nothing about the Mountaineering scenario was MOS-specific (and it’s unlikely that most—if any—of the Marines had previous exposure to the more technical tasks of crossing the gorge, or conducting a cliff ascent), relative experience is not likely a factor.

Since Bridgeport occurred at the end of an already long study, it may be that the general depression of cohesion (from the levels experienced by each constituent MOS, from their Twentynine Palms tenure)—and the sharper drop among females—may be a reflection of suddenly working with new individuals (i.e., different 03XX MOS) and/or falling morale.

N.4.9.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For the Closed MOS (03XX) at Bridgeport, we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 6 females) Squads. We have unit cohesion survey results (five statements per survey) from 43 males (42 in the Control group, and 36 in the High-Density group); statistical results are summarized in the table below (Table 75).

Table 75 - Summary of Males' Unit Cohesion Scores (by Integration Level) – Close MOS (03XX)

MOS	Cohesion	IL	N	Mean	SD	χ2 Statistic	χ2 p-Value
03XX	Q1: Group was united in trying to reach goals	C	141	7.79	1.78	32.57	<0.01*
		HD	72	5.83	2.69		
	Q2: We all take responsibility for task success of the group	C	141	7.82	1.69	34.93	<0.01*
		HD	72	5.74	2.69		
	Q3: Group members have similar aspirations for success	C	141	7.87	1.78	39.78	<0.01*
		HD	72	5.56	2.83		
	Q4: If members had problems with a task, all wanted to help	C	141	7.83	1.82	33.86	<0.01*
		HD	72	5.89	2.60		
	Q5: Members communicated freely about responsibilities	C	141	7.88	1.73	33.29	<0.01*
		HD	72	5.93	2.66		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	141	39.20	8.60	34.57	<0.01*
		HD	72	28.94	12.98		

We found a significant difference between the Control and High-Density groups for all five cohesion statements, as well as for the overall composite cohesion score. When we look more closely at the distributions, we see two very different patterns.

For the Control Group, we see the single most common response at "9" (48% to 55%), with a rather rapid drop off spanning from "8" to "7"—and only sporadic reports below that. With the High-Density groups, however, we see a very pronounced "triple peak", with peaks at the two extrema (21% to 22% at "9" and 13% to 17% at "1), and a "bulge" around the mid-point. This suggests pronounced difference of opinion by males—perhaps an artifact of a multi-MOS unit that had never worked together towards a common goal.

We also graphed the means for each statement as a percentage of how strongly males agreed (positive) or disagreed (negative) with each question's statement while in each group (Figure 15). The average values (by integration level) shown are consistent with what we saw with regards to males' more coherent, and relatively high perceptions of unit cohesion within the Control group, against the disjointed, and highly negative point scores in the High-Density group.

Unfortunately, the survey did not provide for reasons as to why males were so much more likely to choose such profoundly lower cohesion scores when in an integrated group. Given the nature of the scenario, it seems relative experience should not be a factor.

However, the 03XX pool was also composed of an assortment of various "03" MOS—who'd operated in parallel for several months, but not really together. Add to this, the Bridgeport phase taking place after a very long and tiring Twentynine Palms section, it's no wonder cohesion may have suffered. The unknown is why that effect seems to have been so heavily asymmetrical against integrated groups. Incidentally, these neutral results are consistent with females' perception of unit cohesiveness for these trials. As such, there may have well been a serious morale issues at play.

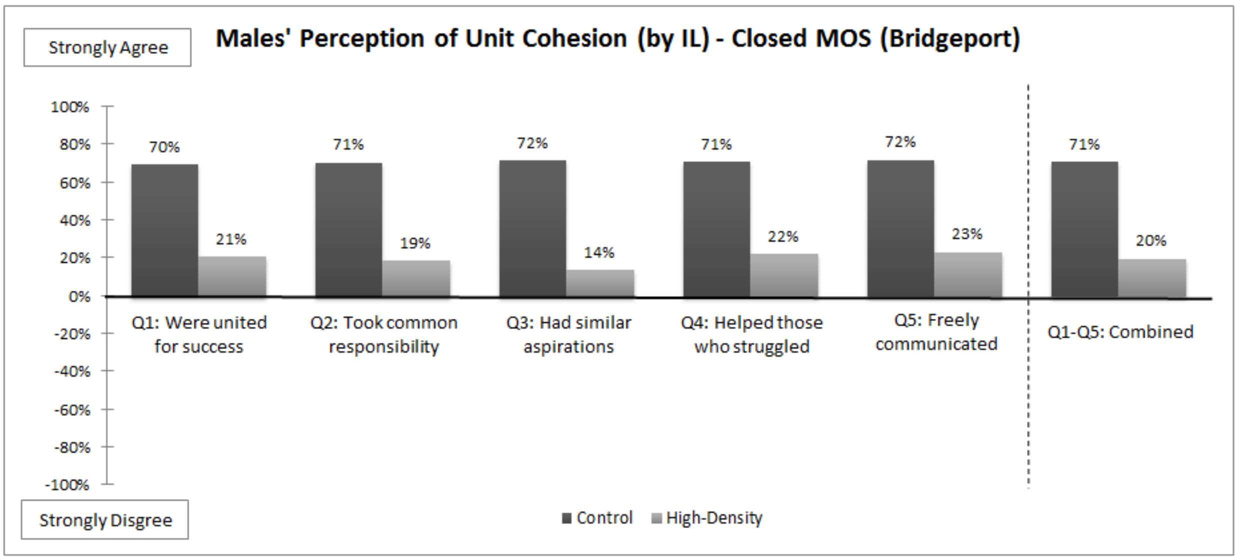


Figure 15 - Males' Perception of Unit Cohesion (by Integration Level) - 03XX

N.4.10 Mountaineering assessment (open MOSs only)

N.4.10.1 Fatigue Results

For the Open MOS (i.e., Provisional Infantry and 1371) volunteers at Bridgeport, we have baseline, mid-trial, and final fatigue self-reports from 45 males and 17 females[25]. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

[25] Four Marines, two males and two females only participated in two or fewer trials (each).

N.4.10.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial), their post-5km Hike (mid-trial), and final (post-trial) fatigue levels. Results are summarized in the table below (Table 76).

Table 76 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) – Open MOS

MOS	Fatigue	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
PI & 1371	Baseline (Pre-Trial)	Bridgeport/1 "Mountain"	M	179	2 / 2	18.3	<0.01*
			F	60	3 / 3		
	Post-5km (Mid-Trial)	Bridgeport/1 "Mountain"	M	176	3 / 3	14.2	<0.01*
			F	59	4 / 3		
	Final (Post-Trial)	Bridgeport/1 "Mountain"	M	179	3 / 2,3	15.5	<0.01*
			F	60	3 / 4		

N.4.10.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We found a significant difference between female and male PI and 1371 volunteers' pre-trial fatigue scores at Bridgeport. Males' responses typically clustered between "1-Fully Alert" and "3-Okay" (91.1%), whereas females had smaller numbers—though still a majority—in this range (76.7%), with more representation at "4-A Little Tired" (15.0%) compared to males (3.4%).

We found few reports in the higher fatigue scores (i.e., "5-Moderately Tired", "6-Extremely Tired", or "7-Exhausted") for either gender—5.6 percent for males (by five Marines, but mostly driven by one), and 8.3 percent for females (by three Marines, but mostly driven by one). Given the short duration of this phase of the study, and the relatively small number of trials per individual, it's not possible to determine whether there might be a pattern over time. However, with most pre-trial scores falling at "4-A Little Tired" or below, we believe most volunteers had sufficient recovery time between trials.

N.4.10.1.1.2 Final Fatigue Scores (by Gender)

In looking at the open MOS volunteers' post-trial responses at Bridgeport, we found significant differences in the distribution of scores by gender. Males showed a (scarce) majority clustered around "2-Very Lively" and "3-Okay" (53.6%), with shallow "tails" at "1-Fully Alert" and "4-A Little Tired" (Table 77). Females showed a very similar pattern, but shifted to the right by one level (i.e., centered at "3-Okay" and "4-A Little Tired", accounting for 65.0%), with the same shallow "tails" to either side. Both groups show a small representation at the two highest levels ("6-Extremelly Tired" and "7-Exhausted").

In looking more closely at these higher scores, we see that these scores represent only a relatively small sub-set of each group (i.e., two of 17 females and six of 45 males), with most contributions to this range being only one report per Marine; only one females and two males had more. However, as all but one male Marine had response ranges that fall within the "2-Very Lively" and "5-Moderately Tired" span, the shift we see towards slightly more fatigued results in the female group is likely reliable.

N.4.10.1.1.3 Post-5km Hike Fatigue Scores (by Gender)

The PI & 1371 volunteers also completed a fatigue survey following the completion of their 5km Hike under load at Bridgeport; we found a significant difference in the distributions of male and female responses for this as well. For both males and females we saw the same pattern of distributions as found post-trial (i.e., males clustered around "2-Very Lively" and "3-Okay" and females shifted a bit higher around "3-Okay" and "4-A Little Tired", with both groups having shallow "tails" to either side).

As with the post-trial results, when we looked more closely at who contributed to the scores for both groups at "6-Extremely Tired" and "7-Exhausted", only three of 17 females and five of 45 males contributed (but only one Marine per group did so more than once). Similar to the final fatigue results, since we only see two Marines (one male and one female) who don't have a range of responses within the "2-Very Lively" and "5-Moderately Tired" span, the post-hike shift towards slightly higher fatigue levels found in females is likely reliable.

Table 77 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) – Open MOS

Distribution of PI & 1371 Fatigue Results (by Gender) Bridgeport				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
PI & 1371	Bridgeport/1 Pre-Trial	M	179	21.8%	39.1%	30.2%	3.4%	2.2%	1.1%	2.2%
		F	60	3.3%	30.0%	43.3%	15.0%	8.3%	0.0%	0.0%
	Bridgeport/1 Post-"5km Hike"	M	176	12.5%	24.4%	33.5%	17.6%	8.0%	1.7%	2.3%
		F	59	0.0%	13.6%	35.6%	30.5%	10.2%	8.5%	1.7%
	Bridgeport/1 Post-Trial	M	179	14.0%	24.0%	29.6%	16.8%	9.5%	2.8%	3.4%
		F	60	5.0%	11.7%	35.0%	30.0%	13.3%	3.3%	1.7%

N.4.10.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. Open MOS (i.e., Provisional Infantry and 1371) volunteers at Bridgeport, we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 6 females) squads. We have unit cohesion survey results (five statements per survey) from 45 males (45 in the Control group, and 34 in the High-Density group); statistical results are summarized in the table below (Table 78).

Table 78 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) – Open MOS

MOS	Fatigue	Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
PI & 1371	Post-5km (Mid-Trial)	Bridgeport/1 "Mountain"	C	117	3 / 3	0.48	0.49
			HD	59	3 / 3		
	Final (Post-Trial)	Bridgeport/1 "Mountain"	C	143	3 / 3	7.3	<0.01*
			HD	71	2 / 1		

N.4.10.1.2.1 Post-5km Hike Fatigue Scores (by Integration Level)

Following the 5km Hike under load at Bridgeport, we did not find any significant difference in males' fatigue responses by Integration Level. In both groups, we see a (close) majority of responses falling between "2-Very Lively" and "3-Okay" (54.7% and 64.4% for Control and High-Density groups, respectively), although the Control group showed an increased likelihood of responses at "4-A Little Tired" (23.1%) than the High-Density group (6.8%). We also see a slightly greater representation in the higher range ("5-Moderately Tired" and higher) for High-Density group (15.3%) than in the Control group (10.3%).

When we look more closely at the contributors to these higher scores, we see that the Control group has a broader base (10 out of 45 Marines—most contributing a single response) than the High-Density group (four out of 35 Marines, largely driven by only two Marines). This suggests that the High-Density results in this range may largely be an artifact caused by a couple of Marines and not a pattern that would persist if we had the opportunity to run more trials without repeating subjects.

N.4.10.1.2.2 Final Fatigue Scores (by Integration Level)

Post-trial fatigue results by Integration Level were significantly different between groups, with Control groups showing a more shallow cluster between "2-Very Lively" and "3-Okay" tapering to "1-Fully Alert" and "4-A Little Tired" than the High-Density group that has a much sharper distribution peaking at "3-Okay" and skewing slightly to the left. However, at the macro level, both groups are very similar in the total representation of scores at "4-A Little Tired" and below (83.4% for Control group, 86.7% for High Density group), which suggests there may be little practical difference between the two groups.

As with the post-hike results, the higher-end scores seen at the end of the trial were more broadly supported within the Control group volunteers (one to two responses by 17 Marines out of 45) compared to the sparse contributions within the High-Density group (four Marines out of 35). As such, we might expect the shape of the upper distribution in the Control group to be preserved (if we had the opportunity to run more trials), whereas the High-Density upper distribution might become less prominent. This would be more consistent with a pattern we saw frequently among Infantry—males in highly-integrated groups typically showing a lower distribution of fatigue scores than all-male groups when there was a task involving a long-duration hike under load, perhaps due to slower paces (see Annex M). However, without additional trials to clarify, this is only theoretical.

Table 79 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) – Open MOS

Distribution of PI & 1371 Fatigue Results (by Integration Level) Bridgeport				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
PI & 1371	Bridgeport/1 Post-"5km Hike"	C	117	12.0%	23.9%	30.8%	23.1%	7.7%	2.6%	0.0%
		HD	59	13.6%	25.4%	39.0%	6.8%	8.5%	0.0%	6.8%
	Bridgeport/1 Post-Trial	C	119	15.1%	25.2%	25.2%	17.6%	11.8%	3.4%	1.7%
		HD	60	11.7%	21.7%	38.3%	15.0%	5.0%	1.7%	6.7%

N.4.10.2 Workload Results

For the Open MOS (i.e., Provisional Infantry and 1371) volunteers at Bridgeport, we have maximum workload self-reports from 45 males and 17 females[26]. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.10.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females. Statistical results are summarized in the table below (Table 80).

Table 80 - Summary of Max Workload Scores (by Gender) – Open MOS

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
PI & 1371	Max Workload for Trial	Bridgeport/1 "Mountain"	M	178	4 / 4	11.0	<0.01*
			F	60	4 / 4		

As in the Closed MOS results, we found a significant difference—by gender—among Open MOS (i.e., PI and 1371) volunteers at Bridgeport. Males' responses typically

[26] Four Marines—two males and two females—only participated in two or fewer trials (each).

spanned from “2-Little to Do” to “5-Extremely Busy” (89.9%), and peaking at “4-Challenging”. Female responses appeared to be more tightly clustered between “3-Active but Easy” and “5-Extremely Busy”, with a sharper peak at “4-Challenging” (96.7%).

As discussed in the Closed MOS section (see N.5.9.2), it’s difficult to determine what “workload” might mean within the context of the Bridgeport Scheme of Maneuver, given that the tasks required were not specific to any particular MOS and the Scheme of Maneuver did not generally allow for compensation mechanisms. It’s possible that the differences we saw in max workload scores between Open MOS males and females may have been closely linked with fatigue, which were also significantly different between males and female. Even relatively low levels of fatigue can degrade concentration and performance of non-rote skills.

Table 81 - Max Workload Score Distributions (by Gender) – Open MOS (PI & 1371)

Distribution of PI & 1371 Workload Results (by Gender) Bridgeport				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
PI & 1371	Bridgeport/1 "Mountain"	M	178	3.4%	14.6%	28.1%	32.6%	14.6%	4.5%	2.2%
		F	60	0.0%	0.0%	20.0%	50.0%	26.7%	1.7%	1.7%

Although we don’t have great fidelity as to what tasks or elements elicited a particular maximum workload report, we do know that all “max” scores were reported in the final survey (as opposed to the post-5km Hike survey)—which suggests that peak “loading” occurred during either the gorge crossing, cliff ascent, or return hike. If females were—in general—a bit more tired following the 5km hike, then it would understandable that (potentially) unfamiliar techniques associated with crossing a gorge and/or rock climbing might be more challenging, even if the actual work “load” remained constant from trial to trial.

N.4.10.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males’ responses remained consistent at different integration levels. For PI and 1371s, we compared males’ responses in Control (i.e., all-male) and High-Density (i.e., 6 females) Squads. We have max workload self-reports from 45 males (45 in the Control group, and 34 in the High-Density group); statistical results are summarized in the table below (Table 82).

Table 82 - Summary of Males' Max Workload Scores (by Integration Level) - Open MOS

MOS	Workload	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
PI & 1371	Max Workload for Trial	Bridgeport/1 "Mountain"	C	118	4 / 4	0.70	0.40
			HD	60	4 / 4		

We did not see a significant difference between males' responses in Control and High-Density groups at Bridgeport for Open MOS (i.e., PI and 1371) volunteers. For both groups, males tend to favor maximum workload scores between "3-Active but Easy" and "4-Challenging", though there is also some representation at all levels.

Table 83 - Males' Max Workload Score Distributions (by Integration Level) – Open MOS

Distribution of PI& 1371 Workload Results (by Integration Level) Bridgeport				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
PI & 1371	Bridgeport/1 "Mountain"	C	118	2.5%	19.5%	25.4%	30.5%	16.9%	4.2%	0.8%
		HD	60	5.0%	5.0%	33.3%	36.7%	10.0%	5.0%	5.0%

N.4.10.3 Cohesion Results

For the Open MOS (PI and 1371) volunteers at Bridgeport, we have Unit Cohesion survey results (five statements per survey) 45 males and 17 females[27]. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.10.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 84).

[27] Four Marines—two males and two females—only participated in two of fewer trials (each).

Table 84 - Summary of Unit Cohesion Scores (by Gender) - Open MOS (PI & 1371)

MOS	Cohesion	Gender	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
PI & 1371	Q1: Group was united in trying to reach goals	M	178	7.98	1.79	3.46	0.06*
		F	60	8.03	0.99		
	Q2: We all take responsibility for task success of the group	M	178	7.99	1.75	4.34	0.04*
		F	60	7.90	1.17		
	Q3: Group members have similar aspirations for success	M	178	7.93	1.79	3.84	0.05*
		F	60	7.87	1.19		
	Q4: If members had problems with a task, all wanted to help	M	178	8.06	1.68	7.41	<0.01*
		F	60	7.87	1.11		
	Q5: Members communicated freely about responsibilities	M	178	8.07	1.66	4.44	0.04*
		F	60	8.05	0.94		
	Composite Score (sum of Q1-Q5)	M	178	40.02	8.38	4.39	0.04*
		F	60	39.64	5.07		

We found a significant difference between 03XX males and females for all five cohesion statements ("Q1-Unity", "Q2-Responsibility", "Q3-Aspirations", "Q4-Help" and "Q5-Comm"), as well as for the overall composite cohesion score. When we look more closely at the distribution of responses for each statement, we a couple of competing patterns.

On one hand, males' responses tend to cluster a bit more strongly at the two highest levels, "8" and "9" (78.1% to 82.6%) than females, who have slightly more shallow but broad (and inclusive) clustering between "7" and "9" (88.3% to 94.9%). Males were also (slightly) more likely to report scores at the lowest three levels, "1" to "3" (3.9% to 5.1%) compared to females (0%).

However, when we look more closely at the contributors to these strongly negative scores (i.e., "1" to "3") among the males, we see that they are from only five Marines (out of 45), and so it's unlikely this is a universal viewpoint (or at least a relatively rare one)—though their exclusion does not change our statistical results. This suggests the significance difference we found is driven by these relatively few (and non-universal) responses—a difference that seems unlikely to be practically relevant, given that we can't be certain all Marines used the same internal criteria for judging cohesion.

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement (Figure 16). In looking at the average values for the Open MOS volunteers, we can see how both genders—despite some nuances as to their specific distributions—appear to be largely consistent (and positive) in their relative perceptions of unit cohesion

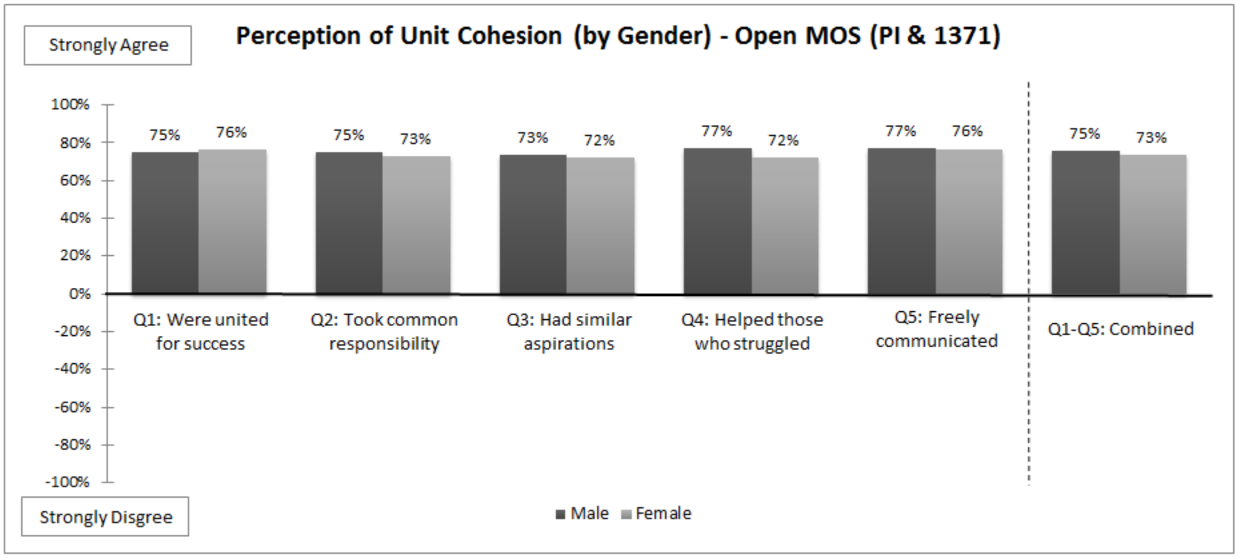


Figure 16 - Perception of Unit Cohesion (by Gender) - Open MOS (PI & 1371)

N.4.10.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For Open MOS (i.e., PI and 1371), we compared males' responses in Control (i.e., all-male) and High-Density (i.e., 6 females) Squads. We have unit cohesion survey results (five statements per survey) from 45 males (45 in the Control group, and 34 in the High-Density group); statistical results are summarized in the table below (Table 85).

Table 85 - Summary of Males' Unit Cohesion Scores (by Integration Level) Open MOS (PI & 1371)

MOS	Cohesion	IL	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
PI & 1371	Q1: Group was united in trying to reach goals	C	118	8.16	1.36	0.40	0.53
		HD	60	7.62	2.40		
	Q2: We all take responsibility for task success of the group	C	118	8.10	1.40	0.17	0.68
		HD	60	7.77	2.28		
	Q3: Group members have similar aspirations for success	C	118	8.08	1.48	0.59	0.44
		HD	60	7.62	2.26		
	Q4: If members had problems with a task, all wanted to help	C	118	8.21	1.29	0.11	0.74
		HD	60	7.75	2.23		
	Q5: Members communicated freely about responsibilities	C	118	8.20	1.32	0.09	0.77
		HD	60	7.82	2.17		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	118	40.76	6.53	0.14	0.70
		HD	60	38.57	11.09		

Unlike the Closed MOS, we did not see a significant difference between the two integration levels, regardless of the question. For both groups, we see the single most represented score at "9", which drops off rather quickly by "6"—across all questions. For the High-Density group, we do see a slight variation in the form of a small, but visible, percentage of responses at the lowest level (7% to 10%)—though this is not enough to significantly diverge the two sets of response.

We also graphed the means for each statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each question's statement (Figure 17). The average values (by integration level) shown are consistent with what we saw with regards to comparable representation for the vast majority of responses, but with a slight dip in the High-Density means due to the presence of some responses at "1". These responses are consistent with what females reported above.

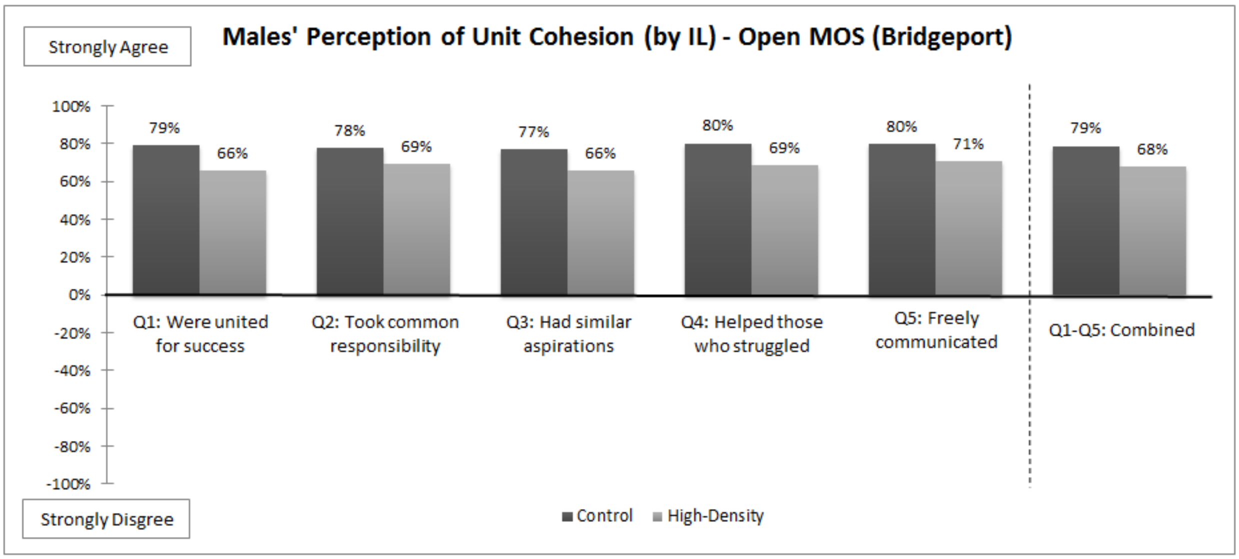


Figure 17 - Males' Perception of Unit Cohesion (by Integration Level) - Open MOS (Bridgeport)

N.4.11 Field Artillery Cannoneer (MOS 0811)

N.4.11.1 Fatigue Results

For the 0811 volunteers at Twentynine Palms, we have baseline, mid-trial, and final fatigue self-reports from 28 males and twelve females[28]. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). If we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—

[28] One male and one female only participated in a single run-cycle.

then we conducted Wilcoxon pairwise comparisons (using p<0.033 to adjust for multiple comparisons).

N.4.11.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial), post-ROSP (mid-trial), and final (post-trial) fatigue levels. Results are summarized in the table below (Table 86).

Table 86 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) - 0811

MOS	Fatigue	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
0811	Baseline (Pre-Trial)	29Palms/1 "Live-Fire"	M	698	3 / 3	0.004	0.95
			F	153	2 / 2		
	Post-RSOP (Mid-Trial)	29Palms/1 "Live-Fire"	M	689	3 / 3	2.8	0.09*
			F	152	3 / 3		
	Final (Post-Trial)	29Palms/1 "Live-Fire"	M	686	3 / 3	6.8	<0.01*
			F	150	3.5 / 4		

N.4.11.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We did not find a significant difference between female and male 0811 volunteers' pre-trial fatigue scores. In both gender groups, pre-trial responses typically clustered between "1-Fully Alert" and "3-Okay" (80.2% and 83.0% for males and females, respectively) (Table 87).

We also found a relatively small, but persistent, number of reports in the higher fatigue levels for both genders (i.e., "5-Moderately Tired", "6-Extremely Tired", and "7-Exhausted")—representing 13 percent of male responses (spread across 20 Marines), and 9.2 percent of female responses (almost exclusively by a single Marine). In many instances, these higher values represented only 1-3 responses for most Marines (15 out of 23) over the course of the entire study.

However, a small number of Marines (seven males and one female) frequently reported such high pre-trial fatigue levels (ranging from six to 18 responses each), over a period of 29 run-cycles. This suggests some sub-set of 0811 Marines were struggling with very high levels of fatigue before beginning a trial.

We should also note, however, that due to a reduction in the number of male 0811s available—either due to medical restrictions or through dropping from the study entirely—males were more likely to participate on a "test" howitzer (i.e., a 6-man crew) than females[29]. While all Marines (on full-duty status) not assigned to a trial unit (i.e., “test” howitzer) were assigned to a spare howitzer in order to maintain consistent physical loading of the 0811 population, this “loading” gun often had more than six crewmembers. As such, some males may have been exposed to more physical stress over the course of the study, which might account for the presence of these higher pre-trial scores.

N.4.11.1.1.2 Post-RSOP Fatigue Scores (by Gender)
While not technically a part of the 0811 trial, volunteers participated in a series of Reconnaissance, Selection, and Occupation of Position (RSOP) tasks (see Annex H) prior to beginning the trial proper. This served the dual purpose of allowing the assigned Section Chief (a direct assignment—not a volunteer) an opportunity to get familiar with his crew prior to the start of the test trial as well as introduce additional physical stress on a crew in order to more closely mimic a realistic scenario.

Upon completion of the RSOP phase, 0811 volunteers completed a fatigue survey. We found a significant difference in response distributions by gender, with a small shift to the right towards “4-A Little Tired” and “5-Moderately Tired” for females (up 10% from pre-trial results), compared to the much smaller rise in this same range for males (up 2% from pre-trial results).

When we looked at individual Marines’ responses to pre-trial and post-RSOP surveys for the same trial, we found a consistent results. Though not representing a majority, we found females were a bit more likely to report a higher scores post-RSOP (30.5%) than males (18.9%), though often an increase of only one.

N.4.11.1.1.3 Final Fatigue Scores (by Gender)
In looking at the volunteers’ post-trial responses, we likewise found a significant difference between genders. While exactly half (50.0%) of female responses remained in lower range (“1-Fully Alert”, “2-Very Lively” or “3-Okay”), a sizeable percentage (46%) reported being “4-A Little Tired” or “5-Moderately Tired” compared to males (27.0%).

Incidentally—and consistent across all three surveys (pre-trial, post-RSOP, and post-trial)—males led females (by a small margin) in the percentage of responses in the highest range (“6-Extremely Tired” and “7-Exhausted”), with 5.4 percent, compared to 4.0 percent for females. However, this cannot be explained as a lack of movement in this highest range from pre-trial results.

[29] Females conducted, on average only 14 run-cycles (with the exclusion of one female who only participated in a single trial). Males, on the other had conducted, on average, 26 trials (with the exception of one male who only participated in one trial).

When we looked at individual Marines' responses, we saw a sizeable number of males who'd completed a trial with a fatigue score of "6-Extremely Tired" or "7-Exhausted" had actually started relatively fresh with a pre-trial score of "3-Okay" or lower (40.1%). Conversely, of the Marines that had started the day extremely fatigued (reporting a pre-trial level of "6-Extremely Tired" or "7-Exhausted"), we saw a similar percentage of males (41.5%) who reported being only "4-A Little Tired" or less. Both sub-populations—i.e., fresh to extremely fatigued, and extremely fatigued to fresh—had representatives in all six billets.

Table 87 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) - 0811

Distribution of 0811 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0811	29Palms/1 Pre-Trial	M	698	30.2%	18.6%	31.4%	6.7%	7.2%	2.3%	3.6%
		F	153	20.3%	33.3%	29.4%	7.8%	7.8%	0.7%	0.7%
	29Palms/1 Post-RSOP	M	689	25.4%	18.7%	35.0%	7.3%	8.6%	2.0%	3.0%
		F	152	14.5%	27.0%	30.9%	15.1%	10.5%	2.0%	0.0%
	29Palms/1 Post-Trial	M	686	13.3%	17.9%	36.4%	14.3%	12.7%	2.6%	2.8%
		F	150	2.0%	28.7%	19.3%	29.3%	16.7%	2.7%	1.3%

N.4.11.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 0811s, we compared males' responses in Control units (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) Artillery Sections. We have post-RSOP and final fatigue self-reports from 28 males; statistical results are summarized in the table below (Table 88).

Table 88 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) - 0811

MOS	Fatigue	Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0811	Post-RSOP (Mid-Trial)	29Palms/1 "Live-Fire"	C	251	3 / 3	1.3	0.53		
			LD	226	3 / 3				
			HD	212	3 / 3				
	Final (Post-Trial)	29Palms/1 "Live-Fire"	C	255	3 / 3	1.65	0.44		
			LD	220	3 / 3				
			HD	211	3 / 3				

N.4.11.1.2.1 Post-RSOP Fatigue Scores (by Integration Level)

Males' responses following the completion of the RSOP phase did not show any significant difference between any of the three integration levels (i.e., Control, Low-Density, and High-Density). As we saw in the "by gender" results above, responses were predominantly clustered between "1-Fully Alert" and "3-Okay" for all three groups (79.4% to 80.9%) (Table 89).

N.4.11.1.2.2 Final Fatigue Scores (by Integration Level)

Males' post-trial responses likewise did not show any significant difference between any of the three integration levels (i.e., Control, Low-Density, and High-Density)—consistent with the "by gender" results above, with a slight shift towards "4-A Little Tired" and "5-Moderately Tired" but largely clustered at "3-Okay" or below. This—like the last of difference for the post-RSOP results—is perhaps not surprising given that the roles to be performed by each billet were carefully controlled, with very little opportunity for the sharing or switching of tasks.

It may be the fatiguing nature of each task within a trial—i.e., its contribution to total, post-trial fatigue—could be influenced by who is filling a particular critical billet. For example, a Marine who's particularly slow performing their portion of a fire mission might allow the other members of the crew more time to recover, making them less fatigued than they might otherwise be if at a faster tempo. However, because we only have overall post-trial fatigue scores—and not post-task scores—we cannot determine if this holds true for this study. Alternatively, a different Scheme of Maneuver plan that allowed for sharing and/or swapping of duties might be able to tease out potential differences based upon critical billets.

Table 89 – Males' Mid-Trial and Final Fatigue Score Distributions (by Integration Level) - 0811

Distribution of 0811 Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0811	29Palms/1 Post-RSOP	C	251	25.9%	19.1%	35.9%	6.8%	7.6%	2.4%	2.4%
		LD	226	26.5%	19.5%	33.6%	4.9%	10.6%	1.3%	3.5%
		HD	212	23.6%	17.5%	35.4%	10.4%	7.5%	2.4%	3.3%
	29Palms/1 Post-Trial	C	255	15.3%	18.0%	36.1%	13.7%	11.8%	2.7%	2.4%
		LD	220	13.2%	15.9%	40.5%	11.8%	14.5%	1.8%	2.3%
		HD	211	10.9%	19.9%	32.7%	17.5%	11.8%	3.3%	3.8%

N.4.11.2 Workload Results

For the 0811 volunteers at Twentynine Palms, we have max workload self-reports from 28 males and 12 females[30]. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \le 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.11.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females. Results are summarized in the table below (Table 90).

Table 90 - Summary of Max Workload Scores (by Gender) – 0811

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
0811	Max Workload for Trial	29Palms/1 "Live-Fire"	M	704	3 / 3	16.3	<0.01*
			F	154	4 / 4		

We found a significant difference between female and male 0811 volunteers' maximum workload scores. While both males and females were found to be strongly clustered at "3-Active but Easy" and "4-Challenging" (79.1% and 77.9%, for males and females, respectively), males more strongly favored the former and females the latter (**Table 91**).

[30] One male and one female only participated in a single trial (each).

We found few of the highest max workload scores for either group (i.e., “6-Overloaded” or “7-Unmanageable”)—only 5.3 percent of male responses and 5.2 percent of female responses. While a larger sub-population of males contributed to these higher max scores (thirteen out 28 males) than females—who had only two Marines contributing to this range—most Marines only did so a couple of times. This suggests most Marines felt capable of accomplishing (if not always easily) the workload assigned them.

As for the difference between the two genders, this may be due to females having less time in the MOS—a distinct disadvantage on an Artillery Section, where all members of the crew must work smoothly together and there was little opportunities for compensation. If experience is a factor, we would expect the difference between the two groups to gradually disappear over time as they became more skilled. This may have also been influenced by the slightly higher levels of fatigue reported by female 0811s (both post-RSOP and post-trial), since even relatively low levels of fatigue can degrade concentration and performance of non-rote skills.

Table 91 - Maximum Workload Score Distributions (by Gender) - 0811

Distribution of 0811 Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0811	29Palms/1 "Live-Fire"	M	704	1.8%	7.0%	46.2%	33.0%	6.8%	2.1%	3.1%
		F	154	0.0%	3.2%	34.4%	43.5%	13.6%	4.5%	0.6%

N.4.11.2.2 Workload Results by Integration Level

In addition to looking at how max workload self-reports compared between males and females, we also examined whether males’ responses remained consistent at different integration levels. For 0811s, we compared males’ responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) Artillery Sections. We have maximum workload self-reports from 28 males (28 males in the Control group, and 27 males in the Low-Density and High-Density groups); statistical results are summarized in the table below (Table 92).

Table 92 - Summary of Males' Maximum Workload Scores (by Integration Level) - 0811

MOS	Workload	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0811	Max Workload for Trial	29Palms/1 “Live-Fire”	C	258	3 / 3	0.21	0.90		
			LD	230	3 / 3				
			HD	216	3 / 3				

We did not see a significant difference between males' responses, regardless of the integration level of the unit they were in. For all three groups, males tend to favor maximum workload scores between "3-Active but Easy" and "4-Challenging", though there is also some representation at all levels.

However, we should also note that the design of Artillery Scheme of Maneuver, which largely prohibited either collaboration or compensation between members of the same crew. As such, we are unable to conclude whether this result would hold under conditions where members would be allowed more leeway in their respective division of labor.

Table 93 - Males' Max Workload Score Distributions (by Integration Level) – 0811

Distribution of 0811 Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0811	29Palms/1 "Live-Fire"	C	258	2.3%	6.6%	46.1%	32.9%	6.6%	2.7%	2.7%
		LD	230	0.4%	6.5%	48.3%	31.7%	8.7%	1.3%	3.0%
		HD	216	2.8%	7.9%	44.0%	34.3%	5.1%	2.3%	3.7%

N.4.11.3 Cohesion Results

For the 0811 volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) 28 males and 12 females[31]. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). If we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.11.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between

[31] One male and one female only participated in a single run-cycle (each).

males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 94).

Table 94 - Summary of Unit Cohesion Scores (by Gender) - 0811

MOS	Location	Cohesion	Gender	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
0811	29Palms	Q1: Group was united in trying to reach goals	M	691	7.59	1.92	7.05	<0.01*
			F	150	8.15	1.37		
		Q2: We all take responsibility for task success of the group	M	691	7.58	1.93	3.63	0.06*
			F	150	8.01	1.58		
		Q3: Group members have similar aspirations for success	M	691	7.57	1.94	4.54	0.03*
			F	150	8.07	1.42		
		Q4: If members had problems with a task, all wanted to help	M	691	7.52	2.01	0.00	0.98
			F	150	7.69	1.77		
		Q5: Members communicated freely about responsibilities	M	691	7.70	1.89	2.82	0.09*
			F	150	8.15	1.37		
		Composite Score (sum of Q1-Q5)	M	691	37.95	9.36	2.55	0.11
			F	150	40.07	6.84		

We found a significant differences between 0811 males and females for four of the five cohesion statements ("Q1-United", "Q2-Responsibility", "Q3-Aspirations", and "Q5-Comm"), but not for "Q4-Help" or the overall composite cohesion score (although the latter is borderline). Overall, females tended towards being a bit more optimistic than males in their perception of their units' cohesion.

When we look more closely at the distributions of each question, we see that males reported a scant majority at the highest level possible, "9" (51.7% to 56.7%), for all questions but one ("Q4-Help"), but then taper off fairly quickly by "6" only to extend out to the lowest level "1" with low-level contributions that, in total, represent 20.4% to 23.7% of all of males' responses.

On the other hand, females tend to peak even more sharply at "9" (52.0% to 56.0%), and taper off much more completely—leaving only 8.0% to 9.3% spread across "1" to "6" for all questions except "Q4-Help". For the latter—the one question that is not significantly different from males—we see a more pessimistic spread (17.3%) in between "1" to "6".

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement (Figure 18).

The average values—by gender—reflect the differences we saw in the distribution of unit cohesion scores—females tend towards slightly more optimistic scores than males, though both are positive.

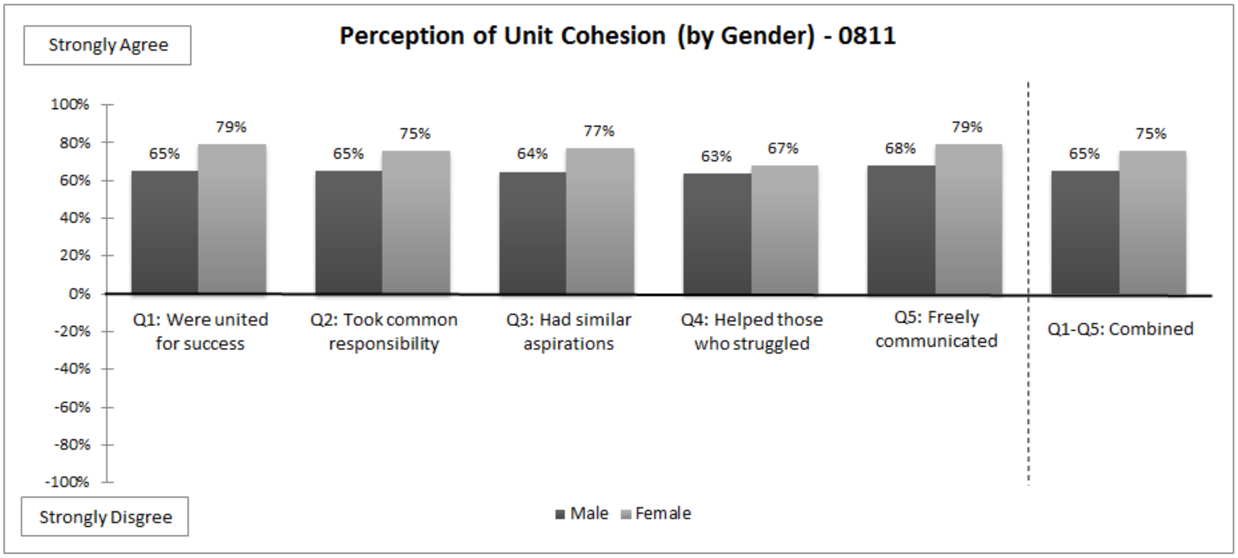


Figure 18 - Perception of Unit Cohesion (by Gender) – 0811

N.4.11.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For 0811s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) Artillery Sections. We have unit cohesion survey results (five statements per survey) from 28 males (28 males in the Control group, and 27 males in the Low-Density and High-Density groups); statistical results are summarized in the table below (Table 95).

Table 95 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 0811

MOS	Cohesion	IL	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0811	Q1: Group was united in trying to reach goals	C	256	7.95	1.90	33.49	<0.01*	(LD-C) -3.20	<0.01*
		LD	223	7.57	1.84			(HD-C) -5.81	<0.01*
		HD	212	7.18	1.95			(HD-LD) -2.47	0.01*
	Q2: We all take responsibility for task success of the group	C	256	7.95	1.88	34.56	<0.01*	(LD-C) -3.06	<0.01*
		LD	223	7.55	1.89			(HD-C) -5.96	<0.01*
		HD	212	7.17	1.96			(HD-LD) -2.59	0.01*
	Q3: Group members have similar aspirations for success	C	256	7.93	1.89	27.68	<0.01*	(LD-C) -3.27	<0.01*
		LD	223	7.45	1.96			(HD-C) -5.30	<0.01*
		HD	212	7.25	1.94			(HD-LD) -1.63	0.10
	Q4: If members had problems with a task, all wanted to help	C	256	7.85	2.01	27.98	<0.01*	(LD-C) -2.95	<0.01*
		LD	223	7.44	2.03			(HD-C) -5.37	<0.01*
		HD	212	7.19	1.94			(HD-LD) -2.07	0.04
	Q5: Members communicated freely about responsibilities	C	256	8.00	1.90	26.92	<0.01*	(LD-C) -3.27	<0.01*
		LD	223	7.61	1.89			(HD-C) -5.17	<0.01*
		HD	212	7.42	1.83			(HD-LD) -1.67	0.10
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	256	39.68	9.40	34.49	<0.01*	(LD-C) -3.71	<0.01*
		LD	223	37.61	1.89			(HD-C) -5.87	<0.01*
		HD	212	36.22	9.06			(HD-LD) -1.96	0.05

We found a statistical difference between the Control group and both integration levels (i.e., High-Density, and Low-Density groups) for all five cohesion statements, as well as for the overall composite cohesion score. We also saw mixed results between High-Density and Low-Density groups; significant for “Q1-Unity”, “Q2-Responsibility”, and “Q3-Aspirations”, but not for the other two statements, nor the overall composite score.

When we looked at the distributions of each question, we see very similar patterns—the majority of responses are heavily skewed to the far right (i.e., “9”, at Significantly Agree), and taper off relatively quickly to the left. Where the differences appear to come in is the "thickness" within the middle ranges (i.e., "7" to "3") that appears to get slightly more pronounced with integration density. The High-Density group is also slightly more likely to have some of the lowest scores, "1" or "2" (4% to 5%).

We also graphed the means for each statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each question's statement (Figure 19). The average values (by integration level) shown are consistent with what we saw with regards to males' somewhat higher perceptions of unit cohesion within the Control group, along with the easy decline in their perception of cohesion when in the Low- and High-Density groups.

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group, although given the fast pace and relatively locked-in responsibilities-by-billet, anyone weaker in key skills might well create an more negative group dynamic.

Since the scenario did not allow for much compensation or collaboration between group members, it wouldn't be too surprising for deficits in skill/deficits in empathy with such struggles to translate (for some) into lack of motivation and/or unity.

Incidentally, males' perceptions in the integrated units are a bit more depressed compared to how females perceived unit cohesiveness in the integrated units.

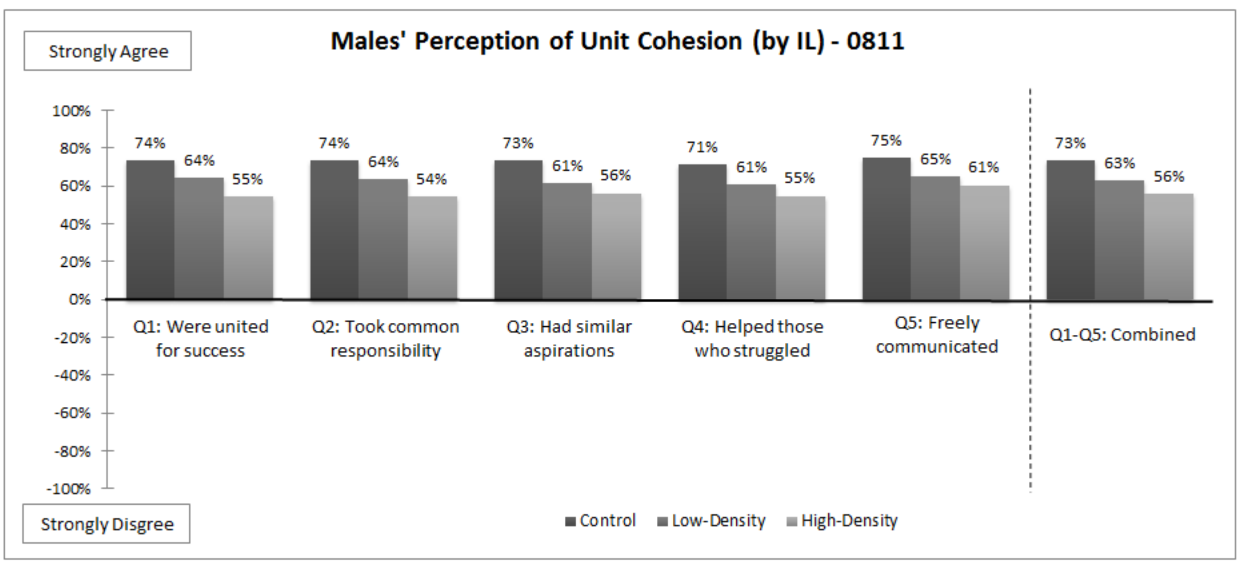


Figure 19 - Males' Perception of Unit Cohesion (by Integration Level) - 0811

N.4.12 LAV Crewman (0313)

N.4.12.1 Fatigue Results

For the 0313 volunteers at Twentynine Palms, we have baseline and final fatigue self-reports from 14 males and seven females. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.12.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial) and final (post-trial, for both days in the run-cycle) fatigue levels. Results are summarized in the table below (Table 96).

Table 96 - Summary of Pre-Trial and Final Fatigue Scores (by Gender) - 0313

MOS	Fatigue	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
0313	Baseline (Pre-Trial)	29Palms/1,2	M	336	1 / 1	11.5	<0.01*
			F	165	2 / 2		
	Final (Post-Trial)	29Palms/1 "Non-Live Fire"	M	171	2 /1	0.56	0.45
			F	84	2 / 2		
		29Palms/2 "Live-Fire"	M	165	2 / 1	3.29	0.07*
			F	81	2 / 2		

N.4.12.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We found a significant difference between female and male 0313 volunteers' pre-trial fatigue scores, with males more likely than females to report being a bit less fatigued. The males' primary responses were "1-Fully Alert" (52.7%), with "2-Very Lively" (32.4%) a strong second; this pattern was flipped for females (33.3% and 51.5%, respectively) (Table 97).

However, given the closeness in verbiage between "1-Fully Alert" and "2-Very Lively"—and that for both genders these two levels encompass approximately 85% of all scores—this difference doesn't reflect any practical difference. From this, we conclude that all 0313 Marines had sufficient recovery time between trials, with no discernable decrease over time.

N.4.12.1.1.2 Final Fatigue Scores (by Gender)

In looking at the volunteers' post-trial responses, we found mixed results: we found no significant difference in scores for non-live fire (i.e., Maintenance) trials ($p=0.45$), but a significant difference for live-fire trials ($p=0.07$). On the face of it, these results seem somewhat counter-intuitive. The Non-Live Fire scenario contained a number of physically demanding tasks, including evacuation of a 204-lb, combat-loaded test dummy from the Vehicle Commander's station, as well as the removal and remounting of scout hatch and side-armor panels and the spare tire assembly.

However, with the exception the CASEVAC event (which was conducted by the Gunner and Driver), all of the Non-Live Fire scenario tasks were conducted by the entire LAV crew, whereas the Live-Fire scenario (which had some physically demanding tasks associated with prepping the vehicle for combat) required many of the tasks to be completed by a single Marine (often the Gunner). As such, the Live-Fire scenario provided much less opportunity for compensation across the trial by crewmembers who might otherwise struggle.

In either event, the difference in distributions for the Live-Fire trials between males and females mirrors the pattern seen with pre-trial results: males predominately reported a final fatigue score of “1-Fully Alert” (49.1%) followed by “2-Very Lively” (33.9%); and females predominately reported “2-Very Lively” (50.6%) followed by “1-Very Lively” (33.3%). We saw little to no representation in the highest levels (“5-Moderately Tired”, “6-Extremely Tired”, or “Exhausted”) for either gender. This suggests that—despite the statistical difference between the genders for the Live-Fire trials—0313 Marines of either sex were not greatly fatigued by either scenario[32].

Table 97 - Pre-Trial and Final Fatigue Score Distributions (by Gender) - 0313

Distribution of 0313 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0313	29Palms/1&2 Pre-Trial	M	336	52.7%	32.4%	12.2%	1.5%	1.2%	0.0%	0.0%
		F	165	33.3%	51.5%	9.1%	4.2%	1.8%	0.0%	0.0%
	29Palms/1 Post-"Maint"	C	171	38.0%	31.0%	22.8%	8.2%	0.0%	0.0%	0.0%
		HD	84	34.5%	46.4%	15.5%	3.6%	0.0%	0.0%	0.0%
	29Palms/2 Post-"Live-Fire"	C	165	49.1%	33.9%	15.2%	1.2%	0.6%	0.0%	0.0%
		HD	81	33.3%	50.6%	11.1%	4.9%	0.0%	0.0%	0.0%

N.4.12.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 0313s, we compared males’ responses in Control (i.e., all-male) versus Low-Density (i.e., 1 female) and High-Density (i.e., 2 females) crews. We have

[32] The 0313 Scheme of Maneuver for the Live-Fire scenario rarely permitted administering a fatigue survey immediately after the conclusion of the last task, Engaging Offensive Targets. LAV crewmembers had to transit off of the range, and the time from the end of the last task to the start of the surveys could vary from as short at 5 minutes to as long as an hour or more (average time was 29 minutes). Delays between completion of a trial and taking the survey could have muted perceptions of fatigue, as Marines had time to recover from their exertions.

final fatigue self-reports from 14 males in Control and Low-Density groups, and five males in High-Density groups[33]; statistical results are summarized in the table below summarized in the table below (Table 98).

Table 98 – Summary of Males' Final Fatigue Scores (by Integration Level) - 0313

MOS	Fatigue	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0313	Final (Post-Trial)	29Palms/1 "Non-Live Fire"	C	87	2 / 1	7.98	0.02*	(LD-C) 0.05	0.96
			LD	56	2 / 1			(HD-C) 2.73	<0.01*
			HD	28	2 / 2			(HD-LD) 2.41	0.02*
		29Palms/2 "Live-Fire"	C	84	1 / 1	2.77	0.25	(LD-C) 0.96	0.34
			LD	54	2 / 1			(HD-C) 1.57	0.12
			HD	27	2 / 1			(HD-LD) 0.82	0.41

N.4.12.1.2.1 Final Fatigue Scores (by Integration Level)

Post-trial fatigue results by Integration Level proved to be mixed. For the Non-Live Fire scenario trials, the males' less-fatigued responses in the High-Density group were significantly differed from the Control group (p<0.01), but not the Low-Density group (p=0.96); there was also a significant difference between the distribution of responses for between the Low-Density and High-Density groups (p=0.02). We found no significant difference in males' responses found between any of the groups for the Live-Fire scenario trials.

This is consistent with the difference in degrees of compensation permitted by the two scenarios. For example, most tasks in the Non-Live Fire scenario were conducted by the entire crew (allowing many opportunities for assistance to occur), whereas the Live-Fire scenario restricted the conduct of most tasks to one (or perhaps two) crewmembers, without allowing for much assistance between volunteers.

As we saw with previous results for this MOS, males final fatigue scores remained heavily clustered in the lowest two levels: "1-Fully Alert" and "2-Very Alert", regardless of Integration Level. This was especially so for the Live-Fire scenario (Table 99). When participating in integrated crews, we see more males reporting the mid-range scores of "3-Okay" and sometimes "4-A Little Tired", which suggests some males may have been

[33] The Vehicle Commander billet in all groups was always filled by a male 0313; only the Driver and Gunner billets could be filled by either sex.

working a little bit harder in the integrated groups, especially when conducting maintenance tasks.

Table 99 – Males' Final Fatigue Score Distributions (by Integration Level) - 0313

Distribution of 0313 Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0313	29Palms/1 Post-"Maint"	C	87	41.4%	32.2%	20.7%	5.7%	0.0%	0.0%	0.0%
		LD	56	44.6%	23.2%	26.8%	5.4%	0.0%	0.0%	0.0%
		HD	28	14.3%	42.9%	21.4%	21.4%	0.0%	0.0%	0.0%
	29Palms/2 Post-"Live-Fire"	C	84	53.6%	34.5%	8.3%	2.4%	1.2%	0.0%	0.0%
		LD	54	46.3%	35.2%	18.5%	0.0%	0.0%	0.0%	0.0%
		HD	27	40.7%	29.6%	29.6%	0.0%	0.0%	0.0%	0.0%

N.4.12.2 Workload Results

For the 0313 volunteers at Twentynine Palms, we have maximum workload self-reports from 14 males and seven females. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.12.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females. Results are summarized in the table below (Table 100).

Table 100 - Summary of Max Workload Scores (by Gender) – 0313

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
0313	Max Workload for Trial	29Palms/1 "Non-Live Fire"	M	171	3 / 3	0.01	0.92
			F	84	3 / 3		
		29Palms/2 "Live-Fire"	M	169	3 / 3	2.56	0.44
			F	83	3 / 3		

We did not a find a significant difference between female and male 0313 volunteers' maximum workload scores for either day in their run-cycle (i.e., Non-Live Fire and Live-

Fire). For both groups, we saw max workload scores heavily clustered at "3-Active but Easy", with a secondary peak at "4-Challenging".

Table 101 - Max Workload Score Distributions (by Gender) - 0313

Distribution of 0313 Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
0313	29Palms/1 "Non-Live Fire"	M	171	0.0%	3.5%	63.2%	32.2%	1.2%	0.0%	0.0%
		F	84	0.0%	3.6%	66.7%	20.2%	9.5%	0.0%	0.0%
	29Palms/2 "Live-Fire"	M	169	0.0%	14.8%	57.4%	27.2%	0.6%	0.0%	0.0%
		F	83	0.0%	7.2%	60.2%	26.5%	6.0%	0.0%	0.0%

We saw no reports—from either gender—at the highest workload levels (i.e., "6-Overloaded" and "7-Unmanageable"). This suggests all Marines felt capable of accomplishing (if not always easily) the workload assigned them.

N.4.12.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males' responses remained consistent at different integration levels. For 0811s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) LAV crews. We have max workload self-reports from 14 males (14 males in the Control and Low-Density groups, and five males in the High-Density group); statistical results are summarized in the table below (Table 102).

Table 102 - Summary of Males' Max Workload Scores (by Integration Level) – 0313

MOS	Workload	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0313	Max Workload for Trial	29Palms/1 "Non-Live Fire"	C	87	3 / 3	10.7	<0.01*	(LD-C) 2.3	0.02*
			LD	56	3 / 3			(HD-C) 3.1	<0.01*
			HD	28	3.5 / 3			(HD-LD) -0.12	0.91
		29Palms/2 "Live-Fire"	C	87	3 / 3	1.8	0.40		
			LD	54	3 / 3				
			HD	28	3 / 3				

We found mixed results when we looked at males' max workload responses by Integration Level. For Non-Live Fire trials, we saw a significant difference between

Control and Integrated groups, but none at all for Live-Fire trials. When we look at the scores males reported between (Table 103), we see that males' in integrated groups (for Non-Live Fire trials) show a shift towards higher maximum workloads—in particular a shift upward from “3-Active by Easy” to “4-Challenging”. A similar pattern appears like it might be emerging for Live-Fire trials—though it's less pronounced, as well as countering, and (very) slight, rise in scores at “2-Little to Do” such that the pattern is not significant.

Table 103 - Males' Max Workload Score Distributions (by Integration Level) – 0313

Distribution of 0313 Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
0313	29Palms/1 "Non-Live Fire"	C	87	0.0%	3.4%	74.7%	21.8%	0.0%	0.0%	0.0%
		LD	28	0.0%	5.4%	51.8%	42.9%	0.0%	0.0%	0.0%
		HD	28	0.0%	0.0%	50.0%	42.9%	7.1%	0.0%	0.0%
	29Palms/2 "Live-Fire"	C	0	0.0%	13.8%	65.5%	19.5%	1.1%	0.0%	0.0%
		LD	87	0.0%	14.8%	50.0%	35.2%	0.0%	0.0%	0.0%
		HD	28	0.0%	17.9%	46.4%	35.7%	0.0%	0.0%	0.0%

As for possible mechanisms, it may be that males' in integrated groups are required to take on a larger workload during the Non-Live Fire scenario, at least one task (if not more). The lack of a difference for Live-Fire trials may be a function of the difference in degrees of compensation permitted by the two scenarios. For example, most tasks in the Non-Live Fire scenario were conducted by the entire crew (allowing many opportunities for assistance to occur), whereas the Live-Fire scenario restricted the conduct of most tasks to one (or perhaps two) crewmembers, without allowing for much assistance between volunteers.

N.4.12.3 Cohesion Results

For the 0313 volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) from 14 males and seven females. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating “Strongly Disagree” and nine indicating “Strongly Agree”).

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—

then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.12.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the table below (Table 104).

Table 104 - Summary of Unit Cohesion Scores (by Gender) - 0313

MOS	Location	Cohesion	Gender	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
0313	29Palms	Q1: Group was united in trying to reach goals	M	164	8.04	1.45	7.74	<0.01*
			F	81	8.57	1.01		
		Q2: We all take responsibility for task success of the group	M	164	8.12	1.33	6.21	0.01*
			F	81	8.47	1.25		
		Q3: Group members have similar aspirations for success	M	164	8.18	1.32	7.88	<0.01*
			F	81	8.62	0.86		
		Q4: If members had problems with a task, all wanted to help	M	164	8.15	1.35	3.35	0.07*
			F	81	8.49	1.06		
		Q5: Members communicated freely about responsibilities	M	164	8.10	1.38	4.47	0.03*
			F	81	8.42	1.18		
		Composite Score (sum of Q1-Q5)	M	164	40.60	6.55	3.84	0.05*
			F	81	42.57	5.12		

We found a significant difference between 0313 males and females for all five cohesion statements ("Q1-Unity", "Q2-Responsibility", "Q3-Aspirations", "Q4-Help" and "Q5-Comm"), as well as for the overall composite cohesion score. When we look more closely at the distribution of responses for each statement, we see that females tended to a bit more optimistic (i.e., higher scores) than males.

For all five questions, female responses peak most sharply at "9" (70.4% to 77.8%)—compared to what we see among males (56.7% to 61.6%)—and then taper off sharply with virtually no responses below "5". Males, on the other hand, also peak at "9" (but not as high) before tapering off less sharply unit popping up again in a small, secondary peak (10.4% to 12.2%) at level "5". We see virtually no responses—for either gender—below "5".

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement (Figure 20). Consistent with what we saw in the distributions of responses, we see that females tend to be more optimistic about their units' cohesion than males—across the board—although both are strongly positive.

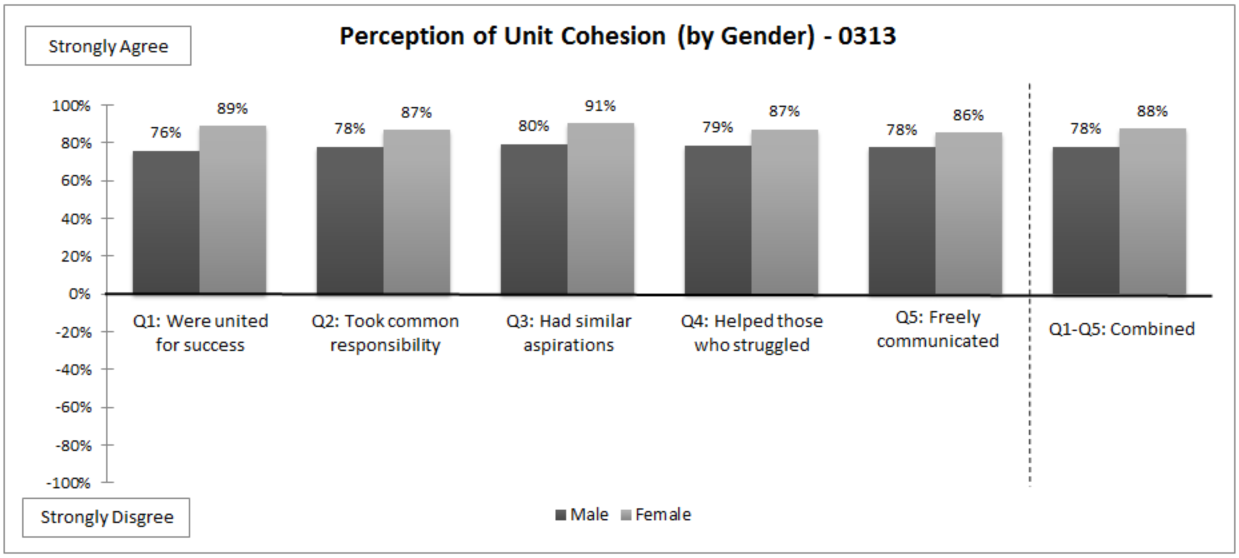


Figure 20 - Perception of Unit Cohesion (by Gender) – 0313

N.4.12.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For 0313s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) LAV crews. We have unit cohesion survey results (five statements per survey) from 14 males (14 males in the Control and Low-Density groups, and five males in the High-Density group); statistical results are summarized in the table below (Table 105).

Table 105 - Summary of Males' Unit Cohesion Scores (by Integration) - 0313

MOS	Cohesion	IL	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
0313	Q1: Group was united in trying to reach goals	C	83	8.35	1.31	13.97	<0.01*	(LD-C) -1.66	0.10
		LD	54	7.93	1.48			(HD-C) -3.80	<0.01*
		HD	27	7.33	1.54			(HD-LD) -1.99	0.05
	Q2: We all take responsibility for task success of the group	C	83	8.43	1.05	10.68	<0.01*	(LD-C) -1.53	0.13
		LD	54	7.98	1.47			(HD-C) -3.30	<0.01*
		HD	27	7.44	1.58			(HD-LD) -1.71	0.09
	Q3: Group members have similar aspirations for success	C	83	8.43	1.15	10.96	<0.01*	(LD-C) -1.82	0.07
		LD	54	8.09	1.35			(HD-C) -3.24	<0.01*
		HD	27	7.59	1.55			(HD-LD) -1.63	0.10
	Q4: If members had problems with a task, all wanted to help	C	83	8.41	1.09	6.84	0.03*	(LD-C) -0.80	0.42
		LD	54	8.07	1.48			(HD-C) -2.68	<0.01*
		HD	27	7.52	1.63			(HD-LD) -1.65	0.10
	Q5: Members communicated freely about responsibilities	C	83	8.40	1.11	8.39	0.02*	(LD-C) -1.68	0.09
		LD	54	7.96	1.47			(HD-C) -2.79	<0.01*
		HD	27	7.44	1.69			(HD-LD) -1.34	0.18
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	83	42.02	5.42	11.06	<0.01*	(LD-C) -1.77	0.08
		LD	54	40.04	6.98			(HD-C) -3.30	<0.01*
		HD	27	37.33	7.68			(HD-LD) -1.65	0.10

We found a statistical difference between the Control group and the High-Density group, but not between the other two pairings, for all five cohesion statements, as well as for the overall composite cohesion score.

When we looked at the distributions of each question, we see very similar patterns—the majority of responses are heavily skewed to the far right (i.e., “9”, at Significantly Agree), and taper off abruptly, with a broad, shallow span (of different thicknesses) in the middle range (i.e., “7” to “5”). The Control group has relatively low levels of responses in the middle range (1% to 7%), which get progressively more pronounced with the density level (i.e., 4%-17% for Low-Density groups, and 7% to 26% for High-Density groups).

We also graphed the means for each statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each question's statement (Figure 21). The average values (by integration level) shown are consistent with what we saw with regards to males' relatively high perceptions of unit cohesion within the Control group, along with the easy decline in their perception of cohesion when in the Low- and High-Density groups.

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group. The number of responses per Integration Level is not ideal (27 and 54, for High- and Low-Density groups, respectively). For that reason, we should be a bit cautious in extrapolating these results to the larger 0313 community.

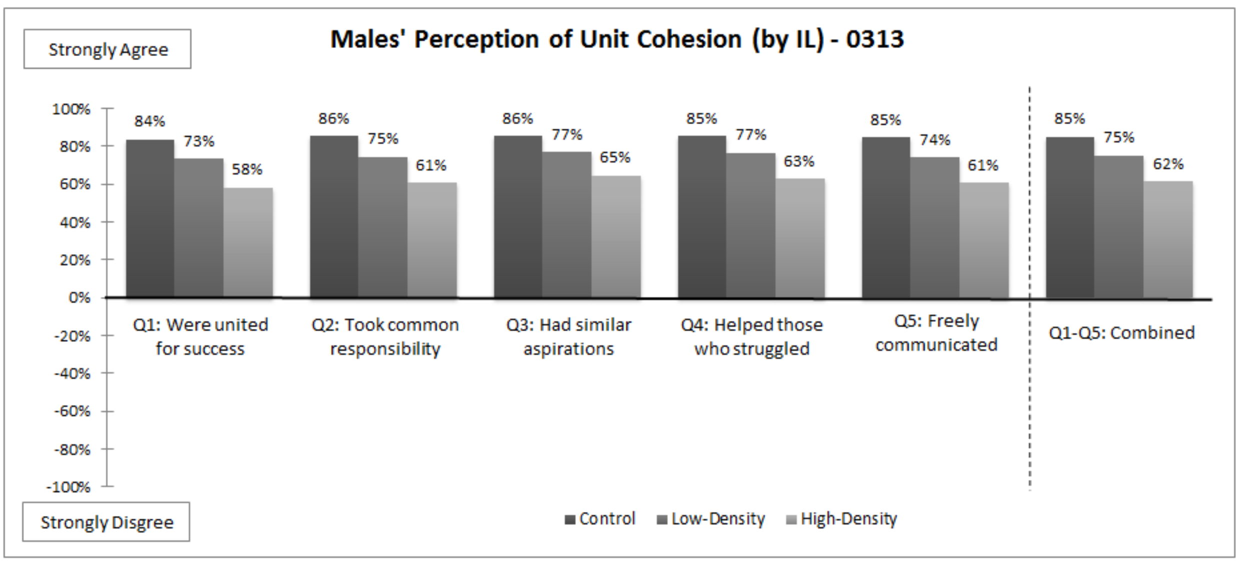


Figure 21 - Males' Perception of Unit Cohesion (by IL) – 0313

We also can't rule it out to potential for differences in experience to play a role in the reduced cohesion perceptions by males in integrated groups. Incidentally, the average values for the integrated groups appear to be quite different from how females perceived unit cohesion.

N.4.13 M1A1 Tank Crewman (MOS 1812)

N.4.13.1 Fatigue Results

For the 1812 volunteers at Twentynine Palms, we have baseline and final fatigue self-reports from 16 males and three females. Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-

Wallis (rank sum) test to check for significance between factors, using the criteria of p ≤ 0.10 for statistical significance (*).

N.4.13.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial) and final (post-trial, for both days in the run-cycle) fatigue levels. Results are summarized in the table below (Table 106).

Table 106 - Summary of Pre-Trial, Mid-Trial, and Final Fatigue Scores (by Gender) - 1812

MOS	Fatigue	Location / Day[34]	Gender	N	Median / Mode	χ2 Statistic	χ2 p-Value
1812	Baseline (Pre-Trial)	29Palms/2,3	M	356	2 / 2	0.24	0.62
			F	72	2 / 2		
	Final (Post-Trial)	29Palms/2 "Non-Live Fire"	M	178	2 / 2	0.28	0.60
			F	35	2 / 2		
		29Palms/3 "Live-Fire"	M	180	2 / 2	0.01	0.91
			F	36	2 / 2		

N.4.13.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We did not find a significant difference between female and male 1812 volunteers' pre-trial fatigue scores—a result consistent the strong peaks at "2-Very Lively" for both genders (Table 107), with nearly all responses at level "3-Okay" or below (94.4% for males, 91.7% for females).

We found only a very small fraction of responses—only 1.1 percent for males (driven largely by one male) and 2.8 percent for females (a single response each by two Marines)—at the "5-Moderately Tired" level, with none at "6-Extremely Tired" or "7-Exhausted". Combined with the lack of any obvious pattern of increasing scores over time suggests that 1812 volunteers had sufficient recovery time between trials.

N.4.13.1.1.2 Final Fatigue Scores

In looking at the volunteers' post-trial responses, we found no significant differences for either the Non-Live Fire scenario trials or the Live-Fire scenario trials. Both males and females showed the same strong peak at "2-Very Lively", with most other responses falling at "1-Fully Alert" or "3-Okay", for both days.

[34] The 1812 trials at Twentynine Palms shared space at Range 500 with 0313 (LAV) and 1833 (AAV) trials. As a results, the three MOS operated on a three-day cycle, where each individual MOS had one non-live fire day, one live-fire day, and one day devoted to general maintenance on the vehicles (non-trial). For 1812s, their trial days were Day 2 (non-live fire) and Day 3 (live-fire) of this cycle.

Table 107 - Pre-Trial, Mid-Trial, and Final Fatigue Score Distributions (by Gender) - 0812

Distribution of 1812 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
1812	29Palms/2&3 Pre-Trial	M	356	25.6%	53.4%	15.4%	4.5%	1.1%	0.0%	0.0%
		F	72	29.2%	41.7%	20.8%	5.6%	2.8%	0.0%	0.0%
	29Palms/2 Post-"Non-Live Fire"	M	178	21.3%	53.4%	16.3%	8.4%	0.6%	0.0%	0.0%
		F	35	17.1%	68.6%	8.6%	2.9%	2.9%	0.0%	0.0%
	29Palms/3 Post-"Live-Fire"	M	180	32.2%	53.3%	12.8%	1.1%	0.6%	0.0%	0.0%
		F	36	33.3%	50.0%	11.1%	5.6%	0.0%	0.0%	0.0%

N.4.13.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 1812s, we compared males' responses in Control (i.e., all-male) and Low-Density (i.e., 1 female) Tank crews, where females might be in one of two billets (Driver or Loader)[35]. We have final fatigue self-reports from 16 males; statistical results are summarized in the table below (Table 108).

Table 108 – Summary of Males' Mid-Trial and Final Fatigue Scores (by Integration Level) - 1812

MOS	Fatigue	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
1812	Final (Post-Trial)	29Palms/2 "Non-Live Fire"	C	108	2 / 2	0.02	0.89	(LD-C) -0.14	0.89
			LD	70	2 / 2				
		29Palms/3 "Live-Fire"	C	108	2 / 2	0.27	0.60	(LD-C) 0.52	0.61
			LD	72	2 / 2				

N.4.13.1.2.1 Final Fatigue Scores (by Integration Level)

Post-trial fatigue results by Integration Level were not significantly different for either trial scenario. In keeping with the distribution of fatigue scores seen in the "by gender" results above, males largely reported fatigue levels at "2-Very Lively" with clustering just above and below (i.e., "1-Fully Alert" and "3-Okay") at the end of each day's trial.

The 1812 Scheme of Maneuver (SoM) incorporated a combination of both crew-level (e.g., Remove/Replace Track Section, Crew Evacuation, Evacuate Wounded Crewman, Conduct Vehicle Recovery, Conduct Ammo Resupply, and Transfer Ammunition) and

[35] The third billet in the three-man crews—Tank Gunner—could only be filled by a Marine experienced in the MOS; as such, this billet was filled by all-male volunteers only.

individual-level tasks (Arm Main Gun, Manipulate Turret, Prep/Reload Commander's Weapon Station, and Engage Defensive/Offensive Targets[36]) (see Annex I).

Crew-level tasks could (notionally) have provided opportunities for males to compensate for females in integrated crews (e.g., share, or swap duties)— making them more tired—whereas individual-level tasks prohibit such compensation, but might have revealed differences by gender if females were taxed more heavily by specific tasks (perhaps allowing additional recovery for non-participants). We found no indication of either condition.

While the mixed nature of the 1812 SoM—with interspersed crew- and individual-level tasks—might have masked some low-level compensation or task-specific fatigue effects for this MOS, the fact that both male and female Marines persistently reported lower fatigue levels ("3-Okay" or below), such effects—if they exist—are likely short-lived and minor.

Table 109 – Males' Final Fatigue Score Distributions (by Integration Level) - 1812

Distribution of 1812 Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	**Location/Day**	**IL**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1812	29Palms/2 Post-"Non-Live Fire"	C	108	22.2%	50.9%	17.6%	9.3%	0.0%	0.0%	0.0%
		LD	70	20.0%	57.1%	14.3%	7.1%	1.4%	0.0%	0.0%
	29Palms/3 Post-"Live-Fire"	C	108	35.2%	49.1%	13.9%	0.9%	0.9%	0.0%	0.0%
		LD	72	27.8%	59.7%	11.1%	1.4%	0.0%	0.0%	0.0%

N.4.13.2 Workload Results

For the 1812 volunteers at Twentynine Palms, we have max workload self-reports from 16 males and three females. Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*).

N.4.13.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females. Results are summarized in the table below (Table 110).

[36] While the entire crew is involved in engaging targets, the bulk of the physical activity falls on the Loader.

Table 110 - Summary of Max Workload Scores (by Gender) – 1812

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
1812	Max Workload for Trial	29Palms/2 "Non-Live Fire"	M	180	3 / 3	9.3	<0.01*
			F	36	4 / 4		
		29Palms/3 "Live-Fire"	M	178	3 / 3	11.2	<0.01*
			F	35	3 / 3		

We found a significant difference between female and male 1812 volunteers' maximum workload scores for both days in their run-cycle (i.e., Non-Live Fire and Live-Fire). Both groups strongly favored maximum workload responses at "3-Active but Easy" (73% for males, and 60.0% for females) for the Live-Fire scenario, but females were a bit more likely to have some responses at "4-Challenging" and less likely to have some at "2-Little to Do" or lower (Table 111).

Table 111 - Max Workload Score Distributions (by Gender) - 1812

Distribution of 1812 Workload Results (by Gender) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	Gender	N	1	2	3	4	5	6	7
1812	29Palms/2 "Non-Live Fire"	M	180	0.0%	1.1%	56.7%	35.6%	6.1%	0.6%	0.0%
		F	36	0.0%	0.0%	30.6%	55.6%	13.9%	0.0%	0.0%
	29Palms/3 "Live-Fire"	M	178	2.2%	7.3%	73.0%	16.9%	0.0%	0.0%	0.6%
		F	35	0.0%	0.0%	60.0%	37.1%	2.9%	0.0%	0.0%

Because there were a number tasks for the 1812 Scheme of Maneuver that were only done by particular billets—and because only males could be Tank Gunners—we also checked to see whether we still saw a significant difference between the two groups when we only compared Marines in the Tank Driver and Tank Loader positions—we did.

It's also possible that the difference between the two groups may be due to females having less time in the MOS; if so, we would expect the difference between the two groups to gradually disappear over time as they became more skilled. Its unlikely fatigue played a role, as we did not see any difference between males and females.

N.4.13.2.2 Workload Results by Integration Level

In addition to looking at how maximum workload self-reports compared between males and females, we also examined whether males' responses remained consistent at different integration levels. For 1812s, we compared males' responses in Control (i.e., all-male) and Low-Density (i.e., 1 females) Tank crews. We have max workload self-reports from 16 males; statistical results are summarized in the table below (Table 112).

Table 112 - Summary of Males' Max Workload Scores (by Integration Level) – 1812

MOS	Workload	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
1812	Max Workload for Trial	29Palms/2 "Non-Live Fire"	C	108	3 / 3	2.0	0.15
			LD	72	3 / 3		
		29Palms/3 "Live Fire"	C	107	3 / 3	4.0	0.05*
			LD	71	3 / 3		

We found mixed results when we looked at males' max workload responses by Integration Level. For Live-Fire trials, we saw a significant difference between Low-Density and Control groups, but not no difference for Non-Live Fire trials. These results are somewhat counter-intuitive, at first look. Of the two scenarios, the non-live fire scenario would seem the most conducive towards compensation in a crew's division of work.

However, when we look more closely at the scores reported (Table 113), we can see a small number of scores at "1-No Demands" among Control groups and a score at "7-Unmanageable" in the Low-Density group. Because the non-parametric test we used (Kruskal-Wallis) looks at the ranking order of groups, these scores at the extrema (and in opposite directions for the two groups)—while relatively tiny—could be driving the significant difference conclusion.

Table 113 - Males' Max Workload Score Distributions (by Integration Level) – 1812

Distribution of 1812 Workload Results (by Integration Level) Twentynine Palms				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
1812	29Palms/2 "Non-Live Fire"	C	108	0.0%	1.9%	60.2%	31.5%	5.6%	0.9%	0.0%
		LD	72	0.0%	0.0%	51.4%	41.7%	6.9%	0.0%	0.0%
	29Palms/3 "Live-Fire"	C	107	3.7%	7.5%	75.7%	13.1%	0.0%	0.0%	0.0%
		LD	71	0.0%	7.0%	69.0%	22.5%	0.0%	0.0%	1.4%

Indeed, if we were to exclude just the four responses found at “1-No Demands” for Live-Fire trials (contributed by four different Marines in the Tank Gunner or Tank Driver billets during the first week of the study), we find that there is no longer a significant difference between Control and Low-Density groups for either day ($\chi^2 = 2.3$, p=0.13).

This certainly makes any determination that a real difference exists in males’ workload perceptions, based upon Integration Level, highly suspect—especially when we did not see a difference during the Non-Live Fire scenario where compensation would be more plausible (based on the Scheme of Maneuver).

N.4.13.3 Cohesion Results

For the 1812 volunteers at Twentynine Palms, we have Unit Cohesion survey results (five statements per survey) from 16 males and three females. Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating “Strongly Disagree” and nine indicating “Strongly Agree”).

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of p ≤ 0.10 for statistical significance (*).

N.4.13.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit’s run-cycle. Statistical results are summarized in the table below (Table 114).

Table 114 - Summary of Unit Cohesion Scores (by Gender) - 1812

MOS	Location	Cohesion	Gender	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
1812	29Palms	Q1: Group was united in trying to reach goals	M	162	8.33	0.90	3.72	0.05*
			F	33	8.64	0.65		
		Q2: We all take responsibility for task success of the group	M	162	8.31	1.00	0.47	0.49
			F	33	8.52	0.67		
		Q3: Group members have similar aspirations for success	M	162	8.29	1.06	1.83	0.18
			F	33	8.61	0.61		
		Q4: If members had problems with a task, all wanted to help	M	162	8.31	0.89	3.34	0.07*
			F	33	8.58	0.79		
		Q5: Members communicated freely about responsibilities	M	162	8.35	0.89	0.89	0.35
			F	33	8.48	0.83		
		Composite Score (sum of Q1-Q5)	M	162	41.59	4.12	2.49	0.11
			F	33	42.82	3.02		

We found a significant difference between 1812 males and females for only two of the five cohesion statements ("Q1-Unity" and "Q4-Help"), and no difference for the overall composite cohesion score (although the latter is borderline). When we look more closely at the distribution of responses for each statement, we see that females tended to a bit more optimistic (i.e., higher scores) than males.

For the two questions where we see a significant difference, we see that females strongly and consistently report cohesion scores in the top two levels (90.0% and 87.9%, for "Q1-Unity" and "Q4-Help", respectively). Males' responses for these same two questions, on the other hand, are more broadly spread between "5" and "9", though they, too, favor the highest two levels (84.6% and 80.9%, for "Q1-Unity" and "Q4-Help", respectively).

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement (Figure 22). The average values (by gender) reflect a consistently high perception of unit cohesion across the board among the 1812 volunteers, despite a very small percentage of responses among males that drive their responses a hair more pessimistic for two questions: "Q1-Unity" and "Q4-Help".

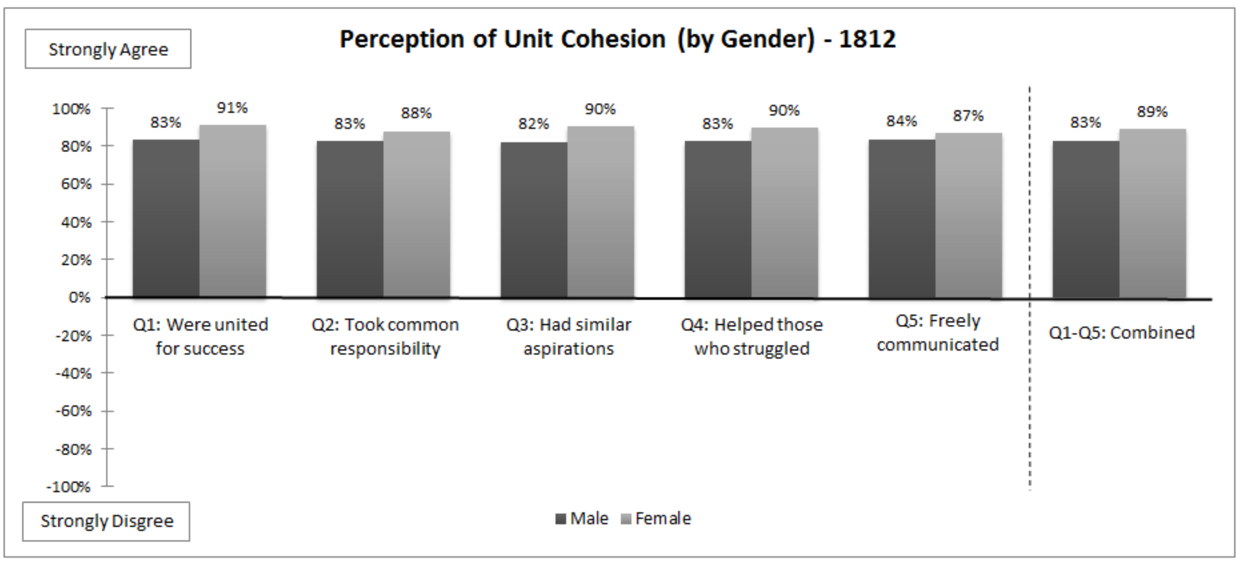


Figure 22 - Perception of Unit Cohesion (by Gender) – 1812

N.4.13.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For 1812s, we compared males' responses in Control (i.e., all-male) and Low-Density (i.e., 1 female) Tank crews. We have unit cohesion survey results (five statements per survey) from 16 males; statistical results are summarized in the table below (Table 115).

Table 115 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 1812

MOS	Cohesion	IL	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
1812	Q1: Group was united in trying to reach goals	C	96	8.53	0.71	10.78	<0.01*
		LD	66	8.03	1.05		
	Q2: We all take responsibility for task success of the group	C	96	8.50	0.83	9.15	<0.01*
		LD	66	8.05	1.16		
	Q3: Group members have similar aspirations for success	C	96	8.46	0.93	8.44	<0.01*
		LD	66	8.05	1.20		
	Q4: If members had problems with a task, all wanted to help	C	96	8.49	0.77	9.30	<0.01*
		LD	66	8.05	0.98		
	Q5: Members communicated freely about responsibilities	C	96	8.52	0.78	8.97	<0.01*
		LD	66	8.11	0.98		
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	96	42.50	3.45	11.04	<0.01*
		LD	66	40.27	4.65		

We found significant differences between the Control and both integrated groups (i.e., Low-Density and High-Density groups) for all five cohesion statements, as well as for the overall composite cohesion score.

When we looked at the distributions of each question, we see very similar patterns—the majority of responses are heavily skewed to the far right (i.e., “9”, at Significantly Agree) that tapers off over a span of several levels. For the Control group, we see a bit more representation at the peak (63% to 66%), with a taper that drops off very quickly at “7”. On the other hand, the Low-Density group has a smaller peak (41% to 44%), and it spans a wider stretch (tapers off at “5” or “6”.

We also graphed the means for each statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each question’s statement (Figure 23). The average values (by integration level) shown are consistent with what we saw with regards to males’ very higher perceptions of unit cohesion within the Control group, with fairly shallow dip to the Low-Density group’s average.

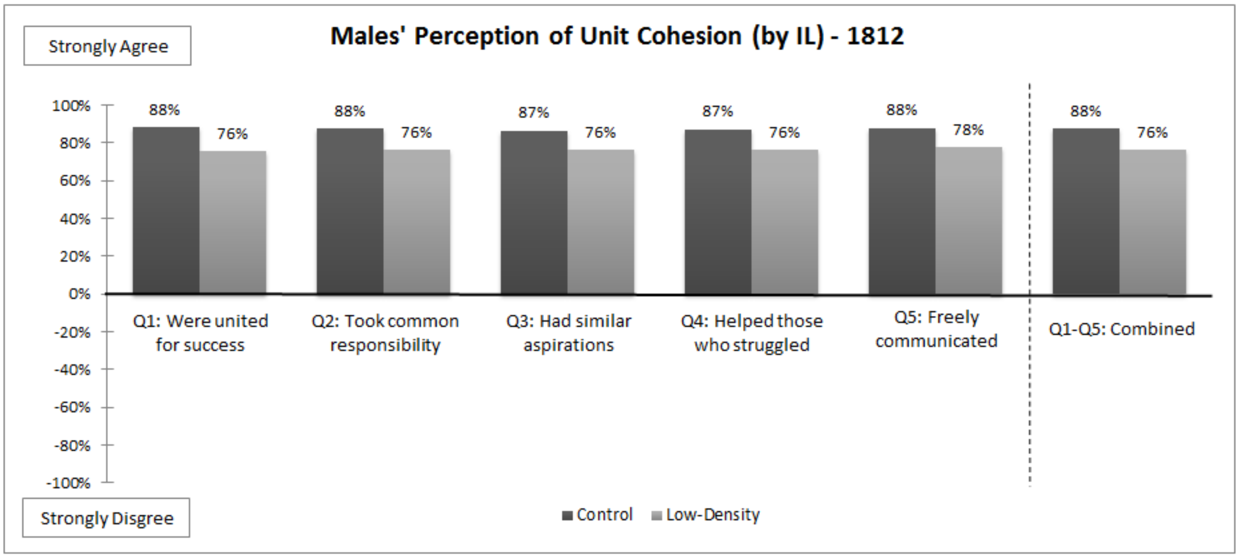


Figure 23 - Males' Perception of Unit Cohesion (by Integration Level) - 1812

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group, but the difference in experience between males and females may have been a factor. We should also keep in mind that these are the result of a relatively small volunteer pool (16 Marines) and their experiences with only three females, who may or may not be representative for the larger 1812 community.

Incidentally, these results differ a fair bit from how females rated their perception of unit cohesiveness in integrated groups—their results were more positive that males' responses in an integrated group. Plus, with the small unit size, some of the males' perceptions may have been colored by a different (vice a truly non-cohesive) group dynamic.

N.4.14 AAV Crewman (MOS 1833)

N.4.14.1 Fatigue Results

For the 1833 volunteers, we have baseline and final fatigue self-reports from 17 males and 10 females (for Twentynine Palms), and 16 males and 10 females (for Pendleton). Because fatigue survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.14.1.1 Fatigue Results by Gender

We first looked at how fatigue self-reports compared between males and females, including their baseline (pre-trial) and final (post-trial, for both days in the run-cycle at Twentynine Palms) fatigue levels. Statistical results are summarized in the table below (Table 116).

Table 116 - Summary of Pre-Trial and Final Fatigue Scores (by Gender) - 1833

MOS	Fatigue	Location/ Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
1833	Baseline (Pre-Trial)	29Palms/ 1&3	M	485	2 / 1	9.5	<0.01*
			F	268	2 / 1		
	Final (Post-Trial)	29Palms/1 "Live Fire"	M	238	2 / 2	0.71	0.40
			F	129	2 / 2		
		29Palms/3 "Non-Live Fire"	M	244	2 / 2	0.00	0.96
			F	134	2 / 2		
	Baseline (Pre-Trial)	Pendleton/1 "Amphib"	M	139	2 / 2	1.6	0.20
			F	77	2 / 2		
	Final (Post-Trial)	Pendleton/1 "Amphib"	M	139	2 / 2	5.6	0.02*
			F	77	2 / 2		

N.4.14.1.1.1 Pre-Trial Fatigue Scores (by Gender)

We found a significant difference between female and male 1833 volunteers' pre-trial fatigue scores for Twentynine Palms, but not at Pendleton where the amphibious portion was conducted. For the Twentynine Palms portions, females were slightly more likely to report pre-trial fatigue levels of "1-Fully Alert" or "Very Lively" (81%) compared to males (72.2%), and slightly less likely than males to report mid-range fatigue levels of "3-Okay" or "4-A Little Tired" (17.9% versus 22.5%).

We found very few of the higher fatigue scores for either gender (i.e., "5-Moderately Tired", "6-Extremely Tired", "7-Exhausted) at Twentynine Palms—only 5.4 percent for males (driven overwhelmingly by one Marine), and 1.1 percent for females (also driven largely by one Marine)—and no obvious pattern of increasing scores over time, suggesting volunteers largely had sufficient recovery time between trials.

At Pendleton, both genders favored pre-trial responses of "1-Fully Alert" to "3-Okay" (88.5% and 90.9%, for males and females respectively), with slightly higher (though still low) representation in the higher fatigue scores (10.1% and 10.4%)—both largely driven by the same two Marines driving the representation in this range for Twentynine Palms. The Pendleton results likewise suggest that most volunteers had sufficient recovery time between trials.

As to why we see a gender difference at Twentynine Palms but not at Pendleton, it may be that the longer duration—and more austere living conditions—presented by the Twentynine Palms' trials was a bit less conducive to recovery (or morale) for a sufficient sub-set of males to produce significance. When we look at who contributed to pre-trial scores of "4-A Little Tired" or higher (representing only 12% of male responses), we see that only eight of out of 17 Marines contributed responses in this range[37]. At the same time, all 17 male Marines did contribute at least some pre-trial scores to the lower range ("3-Okay" and below), which represent the majority of responses (88.0%).

Conversely, when we look at the same fatigue range-split for females, we see that eight of 10 females contributed to their "4-A Little Tired" and higher range (representing only 6 percent of responses), and all contributed to the lower "3-Okay" and below range—representing a slightly higher overall percentage than males (94.0%).

So, while we technically see a significant difference between males and females at Twentynine Palms, it appears that this may be due to periodic shifts to slightly higher fatigue levels by a sub-set of males—which could be influenced by a number of factors—rather than a substantial physiological difference.

N.4.14.1.1.2 Final Fatigue Scores (by Gender)

In looking at the volunteers' post-trial responses, we found mixed results (Table 116): we saw no significant difference post-trial between genders for either day at Twentynine Palms, but we did see a significant difference for Pendleton trials.

At Pendleton, males most often reported post-trial scores of "2-Very Lively" or "1-Fully Alert" (69.1%), with a respectable percentage at "3-Okay" (14.4%). By comparison, females show a slightly flatter, more dispersed distribution with "2-Very Lively" or "Okay" reported most often (59.7%), with respectable percentages split between "1-Fully Alert" (14.3%) and "4-A Little Tired (15.6%). Both groups had comparable (and relatively small) representation at "5-Moderately Tired" and "6-Extremely Tired" (10.1 percent for males, 10.4% for females)—typically single responses by a handful of Marines.

For Twentynine Palms, the distributions look very similar—clustering predominantly between "1-Fully Alert" and "3-Okay" with comparably small representation at "5-Moderately Tired" or "6-Extremely Tired"—but without the statistical difference.

As with the pre-trial results, we must ask why—when the gross shape of male versus female distribution curves appear so similar and at levels suggesting most Marines (most of the time) had energy remaining by the end of a trial—we see a significant

[37] When we look at the same split for females, we see that eight of 10 females contributed to the "4-Okay" and higher range (representing only 6 percent of responses), and

difference at one site but not the other, and whether that difference is operationally relevant and repeatable.

On the face of it, there were certainly a number of physically demanding elements in the Pendleton trials (e.g., lifting the dummy in the water CASEVAC, moving and securing the tie-down chains). However, it's hard to see these being that much harder than the physical elements found during the Twentynine Palms trials (e.g., lifting the dummy in the turret CASEVAC, moving ammo, breaking track, mounting the M2 and Mk19 weapon systems) to warrant Pendleton being significant where Twentynine Palms was not. Indeed, the gross shapes of the fatigue distributions for both locations are remarkably similar.

When we looked at the range of responses an individual Marine reported at Pendleton post-trial, we found that 80 percent of females, and 50 percent of males reported within a range covering three levels or more (e.g., 1 to 4, 2 to 4, 2 to 6)—despite conducting only the single scenario (i.e., amphibious ops). There was even greater variability at Twentynine Palms (90% of females and 65% of males had a response range of three or greater).

With this degree of variability within the same individuals from day-to-day—which may have been influenced by a number of factors including billet, weather, group dynamic, and morale—it's seems less likely that the statistically significant (but hardly operationally distinguishable) difference seen between genders at Pendleton was a real physiological difference, and more the result of relatively small number of subjects, a fewer number of trials, and their high inter-subject variability.

Table 117 - Pre-Trial and Final Fatigue Score Distributions (by Gender) - 1833

Distribution of 1833 Fatigue Results (by Gender) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1833	29Palms/1&3 Pre-Trial	M	485	39.4%	32.8%	15.9%	6.6%	4.9%	0.2%	0.2%
		F	268	48.5%	32.5%	13.1%	4.9%	1.1%	0.0%	0.0%
	29Palms/1 Post-"Live Fire"	M	238	28.6%	39.5%	19.7%	5.9%	6.3%	0.0%	0.0%
		F	129	25.6%	38.8%	20.2%	10.1%	3.9%	1.6%	0.0%
	29Palms/3 Post-"Non-Live Fire"	M	244	25.4%	37.3%	18.4%	11.1%	7.0%	0.4%	0.4%
		F	134	20.9%	40.3%	29.1%	5.2%	4.5%	0.0%	0.0%
	Pendleton/1 Pre-Trial	M	139	28.8%	47.5%	12.2%	4.3%	7.2%	0.0%	0.0%
		F	77	24.7%	39.0%	27.3%	5.2%	3.9%	0.0%	0.0%
	Pendleton/1 Post-"Amphib"	M	139	24.5%	44.6%	14.4%	6.5%	9.4%	0.7%	0.0%
		F	77	14.3%	39.0%	20.8%	15.6%	7.8%	2.6%	0.0%

N.4.14.1.2 Fatigue Results by Integration Level

Next, we looked at whether males fatigue responses remained consistent at different integration levels. For 1833s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) AAV crews. We have final fatigue self-reports from 17 males for Twentynine Palms, and 16 males for Pendleton; statistical results are summarized in the table below (Table 118).

Table 118 – Summary of Males' Final Fatigue Scores (by Integration Level) - 1833

MOS	Fatigue	Location / Day	IL	N	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
1833	Final (Post-Trial)	29Palms/1 "Live-Fire"	C	110	2 / 2	1.3	0.52	(LD-C) 0.05	0.96
			LD	84	2 / 2			(HD-C) -1.1	0.29
			HD	44	2 / 1			(HD-LD) -1.0	0.30
		29Palms/3 "Non-Live Fire"	C	114	2 / 2	1.0	0.59	(LD-C) -0.44	0.66
			LD	84	2 / 2			(HD-C) -1.0	0.30
			HD	46	2 / 2			(HD-LD) -0.59	0.55
	Final (Post-Trial)	Pendleton/1 "Amphib"	C	66	2 / 2	2.3	0.31	(LD-C) -0.27	0.79
			LD	46	2 / 2			(HD-C) 1.3	0.19
			HD	27	2 / 2,5			(HD-LD) 1.4	0.15

N.4.14.1.2.1 Final Fatigue Scores (by Integration Level)

For 1833s, we did not see any significant difference in post-trial fatigue results based on Integration Level for any of the three trial scenarios (Twentynine Palms' Live-Fire and Non-Live Fire, and Pendleton's Amphibious Ops). Across all integration levels and locations, we see males' responses clustering among the three lowest levels (ranging between 63.0% and 93.2%) and peaking at "2-Very Lively".

However, we do see some borderline statistical results for Pendleton in High-Density units, compared to Control and Low-Density crews. This is likely due to the presence of a second, substantial peak at "5-Moderately Tired" (29.6%). When we look more closely at these responses, we see that most (69%) were from two Marines that had reported the same level of fatigue in their pre-trial survey (i.e., "5-Moderately Tired")—i.e., they were just as tired after the trial as they had when they'd started.

In addition, as we examined their history of responses over the course of the study, we found that both persistently reported moderate-to-high levels ("4-Okay" to "7-Exhausted") of pre-trial fatigue at both Twentynine Palms and Pendleton (71.0%). At the same time, they also consistently reported (79% of the time) the exact same (or one higher) post-trial fatigue level, regardless of the integration level of their unit. This suggests that this second peak—and the resulting borderline statistical result—is not a function of Integration Level, but more a characteristic of two male Marines.

Table 119 – Males' Final Fatigue Score Distributions (by Integration Level) - 1833

Distribution of 1833 Fatigue Results (by Integration Level) Twentynine Palms				Fully Alert	Very Lively	Okay	A Little Tired	Moderately Tired	Extremely Tired	Exhausted
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
1833	29Palms/1 Post-"Live Fire"	C	110	25.5%	43.6%	18.2%	5.5%	7.3%	0.0%	0.0%
		LD	84	28.6%	36.9%	20.2%	7.1%	7.1%	0.0%	0.0%
		HD	44	36.4%	34.1%	22.7%	4.5%	2.3%	0.0%	0.0%
	29Palms/3 Post-"Non-Live Fire"	C	114	21.9%	40.4%	16.7%	13.2%	7.0%	0.9%	0.0%
		LD	84	27.4%	35.7%	16.7%	9.5%	9.5%	0.0%	1.2%
		HD	46	30.4%	32.6%	26.1%	8.7%	2.2%	0.0%	0.0%
	Pendleton/1 Post-"Amphib"	C	66	24.2%	47.0%	15.2%	7.6%	4.5%	1.5%	0.0%
		LD	48	22.9%	47.9%	16.7%	4.2%	4.2%	4.2%	0.0%
		HD	27	25.9%	29.6%	7.4%	7.4%	29.6%	0.0%	0.0%

N.4.14.2 Workload Results

For the 1833 volunteers, we have baseline and final fatigue self-reports from 17 males and 10 females (for Twentynine Palms), and 16 males and 10 females (for Pendleton). Because workload survey responses were ordinal in nature, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.14.2.1 Workload Results by Gender

We first looked at how maximum workload self-reports compared between males and females. Results are summarized in the table below (Table 120).

Table 120 - Summary of Max Workload Scores (by Gender) – 1833

MOS	Workload	Location / Day	Gender	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value
1833	Max Workload for Trial	29Palms/1 "Live Fire"	M	236	3 / 3	17.4	<0.01*
			F	130	3.5 / 4		
		29Palms/3 "Non-Live Fire"	M	244	3 / 3	2.2	0.13
			F	134	3 / 3		
		Pendleton/1 "Amphib"	M	139	3 / 3	1.1	0.30
			F	77	3 / 3		

We found mixed results between male and female 1833 volunteers in their maximum workload scores; the results were significant for the “Live-Fire” scenario, but not for the “Non-Live Fire” or “Amphibious” scenarios, though the “Non-Live Fire” results could be considered borderline.

For the non-significant scenarios, both genders were most likely to report a peak workload of “3-Active but Easy”; only for the “Live-Fire Scenario” were females slightly more likely to report their peak workload as “4-Challenging” (though “3-Active but Easy” still had substantial representation).

Distribution of 1833 Workload Results (by Gender) Twentynine Palms & Pendleton				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	**Location/Day**	**Gender**	**N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1833	29Palms/1 "Live-Fire"	M	236	3.0%	27.1%	43.6%	24.6%	1.7%	0.0%	0.0%
		F	130	4.6%	14.6%	30.8%	42.3%	6.2%	1.5%	0.0%
	29Palms/3 "Non-Live Fire"	M	244	1.2%	17.6%	43.4%	34.4%	2.9%	0.4%	0.0%
		F	134	0.7%	14.2%	43.3%	30.6%	8.2%	3.0%	0.0%
	Pendleton/1 "Amphib"	M	139	2.2%	8.6%	44.6%	42.4%	2.2%	0.0%	0.0%
		F	77	0.0%	10.4%	41.6%	36.4%	7.8%	2.6%	1.3%

We saw few of the highest max workload scores for either group (i.e., “6-Overloaded” or “7-Unmanageable”)—only one male response (for the Non-Live Fire scenario) and 2 to 4 percent of female responses (by only three Marines, across all three scenarios). This suggests most Marines felt capable of accomplishing (if not always easily) the workload assigned them.

As far as the statistical difference between males and females for the Live-Fire scenario, it’s unlikely that fatigue was a factor (since the only fatigue effect we saw, by gender, was found during the Amphibious scenario). However, it may be that there is a slight experience bias that is emphasized during the heightened tempo (or pressure) associated with live-fire effects. Without some additional fidelity as to what task or tasks elicited the slightly higher reporting rates, then it’s difficult to say. If experience is the driving factor in the difference, then we would expect this to decrease over time.

N.4.14.2.2 Workload Results by Integration Level

In addition to looking at how max workload self-reports compared between males and females, we also examined whether males’ responses remained consistent at different integration levels. For 1812s, we compared males’ responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) AAV crews. We have maximum workload self-reports from 17 males (17 males in the Control and Low-Density groups, and 16 males in the High-Density group) for Twentynine Palms and 16 males (16 males in the Control and Low-Density groups, and 13 males in the High-Density group) for Pendleton; statistical results are summarized in the table below (Table 121).

Table 121 - Summary of Males' Max Workload Scores (by Integration Level) – 1833

MOS	Workload	Location / Day	IL	N (scores)	Median / Mode	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
1833	Max Workload for Trial	29Palms/1 "Live Fire"	C	108	3 / 3	2.8	0.25		
			LD	84	3 / 3				
			HD	44	3 / 3				
		29Palms/3 "Non-Live Fire"	C	114	3 / 3	0.21	0.90		
			LD	84	3 / 3				
			HD	46	3 / 3				
		Pendleton/1 "Amphib"	C	66	3 / 3	4.5	0.11		
			LD	46	3 / 3				
			HD	27	4 / 4				

We did not see a significant difference between males' responses for any integration level, regardless of the scenario. For the Live-Fire trials, males heavily report maximum workload scores between "2-Little to Do" and "4-Challenging" (95.3% for Control, 96.4% for Low-Density, and 97.7% for High-Density).

On Non-Live Fire trials, we see clustering in this same general range—though with lower numbers at "2-Little to Do", which is consistent with the more physical (and less strictly delineated by billet) nature of the non-live fire scenario.

Finally, with the Pendleton Amphibious trials, we see even lower numbers of reports at "2-Little to Do" for all three groups, but the same heavy concentration at "3-Active but Easy" and "4-Challenging".

This suggests that there are no changes in either compensation or cooperation between Control and integrated groups—at least of a nature to shift males' perception of their max workload. However, the borderline result found for the Pendleton scenario—while not significant—might warrant additional study to see if the results remain constant when more trials are run.

Table 122 - Males' Max Workload Score Distributions (by Integration Level) – 1833

Distribution of 1833 Workload Results (by Integration Level) Twentynine Palms & Pendleton				No Demands	Little to Do	Active but Easy	Challenging	Extremely Busy	Overloaded	Unmanageable
MOS	Location/Day	IL	N	1	2	3	4	5	6	7
1833	29Palms/1 "Live-Fire"	C	106	0.9%	27.4%	39.6%	28.3%	3.8%	0.0%	0.0%
		LD	44	3.6%	22.6%	50.0%	23.8%	0.0%	0.0%	0.0%
		HD	44	2.3%	36.4%	43.2%	18.2%	0.0%	0.0%	0.0%
	29Palms/3 "Non-Live Fire"	C	0	2.6%	19.3%	36.0%	40.4%	1.8%	0.0%	0.0%
		LD	114	0.0%	15.5%	48.8%	32.1%	3.6%	0.0%	0.0%
		HD	46	0.0%	17.4%	52.2%	23.9%	4.3%	2.2%	0.0%
	Pendleton/1 "Amphib"	C	66	3.0%	10.6%	45.5%	40.9%	0.0%	0.0%	0.0%
		LD	0	2.2%	8.7%	47.8%	39.1%	2.2%	0.0%	0.0%
		HD	0	0.0%	3.7%	37.0%	51.9%	7.4%	0.0%	0.0%

N.4.14.3 Cohesion Results

For the 1833 volunteers, we have Unit Cohesion survey results (five statements per survey) from 17 males and 10 females (for Twentynine Palms), and 16 males and 10 females (for Pendleton). Cohesion responses for each statement were set on a nine-point, Likert scale (one indicating "Strongly Disagree" and nine indicating "Strongly Agree").

Because the cohesion responses consistently violated the normality and independence assumptions for parametric data, we restricted our statistical analysis to the non-parametric Kruskal-Wallis (rank sum) test to check for significance between factors, using the criteria of $p \leq 0.10$ for statistical significance (*). Where we had more than two groups to compare—and if the Kruskal-Wallis test showed significance for similarity—then we conducted Wilcoxon pairwise comparisons (using $p<0.033$ to adjust for multiple comparisons).

N.4.14.3.1 Cohesion Results by Gender

We first looked at how cohesion results—for each statement, and for a composite result (the sum of all five responses, which scaled between 5 and 45)—compared between males and females, following the conclusion of their unit's run-cycle. Statistical results are summarized in the tables below for Twentynine Palms (Table 123) and Pendleton (Table 124).

Table 123 - Summary of Unit Cohesion Scores (by Gender) - 1833 (29Palms)

MOS	Cohesion	Gender	N (scores)	Mean	SD	χ2 Statistic	χ2 p-Value
1833	Q1: Group was united in trying to reach goals	M	244	7.98	1.43	0.45	0.50
		F	134	7.97	1.26		
	Q2: We all take responsibility for task success of the group	M	244	7.86	1.55	1.68	0.20
		F	134	8.09	1.28		
	Q3: Group members have similar aspirations for success	M	244	7.90	1.37	0.36	0.55
		F	134	7.98	1.31		
	Q4: If members had problems with a task, all wanted to help	M	244	7.84	1.59	0.94	0.33
		F	134	8.04	1.31		
	Q5: Members communicated freely about responsibilities	M	244	7.94	1.47	0.58	0.45
		F	134	7.85	1.46		
	Composite Score (sum of Q1-Q5)	M	244	39.53	6.83	1.03	0.31
		F	134	39.93	6.20		

For Twentynine Palms, we found no significant difference in between 1833 males and females for all five cohesion statements, as well as no difference for their respective overall composite cohesion scores. For both genders, and across all five statements, we see a very strong preference for cohesion scores in the top two levels (68.9% to 74.6%), with a very small number of sporadic reports at all levels below “5”.

Table 124 - Summary of Unit Cohesion Scores (by Gender) - 1833 (Pendleton)

MOS	Cohesion	Gender	N	Mean	SD	χ2 Statistic	χ2 p-Value
1833	Q1: Group was united in trying to reach goals	M	139	8.20	1.40	4.46	0.03*
		F	77	7.88	1.48		
	Q2: We all take responsibility for task success of the group	M	139	8.19	1.45	6.39	0.01*
		F	77	7.82	1.48		
	Q3: Group members have similar aspirations for success	M	139	8.16	1.49	4.49	0.03*
		F	77	7.86	1.49		
	Q4: If members had problems with a task, all wanted to help	M	139	8.18	1.45	5.28	0.02*
		F	77	7.82	1.49		
	Q5: Members communicated freely about responsibilities	M	139	8.13	1.52	7.78	<0.01*
		F	77	7.71	1.57		
	Composite Score (sum of Q1-Q5)	M	139	40.86	7.11	4.16	0.04*
		F	77	39.09	7.18		

However, when we look at the results for the Pendleton phase of the study (i.e., the amphibious scenario), we see a different result entirely—we find a significant difference between all five cohesion statements (“Q1-United”, “Q2-Responsibility”, “Q3-Aspirations”, “Q4-Help”, and “Q5-Comm”), as well as for the overall composite cohesion score. This is rather unexpected, given that the two populations are almost entirely the same (one male from Twentynine Palms did not continue for the Pendleton phase).

When we look more closely at the distributions of each gender at Pendleton, we still see very strong clustering at the two highest levels—as we saw during Twentynine Palms (71.4% and 85.6%). Where the two phases differ is that females tend to be a bit more evenly split between “8” and “9” than before, and we see a slight bump in male reports of the lowest level (“1”)—all by the same male Marine—and in female reports at “2” and “3” (also by a single female Marine). Based on these revelations, it does not appear that the significant differences we see for the Pendleton phase are necessarily practical differences in the general perception of unit cohesion for 1833s.

We also graphed the means for each survey statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each survey statement, for both Twentynine Palms (Figure 24) and Pendleton (Figure 25).

The average values (by gender)—for both sites—reflect a generally high perception of unit cohesion across the majority of 1833 volunteers, despite a very small percentage of responses by two Marines. The only practical difference between the two sites—one showing no significance (Twentynine Palms) and one that does (Pendleton)—is a slight dip in females’ perception of unit cohesion.

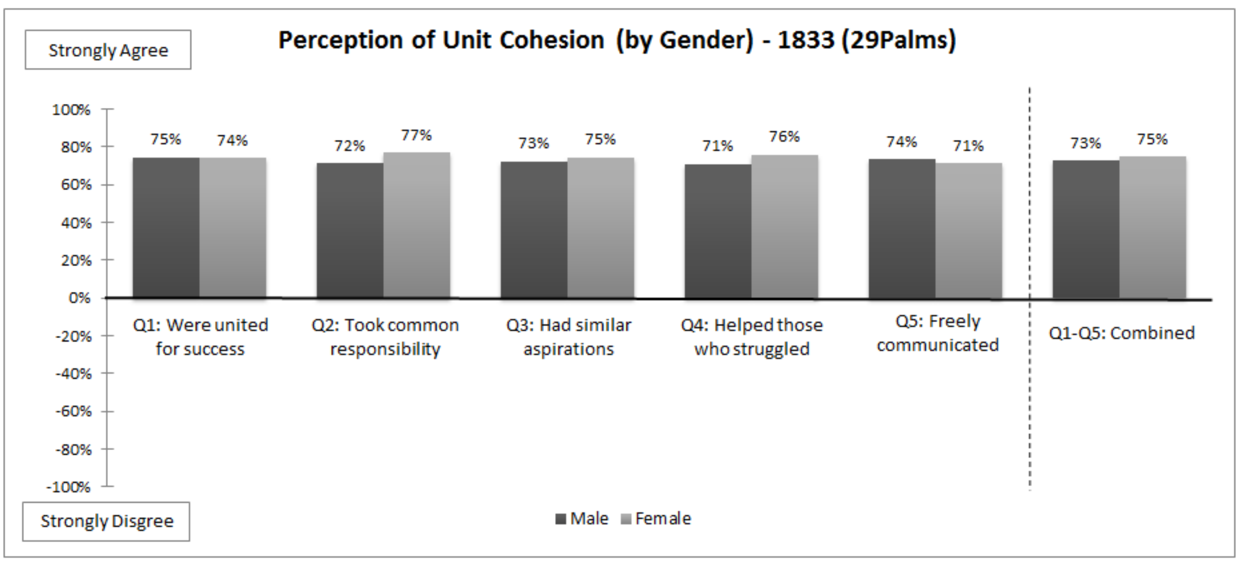


Figure 24 - Perception of Unit Cohesion (by Gender) – 1833 (Twentynine Palms)

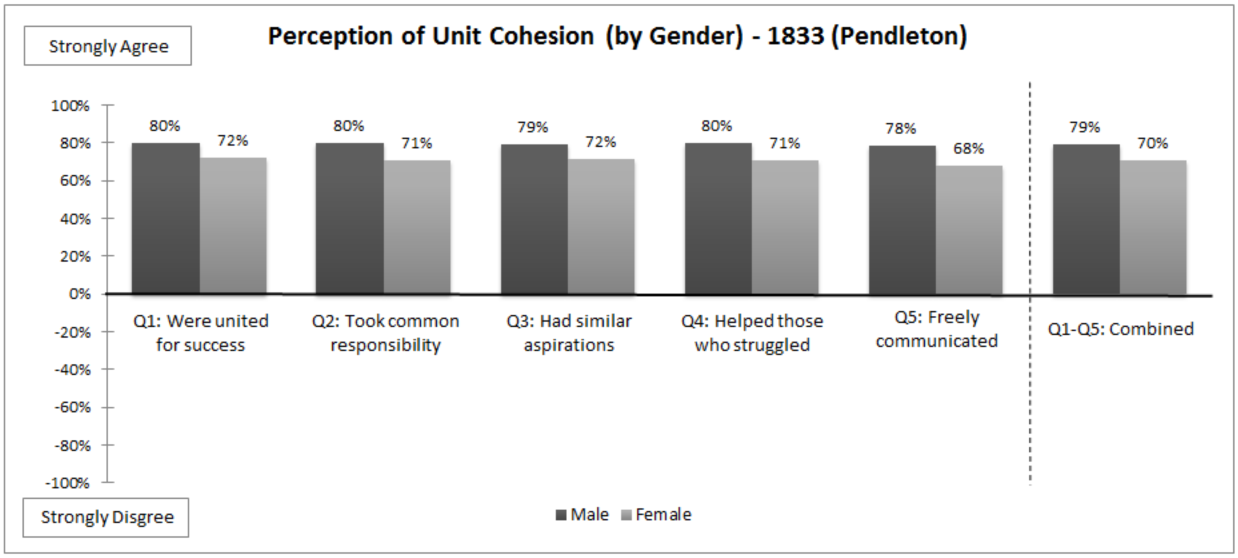


Figure 25 - Perception of Unit Cohesion (by Gender) – 1833 (Pendleton)

N.4.14.3.2 Cohesion Results by Integration Level

In addition to looking at how unit cohesion scores compared between males and females, we also examined whether males' cohesion responses remained consistent at different integration levels. For 1833s, we compared males' responses in Control (i.e., all-male), Low-Density (i.e., 1 female), and High-Density (i.e., 2 females) AAV crews. We have unit cohesion survey results (five statements per survey) from 17 males (17 males in the Control and Low-Density groups, and 16 males in the High-Density group) for Twentynine Palms (Table 125) and 16 males (16 males in the Control and Low-Density groups, and 13 males in the High-Density group) for Pendleton (Table 126).

Table 125 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 1833 (29Palms)

MOS	Cohesion	IL	N	Mean	SD	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
1833	Q1: Group was united in trying to reach goals	C	114	8.39	0.83	12.68	<0.01*	(LD-C) -2.56	0.01*
		LD	84	7.76	1.56			(HD-C) -3.27	<0.01*
		HD	46	7.37	1.97			(HD-LD) -1.03	0.30
	Q2: We all take responsibility for task success of the group	C	114	8.30	0.94	12.97	<0.01*	(LD-C) -2.86	<0.01*
		LD	84	7.58	1.78			(HD-C) -3.10	<0.01*
		HD	46	7.30	2.00			(HD-LD) -0.75	0.45
	Q3: Group members have similar aspirations for success	C	114	8.23	0.96	8.54	0.01*	(LD-C) -2.64	<0.01*
		LD	84	7.69	1.39			(HD-C) -2.14	0.03
		HD	46	7.48	1.93			(HD-LD) -0.07	0.95
	Q4: If members had problems with a task, all wanted to help	C	114	8.16	1.22	5.70	0.06*	(LD-C) -2.05	0.04
		LD	84	7.55	1.86			(HD-C) -1.90	0.06
		HD	46	7.61	1.78			(HD-LD) -0.06	0.95
	Q5: Members communicated freely about responsibilities	C	114	8.36	0.93	15.57	<0.01*	(LD-C) -3.14	<0.01*
		LD	84	7.67	1.65			(HD-C) -2.43	<0.01*
		HD	46	7.41	1.89			(HD-LD) -0.71	0.48
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	114	41.43	4.28	11.73	<0.01*	(LD-C) -2.52	0.01*
		LD	84	38.25	7.57			(HD-C) -3.12	<0.01*
		HD	46	37.17	9.09			(HD-LD) -0.89	0.37

For Twentynine Palms, we found a statistical difference between the Control group and both integration levels (i.e., High-Density, and Low-Density groups) for all five cohesion statements, as well as for the overall composite cohesion score. However, we did not see a difference between the two integration levels.

When we looked at the distributions of each question, we see very similar patterns—the majority of responses are heavily skewed to the far right (i.e., “9”, at Significantly Agree), and taper off relatively quickly to the left. Where the differences occur is in the sharpness of the main peak at “9” and the span of the taper. For Control groups, we see the sharpest peak (52% to 60% across all five questions) that tapers off, predominantly, by “7” with virtually no responses beyond that.

For the two integrated groups, the peaks are a bit more shallow (35% to 45%), and the taper spans a wider (and more negative) set of responses—mostly by “5”, but there are sporadic reports of scores at the lowest level “1” (significantly disagree).

Table 126 - Summary of Males' Unit Cohesion Scores (by Integration Level) - 1833 (Pendleton)

MOS	Cohesion	IL	N	Mean	SD	χ2 Statistic	χ2 p-Value	Z-Test Statistic	Z-Test p-Value
1833	Q1: Group was united in trying to reach goals	C	66	8.62	0.58	12.65	<0.01*	(LD-C) -2.04	0.04
		LD	46	8.28	0.89			(HD-C) 11.81	<0.01*
		HD	27	7.04	2.53			(HD-LD) -1.91	0.06
	Q2: We all take responsibility for task success of the group	C	66	8.59	0.68	11.81	<0.01*	(LD-C) -2.00	0.05
		LD	46	8.30	0.84			(HD-C) -3.25	<0.01*
		HD	27	7.00	2.62			(HD-LD) -1.83	0.07
	Q3: Group members have similar aspirations for success	C	66	8.65	0.57	19.52	<0.01*	(LD-C) -2.11	0.03*
		LD	46	8.33	0.84			(HD-C) -4.23	<0.01*
		HD	27	6.67	2.60			(HD-LD) -2.76	<0.01*
	Q4: If members had problems with a task, all wanted to help	C	66	8.68	0.50	21.16	<0.01*	(LD-C) -2.62	<0.01*
		LD	46	8.30	0.81			(HD-C) -4.32	<0.01*
		HD	27	6.74	2.55			(HD-LD) -2.62	<0.01*
	Q5: Members communicated freely about responsibilities	C	66	8.64	0.60	19.43	<0.01*	(LD-C) -2.25	0.02*
		LD	46	8.28	0.89			(HD-C) -4.21	<0.01*
		HD	27	6.63	2.63			(HD-LD) -2.65	<0.01*
	Composite Score (sum of Q1-Q5) Lowest: 5 | Highest 45	C	66	43.18	2.75	16.18	<0.01*	(LD-C) -2.35	0.02*
		LD	46	41.50	4.08			(HD-C) -3.78	<0.01*
		HD	27	34.07	12.59			(HD-LD) -2.19	0.03*

When we looked the Pendleton results, we do find a statistical difference between the Control group and the High-Density; for the Low-Density group, we see mixed results against both the Control group and the High-Density group.

For this location, the pattern is almost identical to what we saw for Twentynine Palms (i.e., peaks for all groups at "9", with the absolute percentage dropping with integration level and in the span of the taper. For the High-Density group, we also see some indications of a secondary peak at the neutral level, "5", with some sporadic reports at the lowest level, "1" (representing 11% of all High-Density responses).

We also graphed the means for each statement as a percentage of how strongly the volunteers agreed (positive) or disagreed (negative) with each question's statement, for both Twentynine Palms (Figure 26) and Pendleton (Figure 27). The average values (by integration level) shown are consistent with what we saw with regards to males' relatively high perceptions of unit cohesion within the Control group, along with the easy decline in their perception of cohesion when in the Low- and High-Density groups.

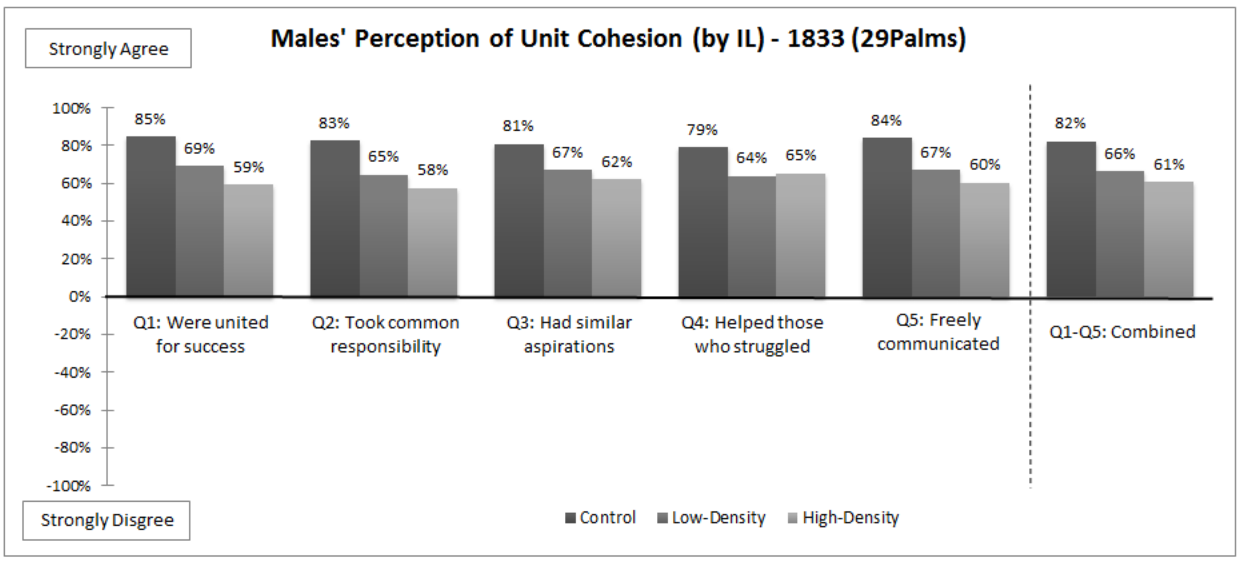


Figure 26 - Males' Perceptions of Unit Cohesion (by Integration Level) - 1833 (29Palms)

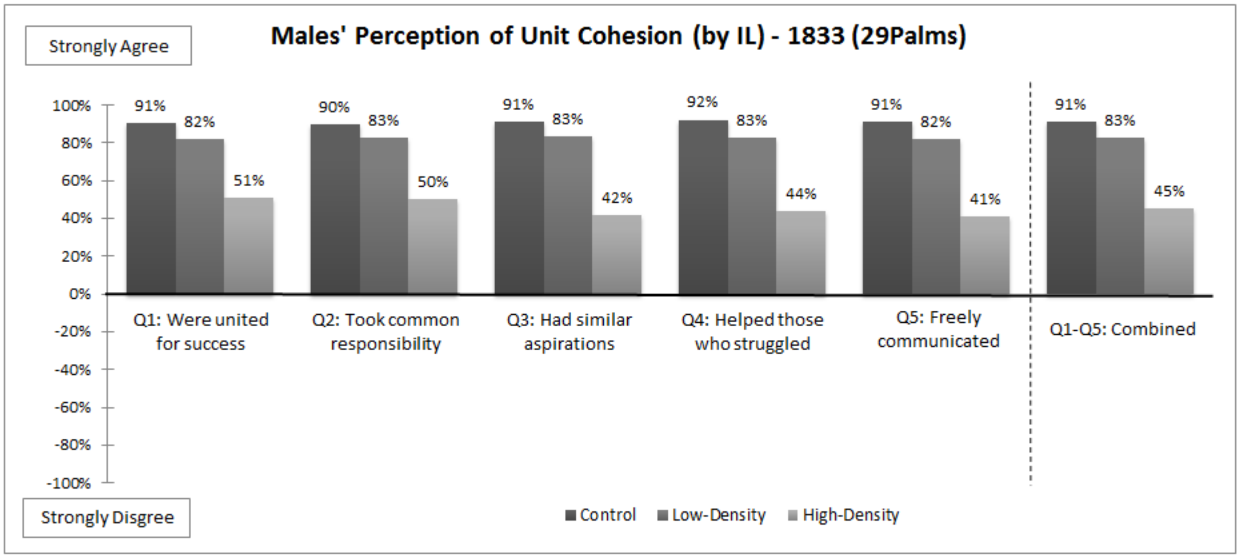


Figure 27 - Males' Perceptions of Unit Cohesion (by Integration Level) - 1833 (Pendleton)

Unfortunately, the survey did not provide for reasons as to why males were more likely to choose lower cohesion scores when in an integrated group, nor what might have changed between Twentynine Palms and Pendleton, given that the populations were almost identical.

Relative experience levels may have played a role—which may also explain the greater volatility of responses for the Amphibious scenario at Pendleton, especially when combined with the relatively low N values for High-Density groups. Incidentally, the sharper drop in males' perceptions in High-Density groups for Pendleton deviates quite a bit from how females perceived their unit cohesion.

Annex O.
Readiness

This annex details the readiness portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment. The sections outline the Data Set Description, and Modeling Results.

O.1 Data Set Description

O.1.1 Data Set Overview

The readiness dataset contains observations for each volunteer's medical readiness at an individual level. Although readiness for each Marine can be summarized as a ratio of the number of full duty days to the number of total duty days, we had much more detail on individual volunteer outcomes and we discuss the analyses of these outcomes below. The total summary measure of readiness may obscure some of the more detailed outcomes, thus, we do not analyze it as a stand-alone result.

The observed time began when the Marine joined the GCEITF and ended at the completion of the experiment. The completion of the experiment for an individual could either be the actual end of the experiment, when a Marine DOR, or when a Marine was dropped (due to medical reasons) from the experiment by GCEITF leadership. For each volunteer, we recorded the following variables, in addition to the USMC personnel variables discussed in Methodology Annex.

- Start and end dates of participation,
- Reason for end of observation (DOR or end of experiment),
- Dates of every visit to the aid station,
- Diagnosis resulting from every visit to the aid station,
- Number of documented visits to the aid station,
- Number of calendar days in which the volunteer wasn't available for participation in the experiment due to injury or illness (termed '*unavailable'* from here on).
- Number of calendar days the volunteer was unavailable due to experimental injury (we will use '*occupational injury'* from here on to describe these),
- Number of calendar days the volunteer was unavailable due to an injury or illness that was not related to his or her participation in the experiment (we term these *'non-occupational injuries'*),

- Lost training days[1] (days lost between joining the units and the start of the experiment),
- Lost Experimental days[2] (days lost between the start and end of the experiment),
- Number of calendar days until the first medical visit (occupational, non-occupational, or overall),
- Number of calendar days until the first occupational injury,

O.1.2 Volunteer Participants

At the beginning of the experiment, we had 104 female volunteers and 283 male volunteers; we only performed statistical modeling on 102 female volunteers and 254 male volunteers. Data errors or omissions caused volunteers to be excluded from modeling analysis.

O.2 Descriptive Statistics

The 283 males were available 98.4% of the time while the 104 females were available 96.8% of the time based on the total quantity of days available divided by the total days. Table O-1 provides a break out of the unavailable days by gender and causality (occupational vs. non-occupational[3]).

Table O-1. Percentage Breakout of Unavailable Days by Gender and Causality

	Occupational	Non-Occupational
Male	40%	60%
Female	80%	20%

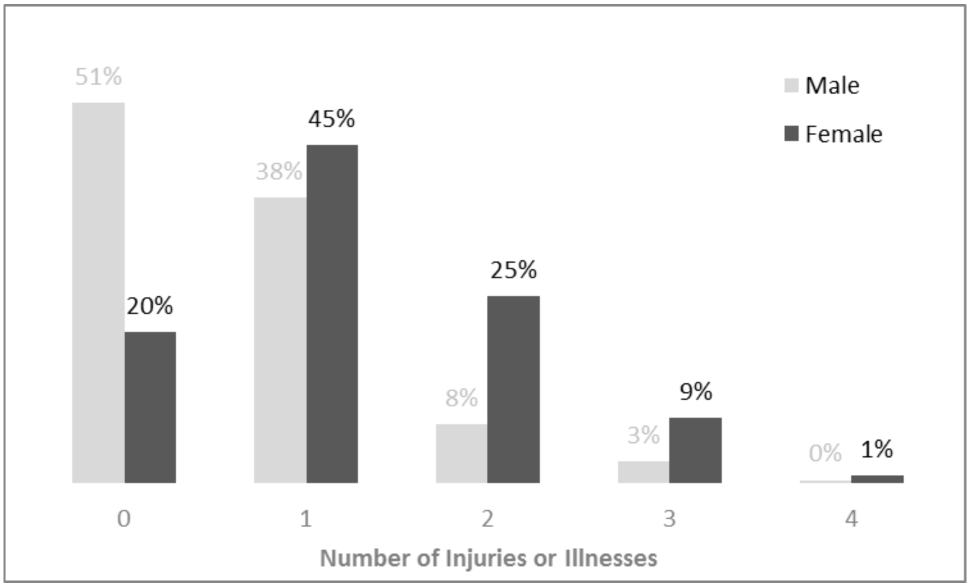
Almost half of the male volunteers (49%) and the majority of female volunteers (80%) experienced an injury or illness during their time assigned to the GCEITF. Some volunteers experienced more than one. Figure O-1 illustrates the percentage of males and females with injuries and/or illnesses ranging from 0 to 4.

[1] Training days include time from when the volunteer joined the unit and until the earlier of DOR date or 02/28/2015.

[2] Experimental days include time from 03/01/2015 and until the earlier of DOR date or the end of experiment for the volunteer's MOS.

[3] Non-occupational categorization includes illnesses (e.g., colds, flu, skin rashes, etc.) whereas occupational does not.

Figure O-1. Percentage of Volunteers, by Gender, with One or More Injuries or Illnesses



O.3 Statistical Modeling Results

O.3.1 Statistical Modeling Results Overview

The modeling section explains the relationships observed in the data in this modeling section. The goal of statistical modeling, as applied here, is to estimate, simultaneously, the effect of gender and other relevant variables on readiness. Refer to the Methodology Annex for a broad overview of the analysis plan and the variables used in the models.

For each outcome, we describe the significant variables in the model and whether these variables are either positively or negatively correlated with the result. A negative correlation indicates an increase in that variable will result in a decrease in the response variable, which is a desired outcome for injury rates and days missed from the experiment but not a desirable outcome for time to injury.

O.3.2 Readiness Method of Analysis

Because of small sample size issues, we group 0311, 0331, 0341, 0351, and 0352 Engineers, as well as Provisional Infantry and Provisional Machine Guns, into a 'hiking MOS' category; Tanks, AAV, and LAV into a 'vehicle MOS' category; and Artillery as its own category, since it is neither a hiking nor a vehicle MOS.

The starting full model for each variable includes a person's gender, MOS type, height, weight, AFQT, GT, PFT run, PFT crunch, CFT MUF, and CFT MTC times. Model selection using AIC was only possible for the DOR analysis. For other outcomes of interest, we only comment on significance of the individual variables in the full model.

Variables reported as significant are concluded to be significant based on at least a one-sided test.

We used the following types of models for the outcomes of interest. For a detailed description of the models and the rationale for their choice, refer to the Methodology Annex.

A zero-inflated negative binomial model, with gender as a predictor of the probability of event, and gender, as well as other variables, predicting the expected count of events, was used to model

- Lost training days,
- Lost experimental days,
- Number of visits to the aid station,
- Days unavailable to participate,
- Days unavailable to participate due to an occupational injury,
- Days unavailable to participate due to a non-occupational injury.

A Cox Proportional Hazards model was used to model

- Days to first medical visit,
- Days to first experimental injury.

Finally, a frailty model was used to model

- Instantaneous injury rates.

O.3.3 Readiness Modeling Results

The only personnel variables that consistently stood out as predictors of readiness were gender and MOS type. In particular, with the exception of non-occupational injuries, outcomes were always worse for female volunteers. Marines in vehicle MOSs tended to have lower injury rates than those in hiking or Artillery MOSs. The CFT MUF time appears several times as a predictor of readiness, so we recommend further investigation of the relationship of this measure with injury rates. Other variables, listed below, appeared statistically significant, but had effects of negligible size.

O.3.3.1 Lost training days

We modeled lost training days using zero-inflated negative binomial regression. The 'event' part of this model is run as a function of gender, and the count part of the model is run as a function of gender, MOS group, and personnel variables. We report significant correlations below.

Females were 18 percentage points more likely to lose training days, and the difference is statistically significant with a two-sided p-value <0.01. Of those who lost training days, however, males lost significantly more as indicated by a two-sided p-value of 0.06. Specifically, of those who lost training days, males lost over 4 times more than females, keeping all else constant.

Additionally, the following variables were found to be significantly negatively correlated to the number of lost training days, for those Marines who lost them:

- Vehicle MOS (relative to hiking MOS),
- CFT MTC time,
- PFT crunch score.

The following variables were found to be significantly positively correlated to the number of lost training days, for those Marines who lost them:

- CFT MANUF time.

O.3.3.2 Lost experimental days

We model lost experimental days using zero-inflated negative binomial regression. The ‘event’ part of this model is run as a function of gender, and the count part of the model is run as a function of gender, MOS group, and personnel variables. We report significant correlations below.

Females were 12.65 percentage points more likely to lose experimental days and the difference is statistically significant with a 2-sided p-value of 0.02. However, of those Marines who lost experimental days, there was no statistically significant difference between how many were lost by male and female Marines.

Additionally, the following variables were found to be significantly negatively correlated to the number of lost experimental days, for those Marines who lost them:

- Vehicle MOS (relative to hiking MOS),
- Weight.

The following variables were found to be significantly positively correlated to the number of lost experimental days, for those Marines who lost them:

- AFQT.

O.3.3.3 Number of visits to the aid station

We model the number of visits to the aid station using zero-inflated negative binomial regression. The ‘event’ part of this model is run as a function of gender, and the count part of the model is run as a function of gender, MOS group, and personnel variables. We report significant correlations below.

Female Marines were 37 percentage points more likely to have a medical visit, which is a statistically significant difference based on a p-value < 0.01. However, of those Marines who had at least one medical visit, male and female Marines had approximately the same number of visits, holding all else constant.

Additionally, the following variables were found to be significantly negatively correlated to the number of visits to the aid station, for those Marines who made them:

- Vehicle MOS (relative to hiking MOS),
- Weight
- GT score

The following variables were found to be significantly positively correlated to the number of visits to the aid station, for those Marines who made them:

- AFQT.

O.3.3.4 Days unavailable to participate

We model days unavailable to participate using zero-inflated negative binomial regression. The 'event' part of this model is run as a function of gender, and the count part of the model is run as a function of gender, MOS group, and personnel variables. We report significant correlations below.

Female Marines were 25 percentage points more likely to not be able to participate in the experiment. However, among the Marines who were unavailable for at least one day, females were unavailable for fewer days than males, keeping other variables constant, and the result is statistically significant in a one-sided test with a p-value of 0.09.

Additionally, the following variables were found to be significantly negatively correlated to the number days unavailable to participate, for those Marines who lost them:

- Vehicle MOS (relative to hiking MOS),
- Weight.

The following variables were found to be significantly positively correlated to the number of days unavailable to participate, for those Marines who lost them:

- CFT MANUF time.

O.3.3.5 Days unavailable to participate due to occupational injury

We model days unavailable to participate due to occupational injury using zero-inflated negative binomial regression. The ‘event’ part of this model is run as a function of gender, and the count part of the model is run as a function of gender, MOS group, and personnel variables. We report significant correlations below.

Female Marines were about 19 percentage points more likely to miss experimental days due to an injury resulting directly from performing experimental tasks, and the difference is statistically significant with a p-value <0.01. Of the Marines who missed days due to such injuries, there are no differences in the number of days missed by males and females, holding all other variables constant.

Additionally, the following variables were found to be significantly negatively correlated to the number of days unavailable to participate due to occupational injury, for those Marines who lost them:

- Vehicle MOS (relative to hiking MOS).

The following variables were found to be significantly positively correlated to the number of days unavailable to participate due to occupational injury, for those Marines who lost them:

- None.

O.3.3.6 Days unavailable to participate due to a non-occupational injury

We model the number of days unavailable to participate due to a non-occupational injury using zero-inflated negative binomial regression. The ‘event’ part of this model is run as a function of gender, and the count part of the model is run as a function of gender, MOS group, and personnel variables. We report significant correlations below.

Because of over-dispersion, the zero inflated part of the model did not estimate standard errors for this result so we cannot comment on statistical significance of the difference in probability, by gender, of missing at least one day due to a non-occupational injury.

However, the negative binomial part of the model identifies that male Marines are likely to miss 12 times the number of days due to a non-occupational injury than female Marines, holding all other variables constant.

Additionally, the following variables were found to be significantly negatively correlated to the number of days unavailable to participate due to a non-occupational injury, for those Marines who lost them:

- Weight.

The following variables were found to be significantly positively correlated to the number of days unavailable to participate due to a non-occupational injury, for those Marines who lost them:

- Artillery (relative to hiking MOS),
- AFQT,
- CFT MANUF time.

O.3.3.7 Days to first medical visit

We model days to first medical visit using a Cox Proportional Hazards model. Time to first medical visits is modeled as a function of gender, MOS group, and personnel variables. We report significant correlations below.

Survival analysis of days to first medical visit showed no difference by gender.

Additionally, the following variables were found to be significantly negatively correlated to the number of days to first medical visit:

- Vehicle MOS (relative to hiking MOS),

The following variables were found to be significantly positively correlated to the number of days to first medical visit:

- GT Score.

O.3.3.8 Days to first experimental injury

We model days to first experimental injury using a Cox Proportional Hazards model. Time to first medical visit is modeled as a function of gender, MOS group, and personnel variables. We report significant correlations below.

Survival analysis of days to first experimental injury showed no significant impact of gender.

Additionally, the following variables were found to be significantly negatively correlated to the number of days to first experimental injury:

- Vehicle MOS (relative to hiking MOS).

The following variables were found to be significantly positively correlated to the number of days to first experimental injury:

- GT score.

O.3.3.9 Frailty analysis of injury rates

We model injury rates using frailty models. Injury rates for individuals are modeled as a function of gender, MOS group, and personnel variables, while taking into account the fact that we have repeated observations for some Marines. Based on AIC and BIC, we chose Weibull distribution for the baseline hazard and Gamma for the frailty function. We report significant correlations below.

We ran three models based on three outcomes: any event that results in a medical visit (AE), only non-occupational injuries (NOI), and any occupational injuries (OI). In all three models, male Marines were statistically significantly less likely to have the negative outcomes we modeled. The hazard rate for men was 0.48 (AE), 0.53 (NOI), and 0.45 (OI) that of females, ceteris paribus.

Additionally, the following variables were found to be significantly negatively correlated to the instantaneous hazard rate in the analysis of all events leading to a medical visit:

- Weight,
- AFQT score.

The following variables were found to be significantly positively to the instantaneous hazard rate in the analysis of all events leading to a medical visit:

- Artillery MOS (relative to hiking MOS).

The following variables were found to be significantly negatively correlated to the instantaneous hazard rate of a non-occupational injury:

- None.

The following variables were found to be significantly positively to the instantaneous hazard rate of a non-occupational injury:

- Artillery MOS (relative to hiking MOS),
- Height,
- CFT MANUF time.

The following variables were found to be significantly negatively correlated to the instantaneous hazard rate of an occupational injury:

- Vehicle MOS (relative to hiking MOS),
- Weight,
- AFQT.

The following variables were found to be significantly positively to the instantaneous hazard rate of an occupational injury:

- None.

Annex P.
Proficiency and Conduct Marks

This annex details the Proficiency and Conduct Mark portion of the Ground Combat Element Integrated Task Force (GCEITF). The sections below explains the purpose of the Proficiency and Conduct Marks in the Marine Corps, logical expectations of impacts on marks given experimental realities, mitigation of negative career impacts, and descriptive statistics.

P.1 Proficiency Marks

Proficiency marks are assigned to Marines who hold the grade of E-1 to E-4. The proficiency mark indicates how well a Marine performed his or her primary duty during the performance period. The marks are provided by a Marine's commander in consultation with the officer or staff noncommissioned officer who supervised the Marine in the performance of his/her duties. Proficiency marks consider the Marine's technical skills, specialized knowledge, and the "whole Marine concept," which is commonly understood as mission accomplishment, leadership, intellect and wisdom, individual character, physical fitness, personal appearance, and completion of professional military education (HQMC, 2000).

P.2 Conduct Marks

Conduct marks are assigned to Marines who hold the grade of E-1 through E-4. The conduct mark indicates a Marine's observance of the law, regulations, conformance to accepted usage and customs, and positive contribution to unit and Corps. General bearing, attitude, interest, reliability, courtesy, cooperation, obedience, adaptability, influence on others, moral fitness, physical fitness as effected by clean and temperate habits, and participation in unit activities not related directly to unit mission are all factors in the evaluation (HQMC, 2000).

P.3 Limitations for Analysis

P.3.1 Proficiency Marks - Logical Expectations

P.3.1.1 MOS Experience

In a non-experimental unit, proficiency marks are assigned based on the performance of duties in the Marine's assigned billet. In the case of the GCEITF experiment, only a few MOSs were operating within their assigned primary MOS (i.e., Males with Closed MOSs and male and female Combat Engineers (MOS 1371)). Female Marine volunteers recruited from the Fleet Marine Force in support of this event possessed an open Primary MOS and had no ground combat-unit experience, and, therefore, did not have any fleet unit-level training experience prior to their assignment to the GCEITF.

Combat arms male Marine volunteers obtained in support of this event had fleet unit-level training experience and, in some cases, real-world deployment experience in their MOS prior to the conduct of this event. The only exception is Male Provisional Infantryman and Provisional Machine Gunners who may or may not have been assigned to closed units prior to the GCEITF experiment. This disparity in occupational experience is noted as a limitation.

P.3.1.2 Billet Assignment

In many cases in the Fleet Marine Force, a junior Marine begins his or her career in a basic billet with minimal complexity, advancing through billets of increasing responsibility as proficiency increases. Per the design of the experiment, personnel assignments to billets were randomized, resulting in an unconventional training period before the start of the experimental phase as well as a change in billet assignment with every new trial during the conduct of the experiment.

P.3.1.3 Physiological Disparities

The GCEITF was designed to measure performance differences in the execution of tasks that were identified as the most physically demanding within each experimental MOS. Furthermore the selected tasks were ones that could be done repetitively and were not necessarily the ones that required the most cognitive/technical skills and therefore did not lend themselves to evaluation of anything other than physicality. The expectation that each Marine perform his or her duties within that MOS lends itself to a stratification of performance levels based, on some level, on the physical capacity to execute the task. It stands to reason that two Marines who are physiologically different from each other might perform a physical task with different levels of effectiveness, and that, theoretically, this would affect the way their evaluating superiors perceived their proficiency.

P.3.2 Conduct Marks: Logical Expectations

Contrary to proficiency, conduct marks are assigned in ways that are generally blind to primary versus billet MOS. Anecdotally, conduct marks tend to be similar to proficiency marks.

P.3.2.1 Conduct Marks: Experimental Nuances

Unlike within Fleet Marine Force Units, the GCEITF volunteers reserved the unusual right to Drop on Request. It has been theorized that this had an inordinate effect on the behavior of some volunteers, and that it imparts a sense of power that undermines traditional military order. In most cases, volunteers acted with the level of professionalism expected of United States Marines. As such, one would anticipate that conduct marks reflected the behavior that volunteers exhibited, but that but some distortion from the norms likely remained.

P.3.2.2 Sample sizes

Proficiency and Conduct marks are assigned to Marines in the grade of E-1 to E-4. This decreased the sample size of the groups being compared and analyzed. Though we provide summary statistics to describe the data, in most cases, sample sizes are sufficiently small to where any extrapolation of these descriptions outside of the GCEITF sample is not appropriate.

P.4 Experimental Mitigation

Per the EAP:

> *"Male and Female Open MOS volunteers participating in the provisional rifle company experiment as well as female volunteer participants for closed MOSs will not be performing in their normal MOS functions for the duration of their assignment to the GCEITF. Duty proficiency marks will be assigned in accordance with the Individual Records Administration Manual (HQMC, 2000) and the commanding officer of the GCEITF will provide due allowance for Marines who will be filling billets inconsistent with their grade and normal MOS skills."*

In practical terms, participation in the experiment cannot harm the participants, including their chances for promotion or retention. Of particular note, when a Marine was operating outside his or her MOS within the ITF, the proficiency and conduct marks assigned were administratively used to compare the Marine to others in that Marine's primary MOS. For instance, a Female Marine serving in the 0311 portion of the experiment, with a primary MOS of 0411, would not administratively (in the context of the greater Marine Corps population) be compared to other 0311s. Instead, her marks count toward her composite score and her promotion potential is gauged against 0411s, regardless of her assigned duties.

The special allowances necessary to protect the volunteers, which are provided by the commanding officer of the GCEITF in the issuance of the proficiency and conduct ratings, preclude their use in any meaningful analysis.

P.5 Descriptive Statistics

Tables P-1 and P-2 provided the descriptive statistics for the GCEITF volunteers.

Table P-1. Descriptive Statistics for Proficiency Marks

		Including DOR individuals			Excluding DOR individuals		
Metric	**Gender**	**Sample Size**	**Mean**	**SD**	**Sample Size**	**Mean**	**SD**
Proficiency Mark – all MOS's (includes DOR individuals)	M	198	4.40	0.17	144	4.42	0.15
	F	71	4.45	0.20	55	4.47	0.20
0311	M	50	4.36	0.19	26	4.42	0.17
	F	8	4.28	0.27	5	4.36	0.30

		Including DOR individuals			[redacted]	Excluding DOR individuals		
Metric	**Gender**	**Sample Size**	**Mean**	**SD**	[redacted]	**Sample Size**	**Mean**	**SD**
0313	M	10	4.53	0.13	[redacted]	9	4.53	0.13
	F	7	4.41	0.04	[redacted]	7	4.41	0.04
0331	M	8	4.41	0.19	[redacted]	4	4.38	0.13
	F	7	4.44	0.19	[redacted]	5	4.48	0.22
0341	M	6	4.45	0.20	[redacted]	4	4.50	0.20
	F	5	4.58	0.19	[redacted]	4	4.55	0.21
0351	M	2	[redacted]	[redacted]	[redacted]	1	[redacted]	[redacted]
	F	3	[redacted]	[redacted]	[redacted]	0	[redacted]	[redacted]
0352	M	4	[redacted]	[redacted]	[redacted]	2	[redacted]	[redacted]
	F	0	[redacted]	[redacted]	[redacted]	0	[redacted]	[redacted]
0811	M	28	4.41	0.13	[redacted]	27	4.42	0.13
	F	13	4.58	0.15	[redacted]	11	4.58	0.15
1371	M	19	4.28	0.19	[redacted]	14	4.32	0.18
	F	9	4.28	0.19	[redacted]	7	4.30	0.21
1812	M	18	4.41	0.19	[redacted]	16	4.43	0.14
	F	2	4.55	0.07	[redacted]	2	4.55	0.07
1833	M	14	4.47	0.12	[redacted]	13	4.47	0.13
	F	5	4.48	0.13	[redacted]	5	4.48	0.13
PI	M	33	4.39	0.14	[redacted]	26	4.42	0.13
	F	10	4.40	0.13	[redacted]	6	4.38	0.15
PMG	M	6	4.42	0.12	[redacted]	2	4.40	0.14
	F	4	4.6	0.22	[redacted]	3	4.70	0.10

Table P-2. Descriptive Statistics for Conduct Marks

		Including DOR individuals			[redacted]	Excluding DOR individuals		
Metric	**Gender**	**Sample Size**	**Mean**	**SD**	[redacted]	**Sample Size**	**Mean**	**SD**
Conduct Mark (includes DOR individuals)	M	198	4.39	0.17	[redacted]	144	4.41	0.15
	F	73	4.45	0.18	[redacted]	55	4.47	0.18
0311	M	50	4.36	0.19	[redacted]	26	4.42	0.17
	F	8	4.31	0.26	[redacted]	5	4.38	0.33
0313	M	10	4.48	0.17	[redacted]	9	4.48	0.18
	F	7	4.40	0.06	[redacted]	7	4.40	0.06
0331	M	8	4.41	0.20	[redacted]	4	4.43	0.13
	F	7	4.47	0.16	[redacted]	5	4.52	0.16
0341	M	6	4.47	0.23	[redacted]	4	4.50	0.27
	F	5	4.58	0.18	[redacted]	4	4.50	0.19
0351	M	2	[redacted]	[redacted]	[redacted]	1	[redacted]	[redacted]
	F	3	[redacted]	[redacted]	[redacted]	0	[redacted]	[redacted]
0352	M	4	[redacted]	[redacted]	[redacted]	2	[redacted]	[redacted]
	F	0	[redacted]	[redacted]	[redacted]	0	[redacted]	[redacted]

		Including DOR individuals			Excluding DOR individuals		
Metric	**Gender**	**Sample Size**	**Mean**	**SD**	**Sample Size**	**Mean**	**SD**
0811	M	28	4.40	0.11	27	4.41	0.11
	F	13	4.48	0.11	11	4.48	0.12
1371	M	19	4.29	0.18	14	4.32	0.18
	F	9	4.28	0.22	7	4.31	0.23
1812	M	18	4.36	0.15	16	4.38	0.10
	F	2	4.60	0.14	2	4.60	0.14
1833	M	14	4.48	0.16	13	4.48	0.17
	F	5	4.56	0.11	5	4.56	0.11
PI	M	33	4.38	0.14	26	4.40	0.15
	F	10	4.41	0.12	6	4.43	0.15
PMG	M	6	4.43	0.14	2	4.40	0.00
	F	4	4.58	0.21	3	4.67	0.12

THIS PAGE INTENTIONALLY LEFT BLANK

Annex Q.
GCEITF Population

This annex describes the population of the GCEITF volunteers when compared to today's Marine Corps Active Component who possess the same parameters as the GCEITF population, i.e., E-1 to E-5s, full-duty medical status. The sections outline the Population Overview, Assumptions, Limitations, Descriptive Statistics, Results, and Insights.

Q.1 GCEITF Population Overview

Q.1.1 Background

The ability to generalize the results of the GCEITF experiment to the wider Marine Corps is, in part, dependent upon how representative the GCEITF participants were of the total population of Marines.

Q.1.2 Purpose

This annex presents comparative analysis results for selected personnel variables judged to be indicative of relevant military capabilities. These results inform judgment about the repeatability of the experimental results and the expectation that observed integration effects would be observed throughout a similarly integrated Marine Corps.

Q.1.3 Analytic Objective

The fundamental question addressed by this analysis is:

How representative are the GCEITF Marine volunteers when compared to the Marine Corps' total Active Component population of Marines with the same parameters?

Q.1.3.1 Participant Parameters

First, to answer this question, the comparisons between the GCEITF volunteer sample and the Marine Corps' total active component population (query database pulled June 2014) had to be of similar parameters to include:

- The total Marine Corp's population was defined using the Active Component, whereas the GCEITF sample contains both Active and Reserve Component Marines.
- Pay grade of E-5 and below.
- Full-duty status (i.e., not currently on limited/light duty or pending physical evaluation board).

- For MOS-to-MOS comparisons, Marines must hold the same Primary MOS (PMOS). MOS-to-MOS comparisons were done by gender (i.e., Female 1371 Combat Engineers within the GCEITF were compared to female Marines who also possess the 1371 PMOS within the total Marine Corps, given the same parameters; Male PI and PIMG Marines were compared to male open MOS Marines within the total Marine Corps, given the same parameters)

From this point on, this annex will refer to the Marine Corps' total Active Component population used for comparison as the total Marine Corps Population.

The GCEITF population encompassed all volunteers who actively participated in the experiment. While the number of volunteers was in constant flux, the maximum active participation by volunteers at any one time was 382 (278 males/104 females).

Q.1.3.2 Personnel Variables used for comparison

Second, to answer this question, the comparisons between the GCEITF volunteer sample and the Marine Corps' total active component population were based on variables that are common among all Marines, regardless of gender. The tables and graphs present the mean of each variable. While the PFT/CFT score used in the comparisons are gender-normed, the component scores of PFT and CFT events that are continuous and/or gender-neutral (i.e., run times (continuous time captured), crunch score (where 100 is the maximum score regardless of gender) were used to give greater insight into the individual personnel variables of the Marines within the two groups. The 12 personnel variables used include:

- Age
- Height
- Weight
- AFQT
- GT
- PFT, Total Score
- PFT, Crunches
- PFT, Run Time
- CFT, Total Score
- CFT, Maneuver under Fire
- CFT, Movement to Contact
- Rifle Range Score

Q.1.3.3 Analytical comparisons

Lastly, to answer this question, several comparisons were conducted. All comparisons given for the below personnel variables are done so between the GCEITF population and the total Marine Corps population with the same parameters, aggregated and by PMOS. For each of the 12 personnel variables, 15 comparisons were made yielding a total of 180 comparisons. The comparisons are as follows:

Q.1.3.3.1 Objective 1 (Closed MOSs) Comparisons

Q.1.3.3.1.1 *Male-Male Marines (Closed MOSs)*

GCEITF male Marines participating within their closed MOSs were compared in two ways. They were compared to the total male Marine Corps population and they were compared to the male Marines currently serving with the same MOS.

Q.1.3.3.1.2 *Female-Female Marines (Closed MOSs)*

GCEITF female Marines participating within their closed MOSs were compared to the total female Marine Corps population. Comparing females to other females serving in closed MOSs wasn't possible, as that population currently does not exist. An open proxy MOS was also not used.

Q.1.3.3.2 Objective 2 (Open MOSs in Closed MOS Units) Comparisons

Q.1.3.3.2.1 *Male-Male Marines (Open MOSs)*

GCEITF male Marines participating within their open MOSs, as part of Provisional Infantry (PI) or as Provisional Infantry Machinegunners (PIMG), were compared in two ways. They were compared to the total male Marine Corps population and they were compared to the male Marines currently serving in open MOSs.

GCEITF male Marines participating within the 1371 Combat Engineer MOSs were also compared in two ways. They were compared to the total male Marine Corps population and they were compared to the male Marines currently serving in the 1371 MOS.

Q.1.3.3.2.2 *Female-Female Marines (Open MOSs)*

GCEITF female Marines participating within their open MOSs, as part of Provisional Infantry, were compared to the total female Marine Corps population.

GCEITF female Marines participating within the 1371 Combat Engineer MOSs were compared in two ways. They were compared to the total female Marine Corps population and they were compared to the female Marines currently serving in the 1371 MOS.

Q.1.3.4 Analytical Methodology

The personnel variables were compared using a non-parametric Kolmogorov-Smirnov Test (KST) for differences. Observed differences were judged to be statistically significant if the KST produced a p-value less than or equal to 0.10.

Q.2 Assumptions

Conclusions depend on the validity of assumptions. These assumptions include:

- The Marine Corps Reserve Component Marines are similar both physically and mentally in the performance of Marine Corps fitness and cognitive tests as the Active Marine Component.
- When comparison of continuous/gender-neutral personnel variables (i.e., PFT run time, AFQT score) were used for analysis, it is assumed that the Marines times and/or scores reflect the best effort of the individual.

Q.3 Limitations

The following limitations are noted:

- Because a population of female Marines serving in currently closed MOSs does not exist, this Annex is limited to only comparing those GCEITF female Marines who participated in Closed MOS experiments to the total female Marine Corps population, which consists of Open MOSs only. There is no MOS that can serve as a proxy MOS for comparison purposes.

Q.4 Descriptive and Basic Inferential Statistics

Q.4.1 Descriptive Statistics Overview

This section includes a comparative analysis of personnel variables among GCEITF volunteers and the larger Marine Corps based on the results of descriptive statistics, non-parametric Kolmogorov-Smirnov Test (KST) for differences, and density plots. This section presents results that are statistically significant for each gender and for MOS. The Population Appendix contains the descriptive statistics for the remainder of the personnel variables.

Q.4.2 Results overview

One hundred eighty comparisons were conducted and only 30 provided evidence of statistically significant differences between the MOSs analyzed between GCEITF and the total Marine Corps population. Of those 30 differences, the following results are presented for MOS-to-MOS comparisons:

- 14 were observed between male Closed MOS-MOS comparisons.
- 4 were observed between male Open MOS-to-MOS comparisons.
- 4 were observed between male 1371 MOS-to-MOS comparisons.

- 4 between the female GCEITF volunteers and the total female Marine Corps population.
- 4 were observed between female 1371 MOS-to-MOS comparisons.

The differences noted were all within the normal limits of any particular variable. With the exception of very few differences, the GCEITF volunteers were representative of the total population of the Marine Corps. As such, the experimental results can be generalized to the wider Marine Corps with confidence in the repeatability of the experiment and our expectation that performance differences observed in the experiment would also appear in a Marine Corps similarly integrated.

Q.4.2.1 Age

The table Q-1 and Q-2 below display the results for comparing the age of the male GCEITF volunteers to that of the total Marine Corps population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figures Q-1 through Q-4 are density plots that display the density of calculated values in a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps population.

Table Q-1 – GCEITF Male Age Compared to Population Male Age

Male Age												
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	278	22.45	2.45	18.80	21.85	31.60	79212	23.03	3.03	17.40	22.30	47.10	0.02*

Table Q-2. Significant Male Age Comparisons by MOS

Male Age													
MOS (w/ DOR)	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0352	16	21.71	1.61	20.00	21.25	25.60	790	23.35	3.24	18.50	22.60	36.30	0.07*
1812	19	23.12	2.08	19.80	22.90	27.20	379	22.87	2.78	18.70	22.00	34.20	0.05*
PI&PMG	55	21.69	2.01	19.40	20.90	28.00	58368	23.18	3.04	17.70	22.50	44.40	< 0.01*

Figure Q-1 – Age by Population for Males

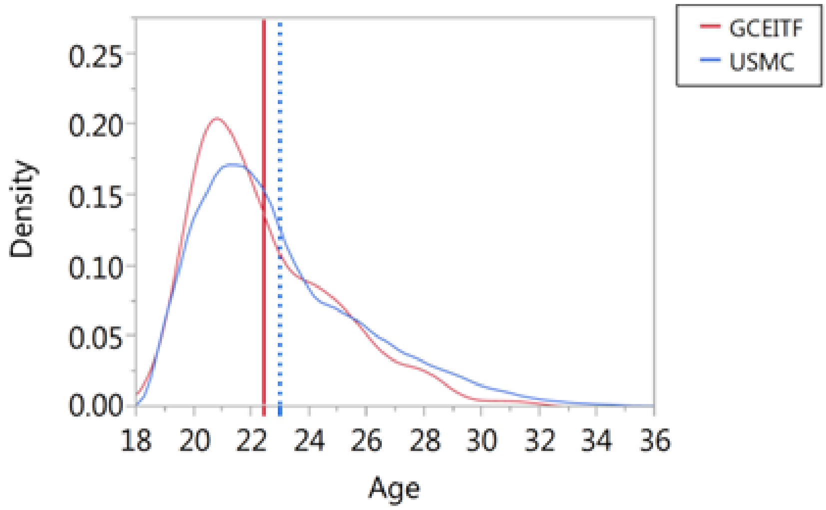


Figure Q-2. Age by Population for Males MOS 0352

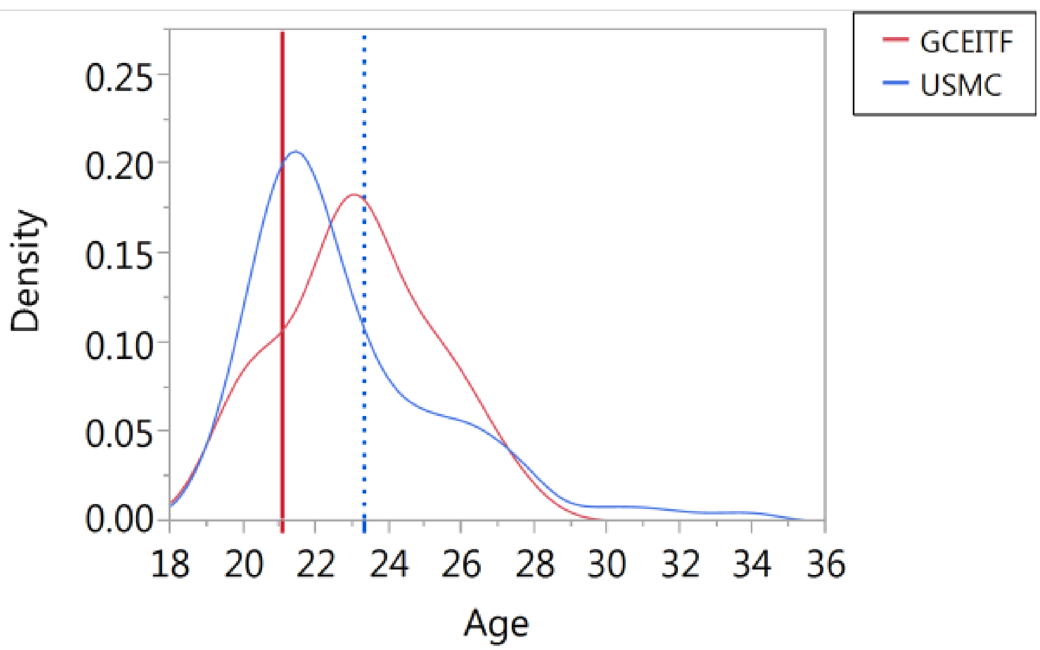


Figure Q-3. Age by Population for Males MOS 1812

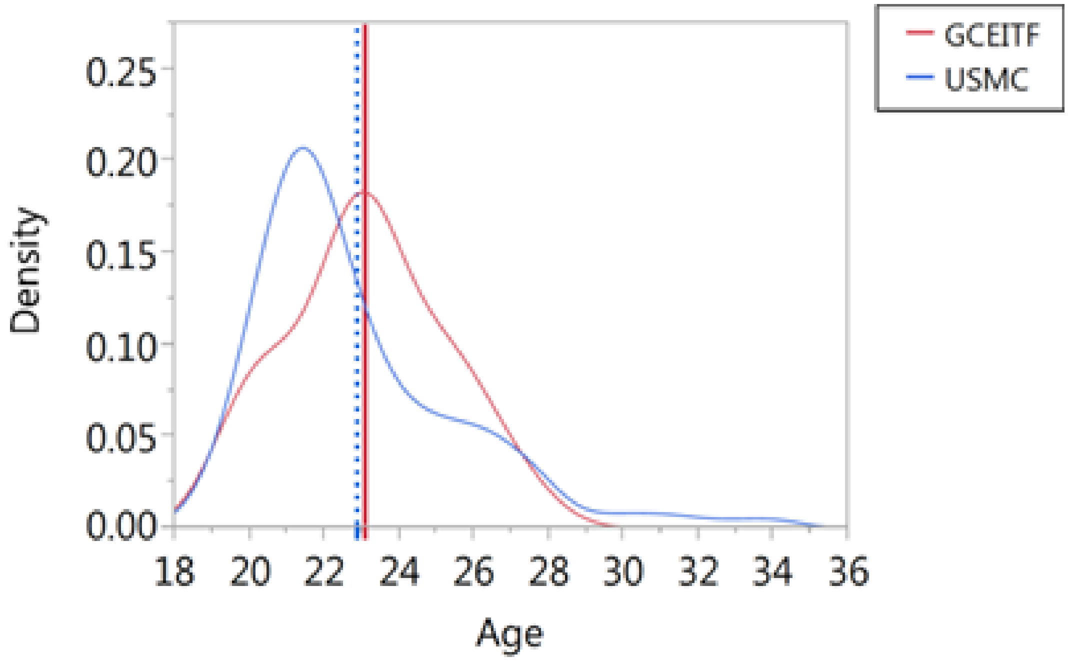
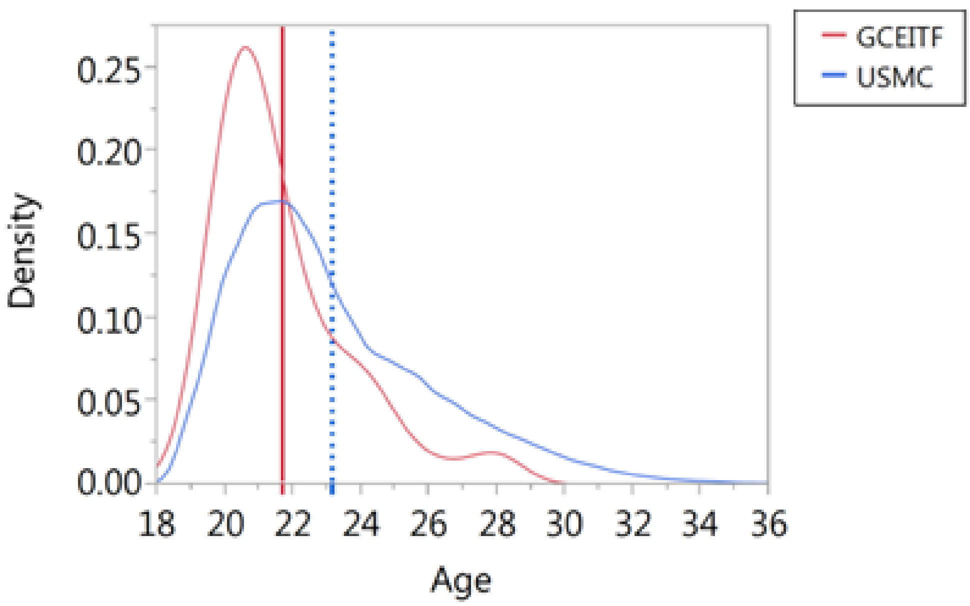


Figure Q-4. Age by Population for Males MOS PI & PIMG



The average age of the GCEITF male was 22.45 years, and the average age of a male in the Marine Corps Population was 23.03 years. Based on the p-value of the K-S test being less than 0.10, this age difference is statistically significant. The GCEITF males were statistically significantly younger than the Marine Corps Population, by approximately 6 months. The following MOSs had statistically significant differences:

- 0352 GCEITF males' average age was 21.71 years and the average age of a male in the Marine Corps 0352 population was 23.35 years. The GCEITF males were statistically significantly younger, by approximately 20 months.
- 1812 GCEITF males' average age was 23.12 years and the average age of a male in the Marine Corps 1812 population was 22.87 years. The GCEITF males were statistically significantly older, by approximately 3 months.
- PI & PIMG GCEITF males' average age was 21.69 years and the average age of a male in the Marine Corps open MOS population was 23.18 years. The GCEITF males were statistically significantly younger, by approximately 18 months.

There is no statistically significant difference in age between female Marines in the GCEITF when compared to females in the total Marine Corps Population or in any of the female MOS-to-MOS comparisons.

Q.4.2.2 Height

There is no statistically significant difference in height between Marines in the GCEITF when compared to the Marine Corps Population or in any of the MOS-to-MOS comparisons.

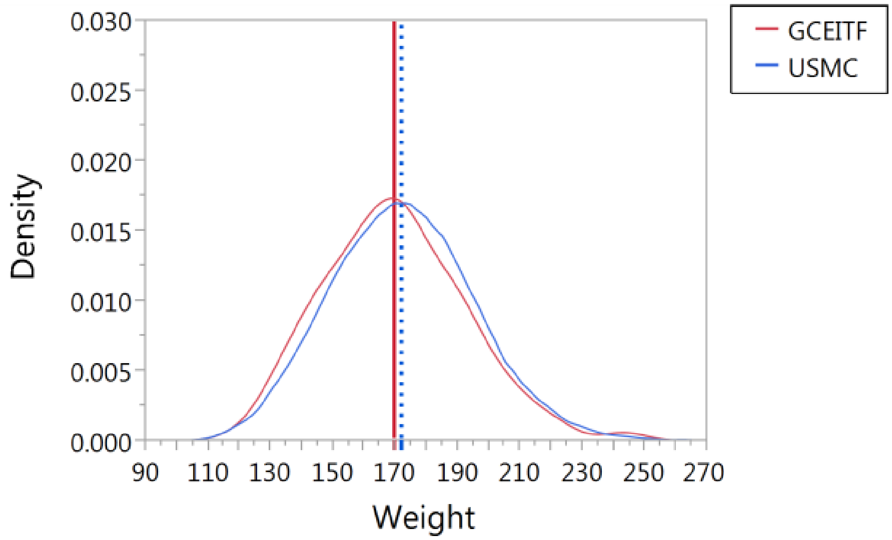
Q.4.2.3 Weight

Table Q-3 displays the results from the comparison of the weight of the male GCEITF volunteers to that of males in the Marine Corps Population given the same parameters. The table displays non-parametric KST results and p-values suggesting statistical significance. Figure Q-5 is a density plots that display the density of calculated values in a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps Population.

Table Q-3. GCEITF Male Weight Compared to Population Male Weight

Male													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	278	169.79	22.92	117.00	170.00	246.00	79198	172.36	23.46	92.00	172.00	295.00	0.08*

Figure Q-5. Weight by Population for Males



The average weight of the GCEITF male was 169.79 lb and the average weight of a male in the total Marine Corps Population was 172.36 lb. Based on the p-value of the K-S test being less than .10, this difference is statistically significant. The GCEITF males were statistically significantly lighter than the total Marine Corps Population, by approximately 2.5 lb. There are no statistically significant differences in any of the MOS-to-MOS comparisons.

There is no statistically significant difference in weight between females in the GCEITF when compared to the total Marine Corps Population or in any of the female MOS-to-MOS comparisons.

Q.4.2.4 AFQT

Table Q-4 through Q-6 below display the results for comparing the AFQT of the male and female GCEITF Volunteers to that of the total Marine Corps Population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figure Q-6 through Q-10 are density plots that display the density of calculated values to a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps population.

Table Q-4. GCEITF Male AFQT Compared to Male Population AFQT

Male													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	277	60.56	19.19	22.00	59.00	99.00	79199	62.63	17.79	4.00	62.00	99.00	0.03*

Table Q-5. Significant Male AFQT Comparisons by MOS

Male AFQT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD		Median	Max	N	Mean	SD	Min	Median	Max	P-value
0313	14	70.64	20.59	38.0	77.0	95.0	447	59.72	17.74	23.0	56.0	99.0	0.05*
1833	20	61.90	18.27	31.0	69.0	89.0	906	54.93	16.82	20.0	52.0	98.0	0.04*
PI&PMG	55	57.96	17.86	28.0	54.0	95.0	58356	63.24	17.73	9.0	63.0	99.0	0.04*

Table Q-6. Significant Female AFQT Comparisons by MOS

Female AFQT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
1371	9	74.00	17.73	39	75	92	124	60.56	16.86	33	59.5	95	0.04*

Figure Q-6. AFQT by Population for Males

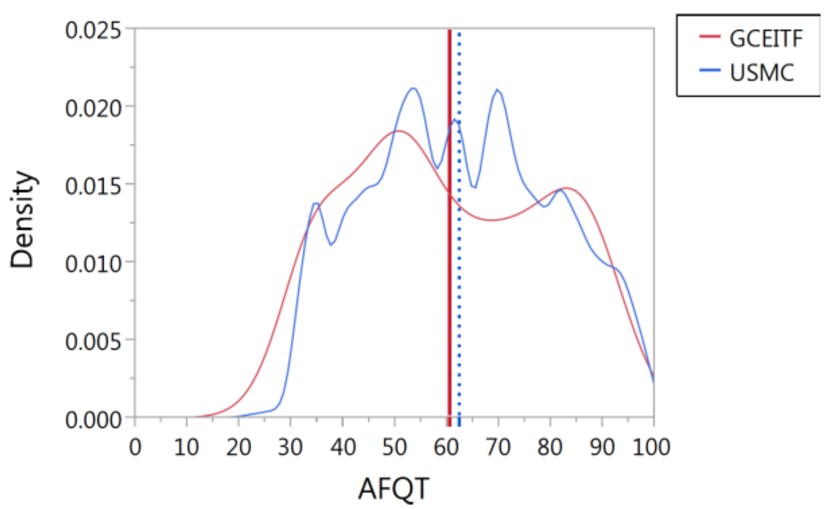


Figure Q-7. AFQT by Population for Males MOS 0313

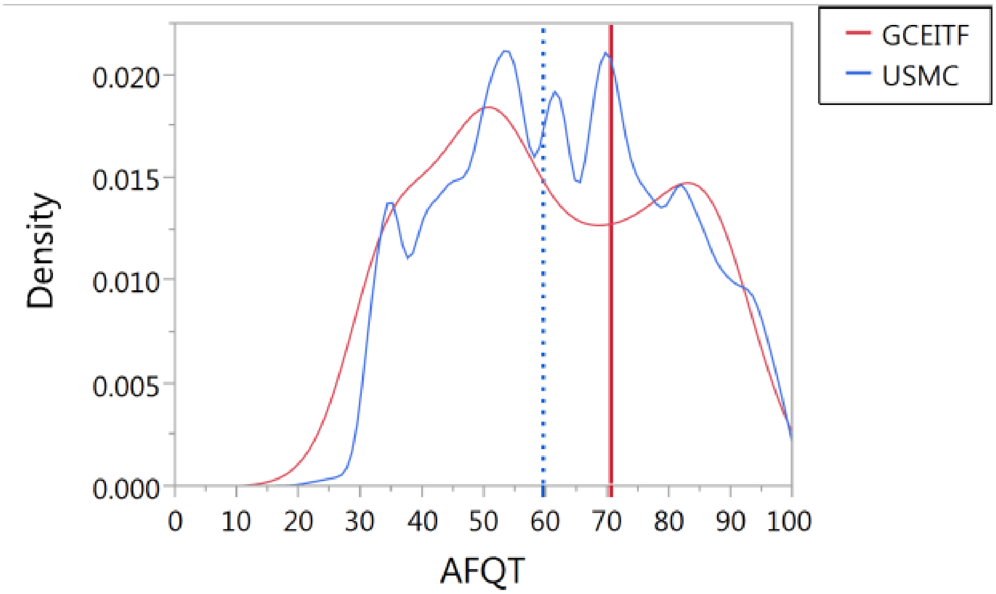


Figure Q-8. AFQT by Population for Males MOS 1833

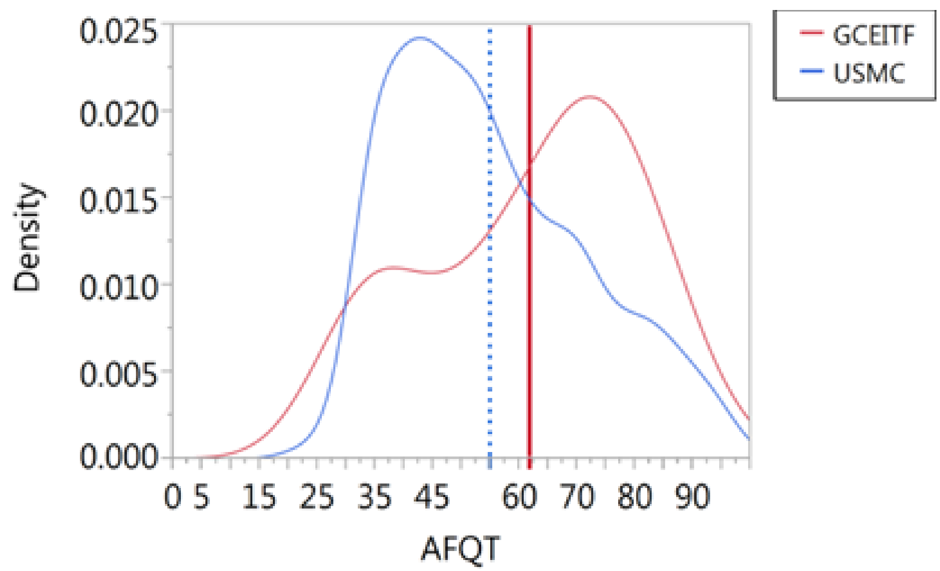


Figure Q-9. AFQT by Population for Males MOS PI & PIMG

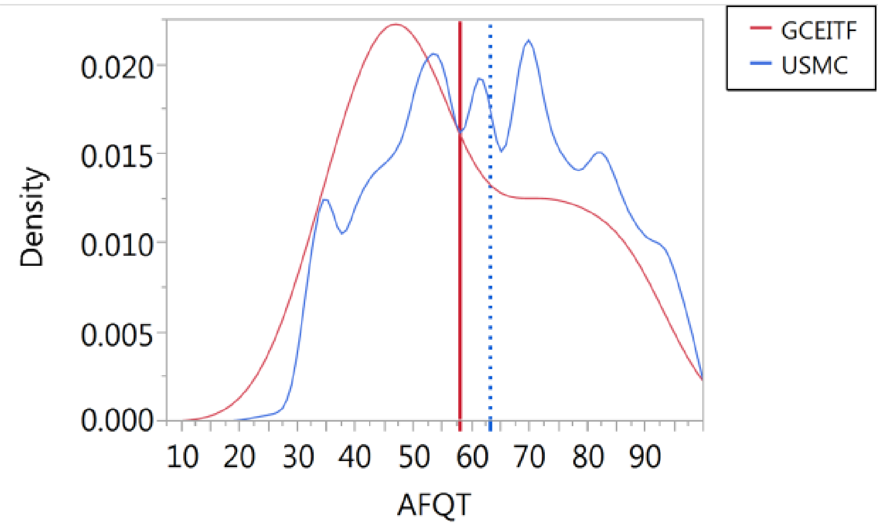
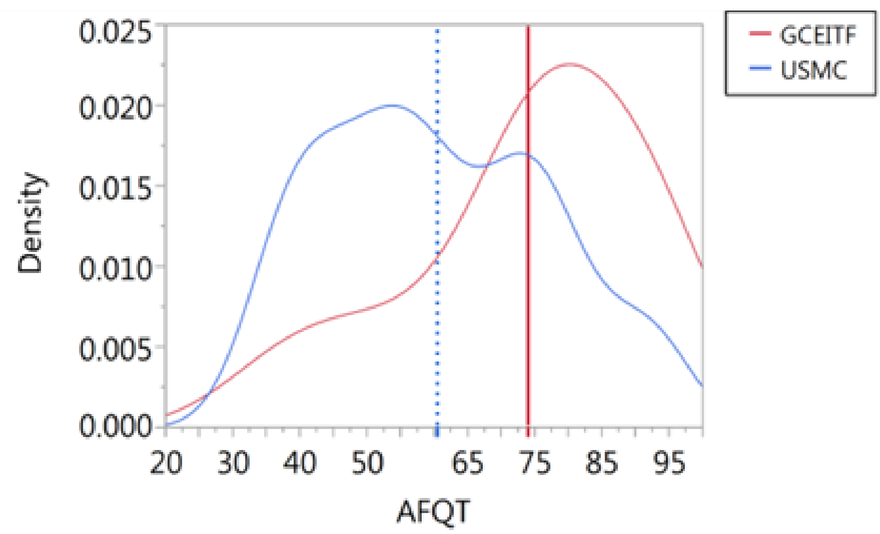


Figure Q-10. AFQT by Population for Females MOS 1371



The average AFQT score of the GCEITF male was 60.56 and the average AFQT score of a male in the Marine Corps Population was 62.63. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF male had statistically significant lower scores when compared to the Marine Corps Population, by approximately 2 points. The following MOSs had statistically significant differences:

- 0313 GCEITF males' average AFQT score was 70.64 and the average AFQT score for a male in the Marine Corps 0313 population was 59.72. The GCEITF males' AFQT scores were statistically significantly greater, by approximately 11 points.

- 1833 GCEITF males' average AFQT score was 61.90 and the average AFQT score for a male in the Marine Corps 1833 population score was 54.93. The GCEITF males' AFQT scores were statistically significantly greater, by approximately 7 points.
- PI & PIMG GCEITF males' average AFQT score was 57.96 and the average AFQT score of a male in the Marine Corps open MOS population was 63.24. The GCEITF males' AFQT scores were statistically significantly less, by approximately 5 points.

There is no statistically significant difference in AFQT score between female Marines in the GCEITF when compared to the total Marine Corps Population. The following MOS had a statistically significant difference:

- 1371 GCEITF females' average AFQT score was 74.00 and the average AFQT score for a female in the Marine Corps 1371 population was 60.56. The GCEITF females' AFQT scores were statistically significantly greater, by approximately 13 points.

Q.4.2.5 GT

Tables Q-7 and Q-8 below display the results for comparing the GT score of the male GCEITF Volunteers to that of the males in the Marine Corps Population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figure Q-11 and Q-12 are density plots that display the density of calculated values to a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps population.

Table Q-7. GCEITF Male GT Score Compared to Male Population GT Score

Male													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	272	106.79	12.17	80.00	106.00	136.00	79207	108.99	11.73	12	108	159	< 0.01*

Table Q-8. Significant Male GT Comparisons by MOS

Male GT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value

PI&PMG	54	104.44	12.47	81	102.0	136	58363	109.33	11.73	12	108	159	0.01*

Figure Q-11. GT by Population for Males

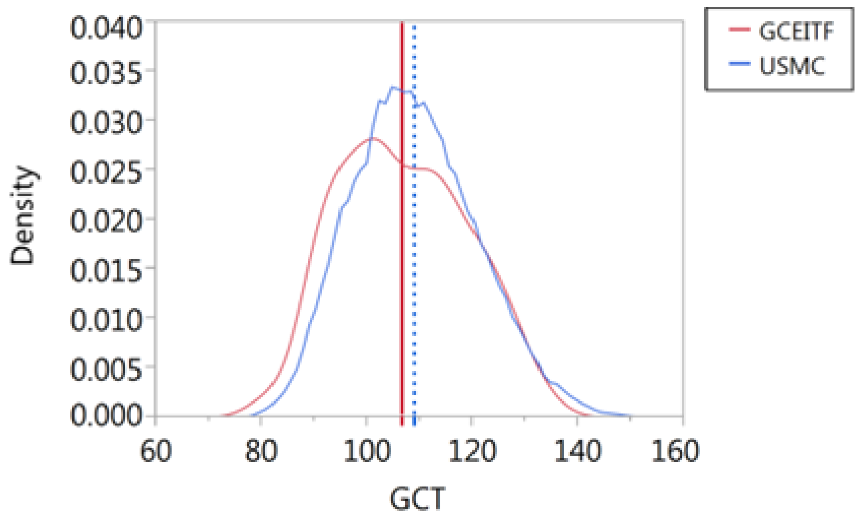
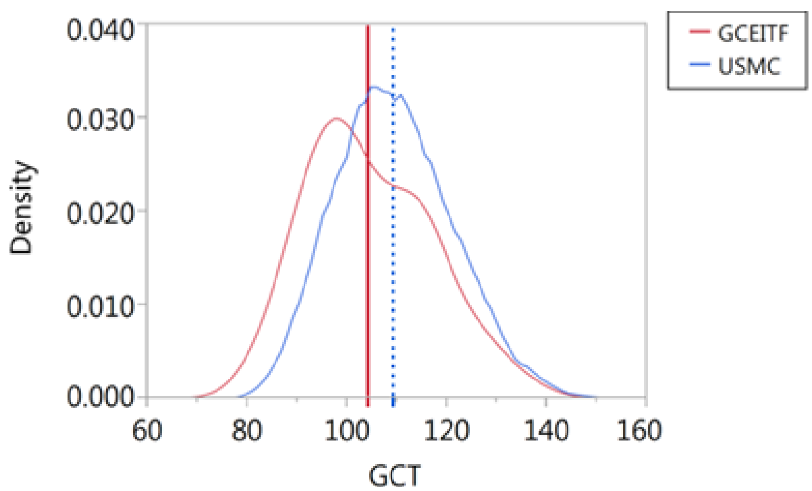


Figure Q-12. GT by Population for Males for MOS PI & PIMG



The average GT score of the GCEITF male was 106.79 and the average GT score of a male in the Marine Corps Population was 108.99. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF males scored lower and the difference is statistically significant compared to the males in the Marine Corps Population, by approximately 2.2 points. The following MOS had a statistically significant difference:

- PI & PIMG GCEITF males' average GT score was 104.44 and the average GT score of a male in the Marine Corps open MOS population was 109.33. The GCEITF males' GT scores were statistically significantly lower, by approximately 5 points.

There is no statistically significant difference in GT scores between females in the GCEITF when compared to females in the Marine Corps Population. There are no statistically significant differences in any of the female MOS-to-MOS comparisons.

Q.4.2.6 Physical Fitness Test (PFT) Score

Table Q-9 through Q-11 below display the results for comparing the Physical Fitness Test (PFT) Scores of the male and female GCEITF Volunteers to that of the total Marine Corps Population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figure Q-13 through Q-17 are density plots that display the density of calculated values in a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps Population.

Table Q-9. Significant Male PFT Score Comparisons by MOS

Male PFT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0331	13	237.00	27.61	181.0	243.0	281.0	1532	263.84	25.69	159.0	271.0	300.0	< 0.01*
0811	35	239.06	34.67	165.0	246.0	294.0	1179	254.45	27.38	136.0	259.0	300.0	< 0.01*
1371	19	244.58	34.09	194.0	247.0	300.0	1876	255.41	28.03	138.0	261.0	300.0	0.09*

Table Q-10. GCEITF Female PFT Score Compared to Female Population PFT Score

Female													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	104	275.72	20.03	211.00	280.00	300.00	6419	255.51	31.36	125	263	300	< 0.01*

Table Q-11. Significant Female PFT Score Comparisons by MOS

Female PFT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
1371	9	281.89	14.59	253	285	300	123	259.95	31.18	160	266	300	0.08*

Figure Q-13. PFT by Population for Males MOS 0331

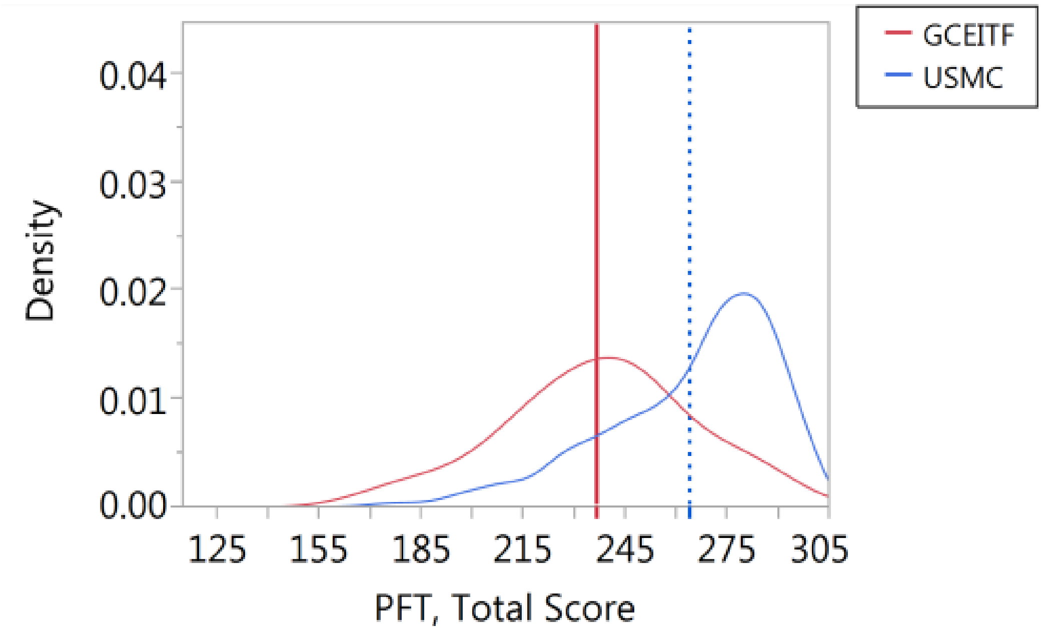


Figure Q-14. PFT by Population for Males MOS 0811

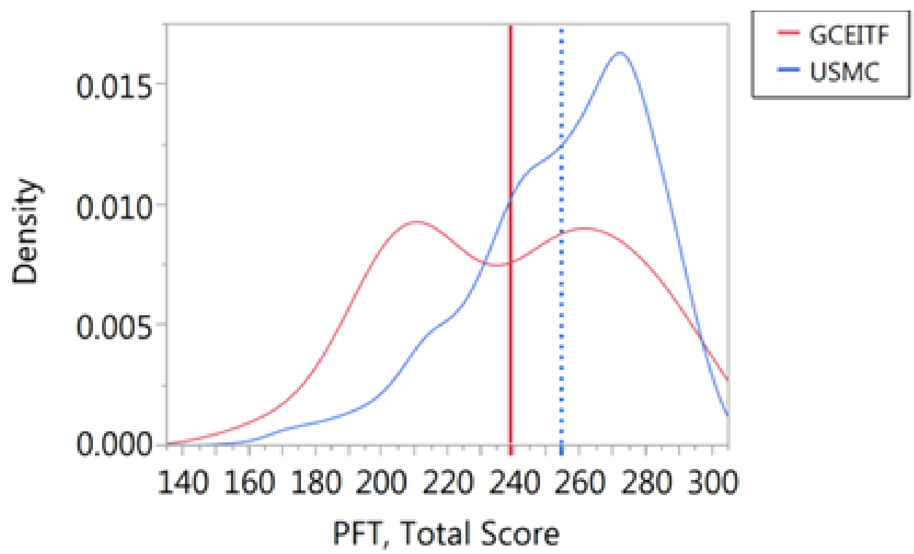


Figure Q-15. PFT by Population for Males MOS 1371

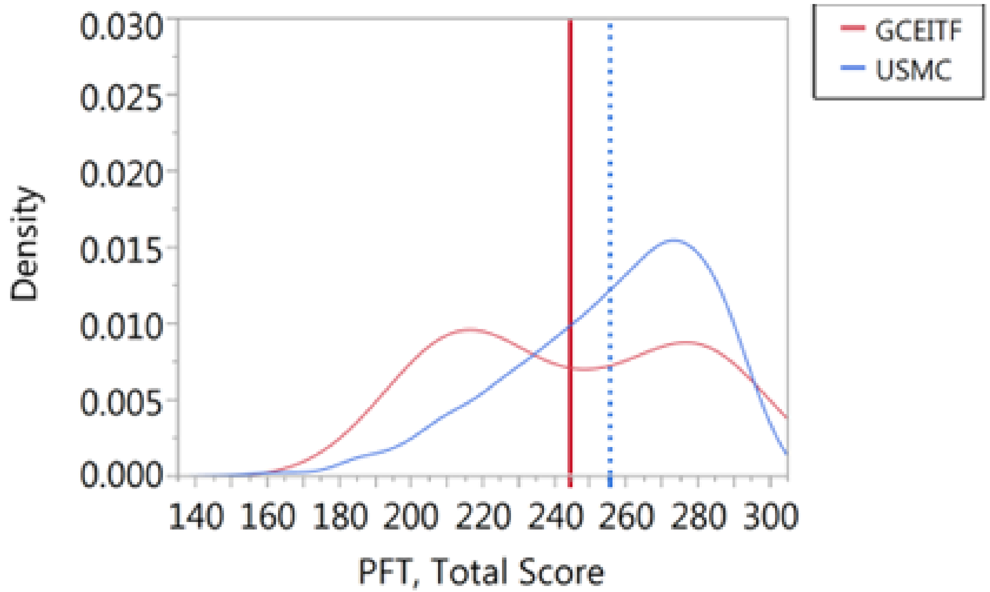


Figure Q-16. PFT by Population for Females

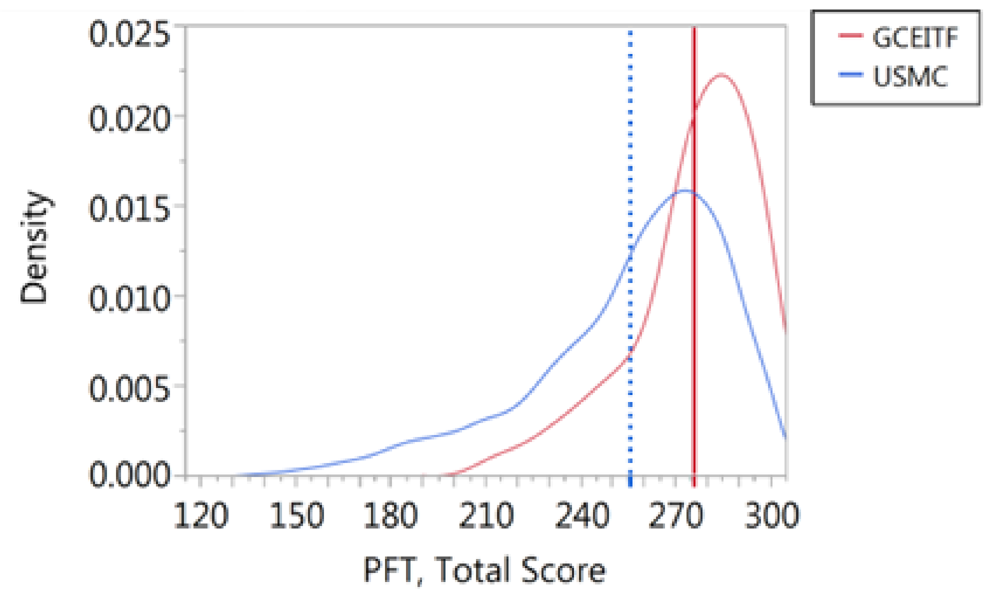
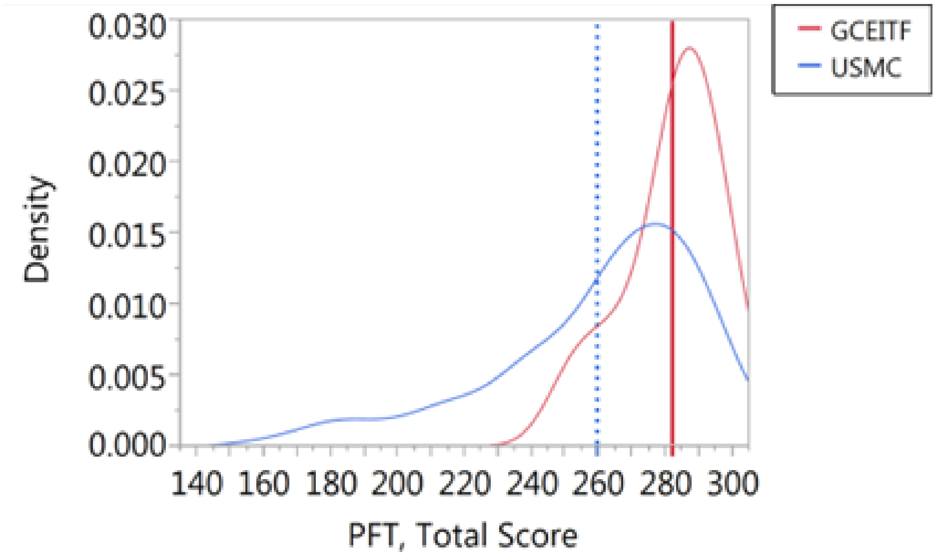


Figure Q-17. PFT by Population for Females MOS 1371



There is no statistically significant difference in PFT score between males in the GCEITF when compared to males in the total Marine Corps Population. The following MOSs had statistically significant differences:

- 0331 GCEITF males' average PFT score was 237 and the average PFT score for males in the Marine Corps 0331 population was 263.83. The GCEITF males' average PFT score was statistically significantly lower, by approximately 27 points.
- 0811 GCEITF males' average PFT score was 239.06 and the average PFT score for males in the Marine Corps 0811 population was 254.45. The GCEITF males' average PFT score was statistically significantly lower, by approximately 15 points.
- 1371 ITF males' average PFT score was 244.58 and the average PFT score for males in the Marine Corps 1371 population was 255.41. The GCEITF males' average PFT score was statistically significantly lower, by approximately 11 points.

The average PFT score of the GCEITF female was 275.72 and the average PFT score of a female in the Marine Corps Population was 255.51. Based on the p-value of the K-S test being greater than 0.10, this difference is statistically significant. The GCEITF females' PFT scores were statistically significantly higher than the females in the Marine Corps Population, by approximately 20 points. The following MOS had a statistically significant difference:

- 1371 GCEITF females' average PFT score was 281.89 and the average PFT score for females in the Marine Corps 1371 population was 259.95. The GCEITF females' average PFT score was statistically significantly higher, by approximately 22 points.

Q.4.2.7 PFT Crunches

There is no statistically significant difference in PFT crunches between Marines in the GCEITF when compared to the total Marine Corps Population. There are no statistically significant differences in any of the MOS-to-MOS comparisons.

Q.4.2.8 PFT Run time

Table Q-12 through Q-14 below display the results from the comparisons of the PFT run times of the male and female GCEITF Volunteers to that of the total Marine Corps Population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figure Q-18 through Q-20 are density plots that display the density of calculated values in a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps population.

Table Q-12. Significant Male Run Time Comparison by MOS

Male Run Time (minutes)													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
1812	19	23.91	1.93	20.20	24.00	29.00	367	22.48	2.03	17.22	22.35	30.25	<0.01*

Table Q-13. GCEITF Female Run Time Compared to Female Population Run Time

Female													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	103	23.79	2.20	19.3	23.67	30.7	6397	25.76	2.55	17.55	25.70	40.98	<.01*

Table Q-14. Significant Female Run Time Comparison by MOS

Female													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
1371	9	22.99	1.34	20.53	23.43	24.67	121	25.27	2.55	19.85	25.25	32.67	<0.01*

Figure Q-18. PFT Run Time by Population for Males MOS 1812

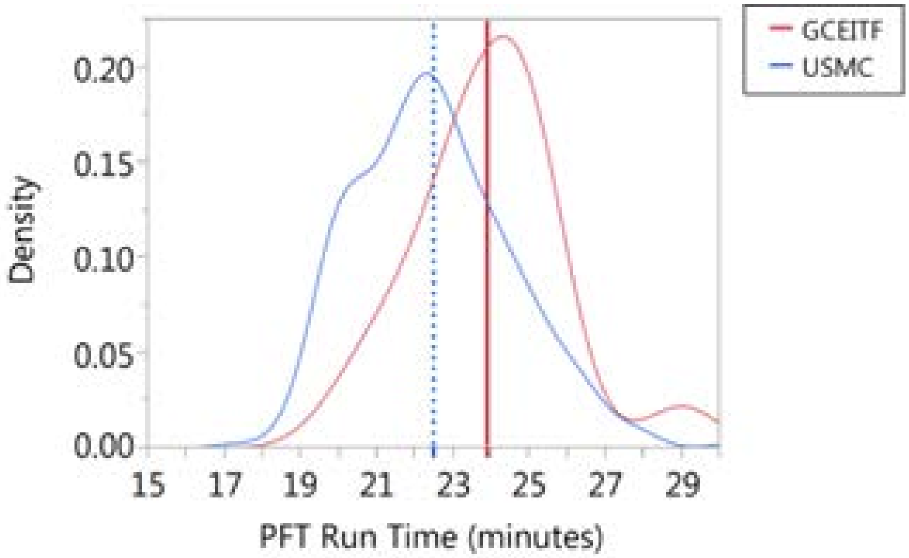


Figure Q-19. PFT Rum time by Population for Females

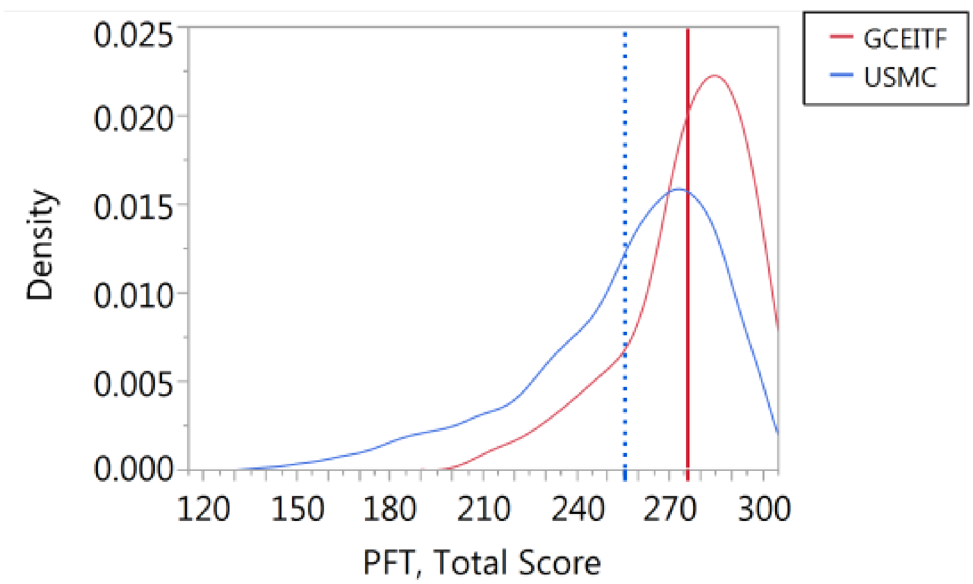
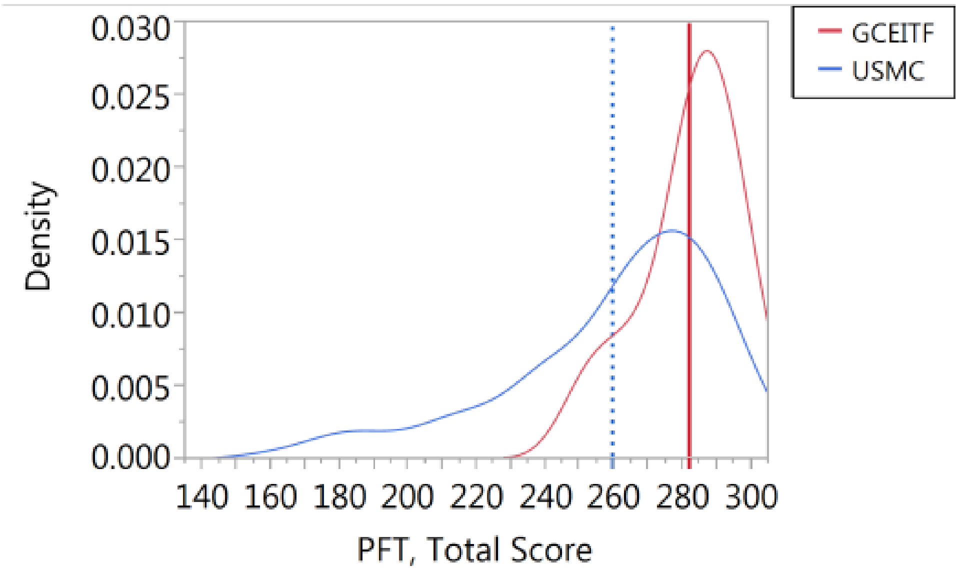


Figure Q-20. PFT Run Time by Population for Females MOS 1371



There is no statistically significant difference in PFT run times between males in the GCEITF when compared to males in the Marine Corps Population. The following MOSs had statistically significant differences:

- 1812 GCEITF males' average PFT run time was 23.91 minutes and the average PFT run time for males in the Marine Corps 1812 population was 22.48 minutes. The GCEITF males' average PFT run time was statistically significantly slower, by approximately 1:25.

The average PFT run time of the GCEITF female was 23.79 minutes and the average PFT run time of a female in the Marine Corps Population was 25.76 minutes. Based on the p-value of the K-S test being greater than 0.10, the difference is statistically significant and faster, on average, by approximately 2 minutes. The following MOSs had statistically significant differences:

- 1371GCEITF females' average PFT run time was 22.99 minutes and the average PFT run time for females in the Marine Corps 1371 population was 25.72. The GCEITF Marines' average PFT run time was statistically significantly faster.

Q.4.2.9 Combat Fitness Test (CFT) score

Table Q-15 through Q-17 below display the results for comparing the CFT scores of the male and female GCEITF Volunteers to that of the total Marine Corps Population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figure Q-21 through Q-26 are density plots that display the density of calculated values in a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps population.

Table Q-15. GCEITF Male CFT Score Compared to Male Population CFT Score

Male													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	275	287.44	11.19	236.00	290.00	300.00	78567	288.29	11.83	148.00	291.00	300.00	0.04*

Table Q-16. Signficant Male CFT Score Comparisons by MOS

Male													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	63	287.41	11.47	258.0	291.0	300.0	9732	290.93	9.99	214.0	294.0	300.0	0.04*
0341	18	284.56	10.33	261.0	287.5	297.0	1580	291.42	10.27	157.0	294.0	300.0	0.01*
0811	35	285.49	9.69	262.0	286.0	300.0	1185	288.59	10.76	226.0	291.0	300.0	0.06*
1371	19	283.68	11.35	265.0	283.0	300.0	1872	289.20	10.71	222.0	292.0	300.0	0.02*

Table Q-17. GCEITF Female CFT Score Compared to Female Population CFT Score

Female													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	103	292.98	8.37	266	297	300	6476	286.83	13.17	143	290	300	< 0.01*

Figure Q-21. CFT by Population for Males

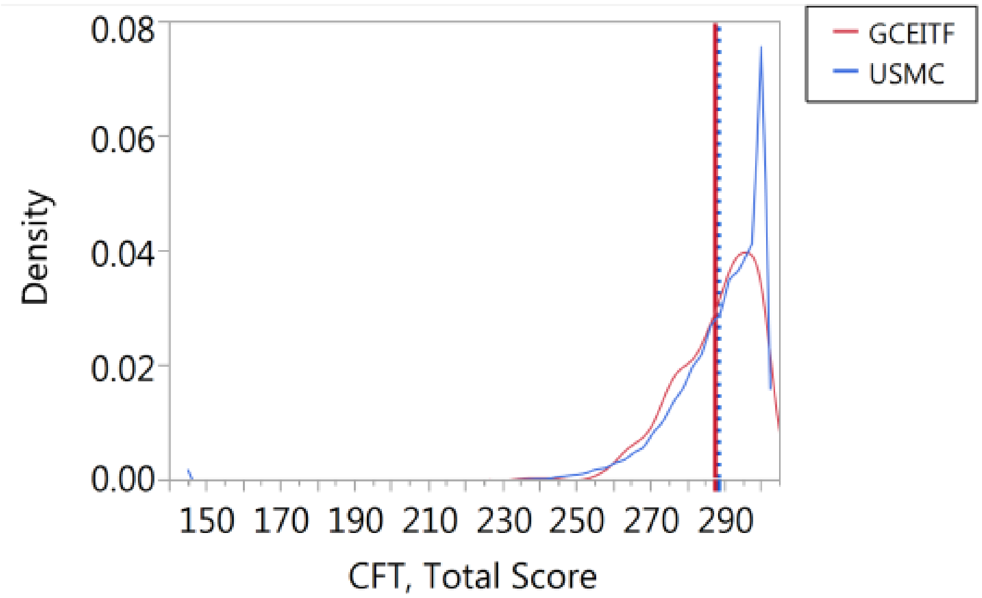


Figure Q-22. CFT by Population for Males MOS 0311

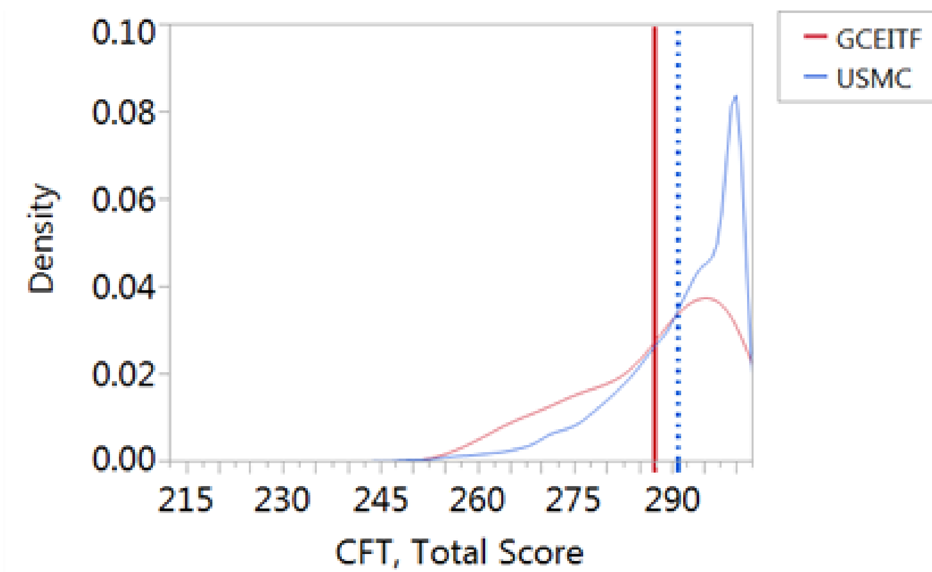


Figure Q-23. CFT by Population for Males MOS 0341

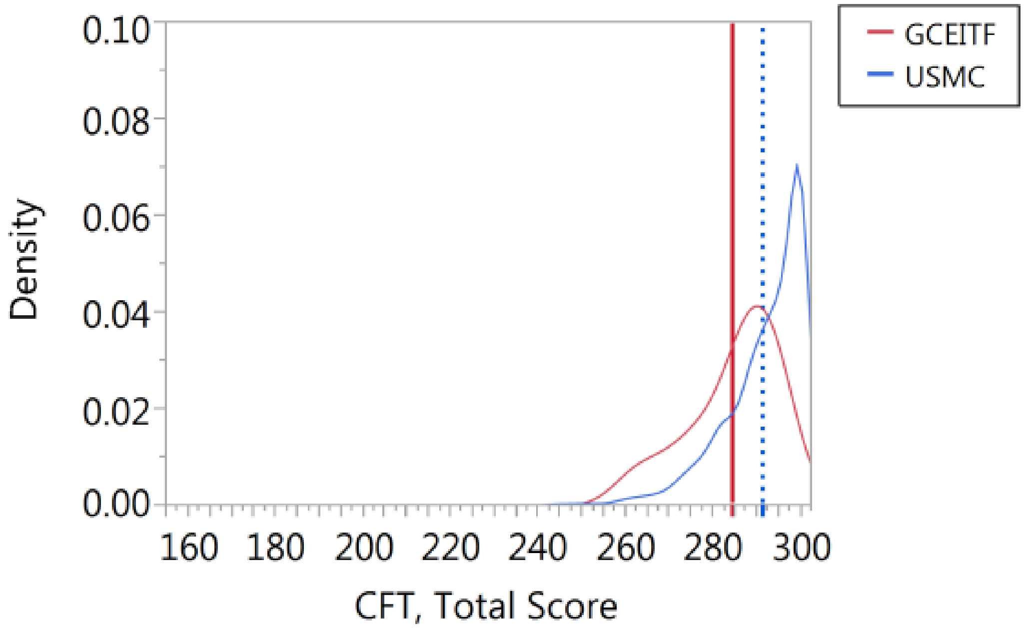


Figure Q-24. CFT by Population for Males MOS 0811

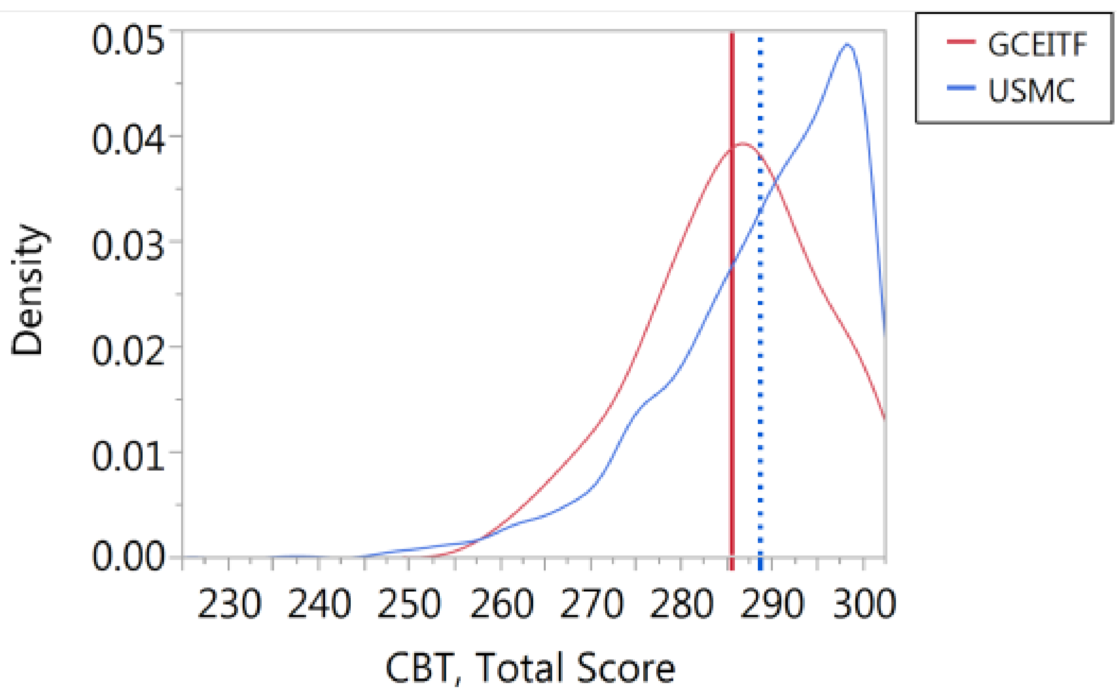


Figure Q-25. CFT by Population for Males MOS 1371

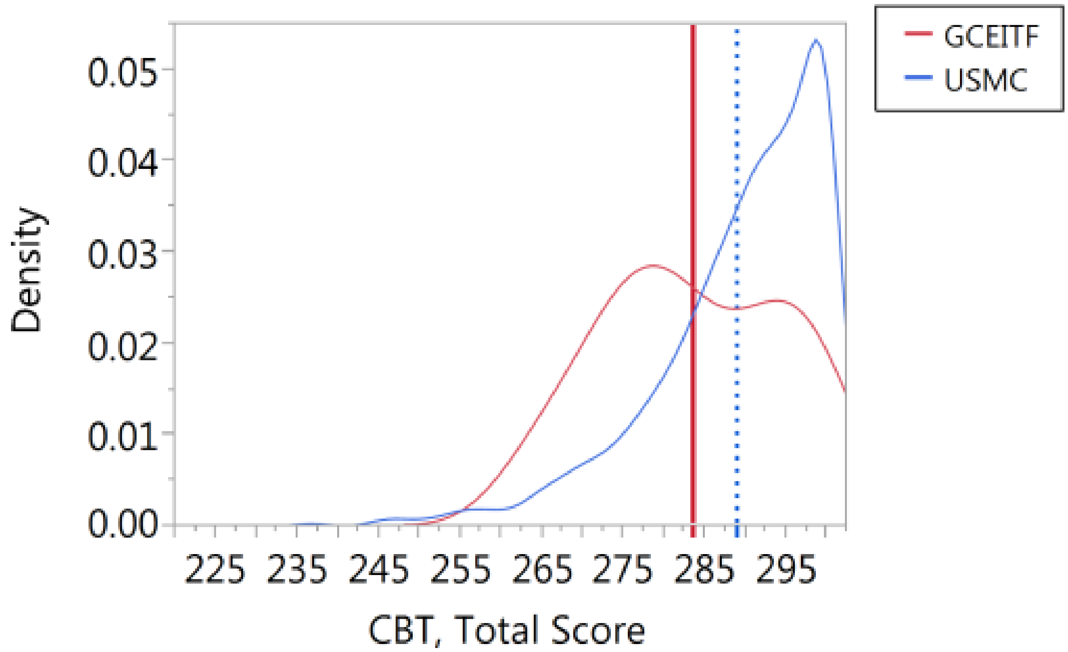
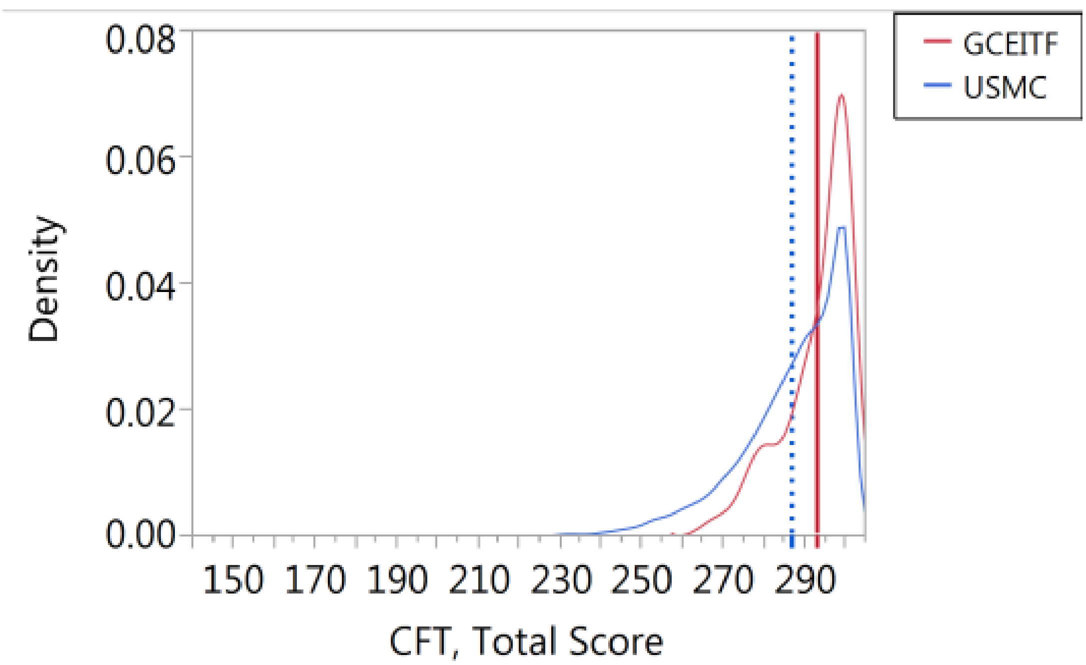


Figure Q-26. CFT by Population for Females



The average CFT score of the GCEITF male was 287.44 and the average CFT score of a male in the Marine Corps Population was 288.29. Based on the p-value of the K-S test being greater than 0.10, the difference is statistically significant and better, on average, by approximately 1 point. The following MOSs had statistically significant differences:

- 0311 GCEITF males' average CFT score was 284.56 and the average CFT score for males in the Marine Corps 0311 population was 290.42. The GCEITF males' average CFT score was statistically significantly lower, by approximately 4 points.
- 0341 GCEITF males' average CFT score was 287.41 and the average CFT score for males in the Marine Corps 0341 population was 290.93. The GCEITF males' average CFT score was statistically significantly lower, by approximately 4 points.
- 0811 GCEITF males' average CFT score was 285.49 and the average CFT score for males in the Marine Corps 0811 population was 288.59. The GCEITF males' average CFT score was statistically significantly lower, by approximately 3 points.
- 1371 GCEITF males' average CFT score was 283.68 and the average CFT score for males in the Marine Corps 1371 population was 289.20. The GCEITF males' average CFT score was statistically significantly lower, by approximately 6 points.

The average CFT score of the GCEITF female was 292.98 and the average CFT score of a female in the Marine Corps Population was 286.83. Based on the p-value of the K-S test being greater than 0.10, there is a statistically significant difference between the two groups. The GCEITF females' CFT scores are higher, on average, by approximately 6 points. There are no statistically significant differences in any of the MOS-to-MOS comparisons.

Q.4.2.10 CFT Maneuver Under Fire (MUF) time

Table Q-18 through Q-20 below display the results for the comparisons of the CFT MUF time of the male and female GCEITF Volunteers to that of the total Marine Corps Population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figure Q-27 through Q-31 are density plots that display the density of calculated values in a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps population.

Table Q-18. GCEITF Male MUF Time Compared to Male Population MUF Time

Male													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	275	2.44	0.31	1.80	2.40	3.75	78688	2.40	0.32	1.08	2.37	8.00	0.06*

Table Q-19. Significant Male MUF Time Comparisons by MOS

MOS	GCEITF N	GCEITF Mean	GCEITF SD	GCEITF Min	GCEITF Median	GCEITF Max	USMC N	USMC Mean	USMC SD	USMC Min	USMC Median	USMC Max	KS Test P-value
0311	63	2.44	0.34	1.8	2.4	3.8	9738	2.32	0.29	1.5	2.3	5.3	<0.01*
0341	18	2.48	0.23	2.0	2.5	2.9	1582	2.32	0.31	1.6	2.3	8.0	<0.01*
1371	19	2.58	0.30	1.9	2.6	3.0	1876	2.37	0.28	1.7	2.3	3.8	<0.01*

(Male)

Table Q-20. GCEITF Female MUF Time Compared to Female Population MUF Time

All MOS	GCEITF N	GCEITF Mean	GCEITF SD	GCEITF Min	GCEITF Median	GCEITF Max	USMC N	USMC Mean	USMC SD	USMC Min	USMC Median	USMC Max	KS Test P-value
DOR incl	103	3.09	0.42	2.2	3.08	4.1	6489	3.27	0.47	2	3	7	<0.01*

(Female)

Figure Q-27. CFT MUF by Population for Males

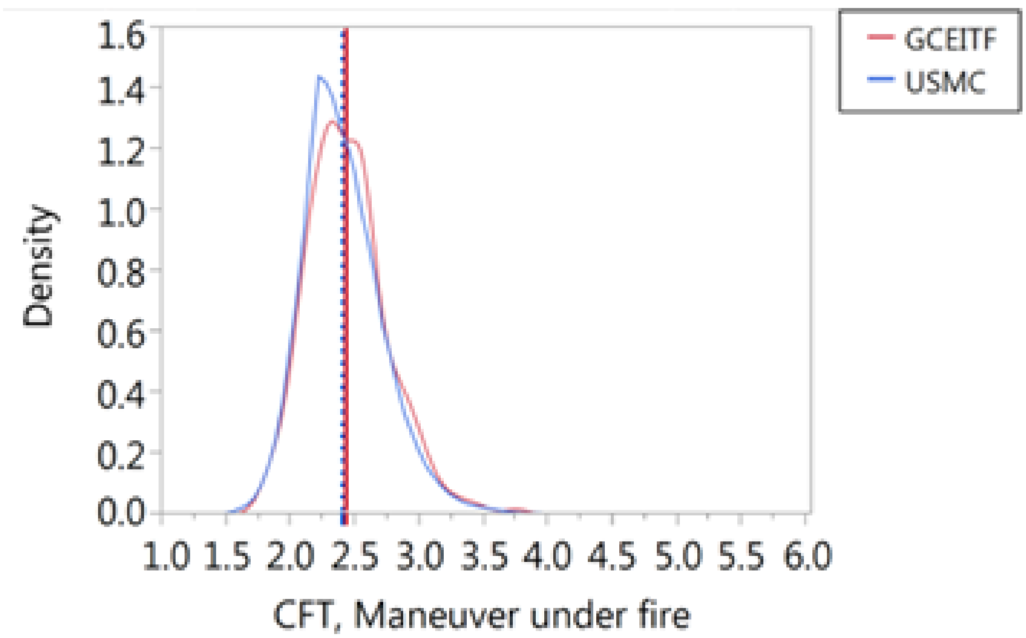


Figure Q-28. CFT MUF by Population for Males MOS 0311

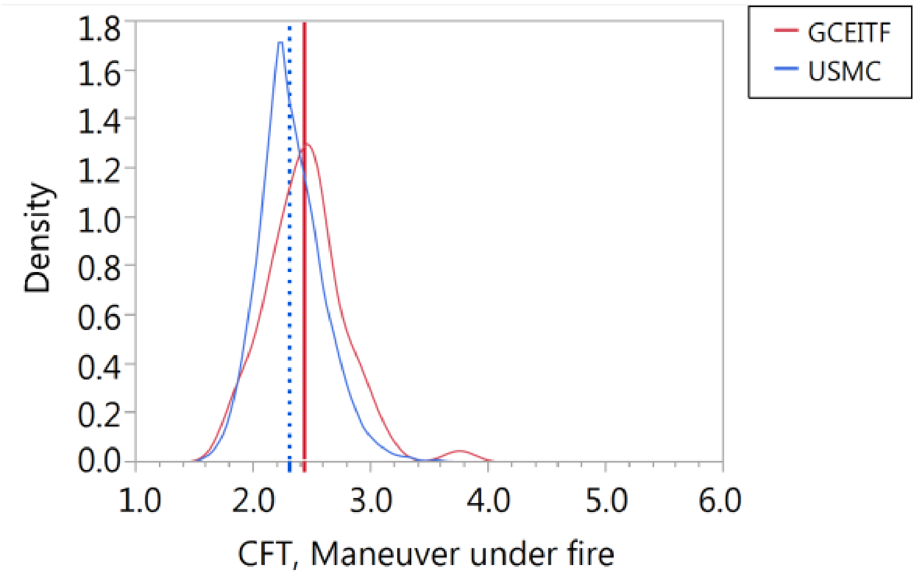


Figure Q-29. CFT MUF by Population for Males MOS 0341

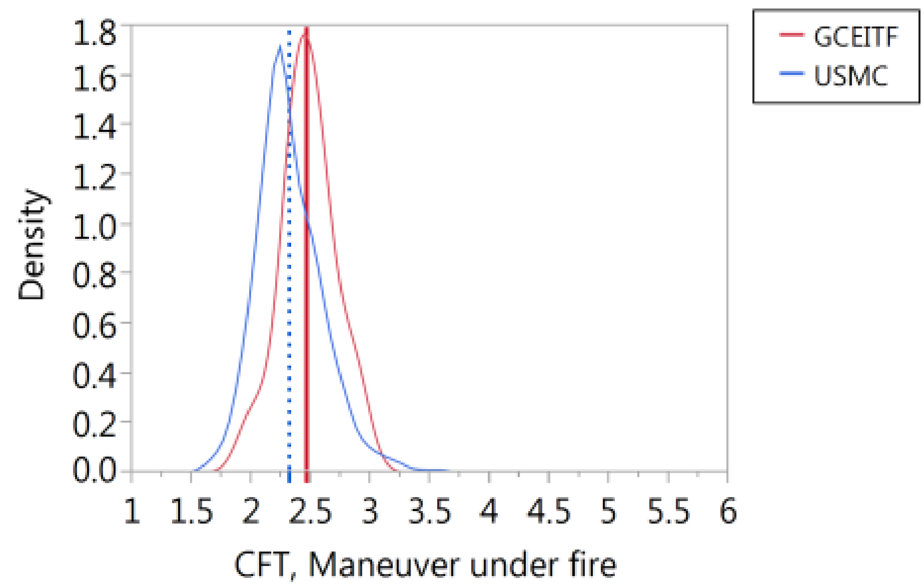


Figure Q-30. CFT MUF by Population for Males MOS 1371

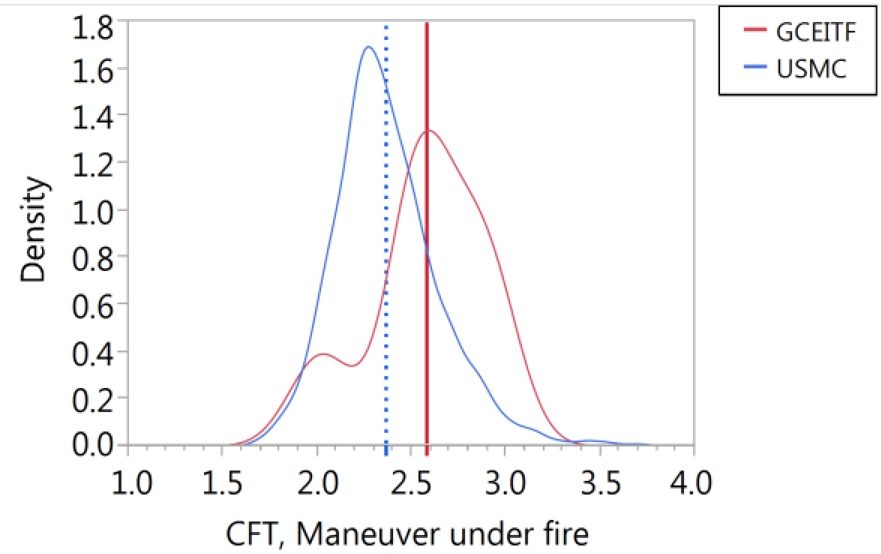
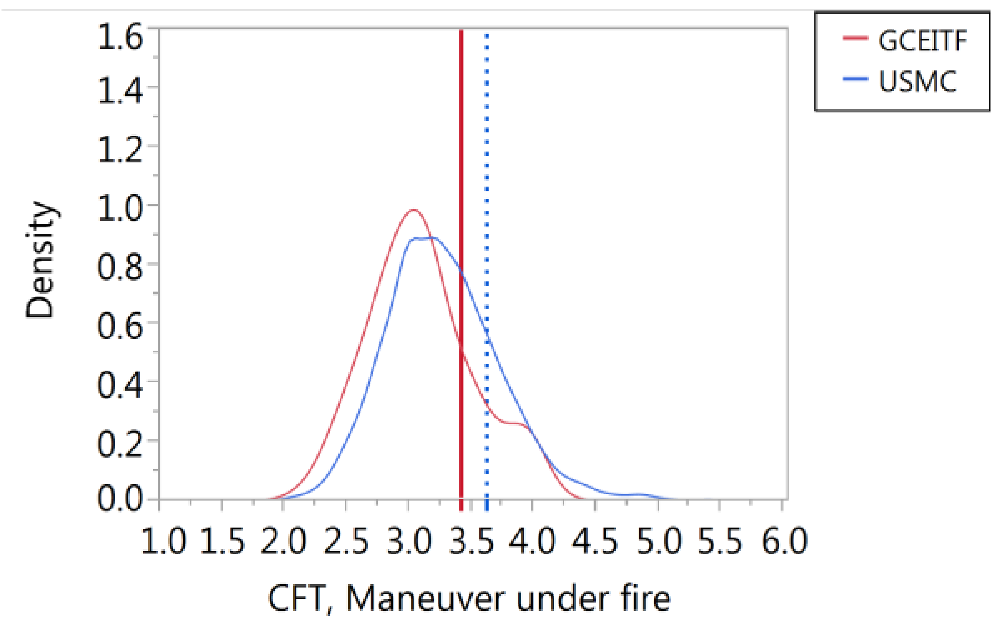


Figure Q-31. CFT MUF by Population for Females



The average CFT MUF time of the GCEITF male was 2.44 minutes and the average CFT MUF time of a male in the Marine Corps Population was 2.4 minutes. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF males' CFT MUF times were statistically significantly slower than the males in the Marine Corps Population times, by approximately 2 seconds. The following MOSs had statistically significant differences:

- 0311 GCEITF males' average CFT MUF time was 2.44 minutes and the average CFT MUF time for males in the Marine Corps 0311 population was 2.32 minutes.

The GCEITF males' average CFT MUF time was statistically significantly slower, by approximately 7 seconds.

- 0341 GCEITF males' average CFT MUF time was 2.48 minutes and the average CFT MUF time for males in the Marine Corps 0341population was 2.32 minutes. The GCEITF males' average CFT MUF time was statistically significantly slower, by approximately 9 seconds.
- 1371 GCEITF males' average CFT MUF time was 2.58 minutes and the average CFT MUF time for males in the Marine Corps 1371population was 2.37 minutes. The GCEITF males' average CFT MUF time was statistically significantly slower, by approximately 12 seconds.

The average CFT MUF time of the GCEITF female was 3.09 minutes and the average CFT MUF time of a female in the Marine Corps Population was 3.27 minutes. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF females' CFT MUF times were statistically significantly faster than for females in the Marine Corps Population, by approximately 11 seconds. There are no statistically significant differences in any of the MOS-to-MOS comparisons.

Q.4.2.11 CFT Movement to Contact (MTC) time

Table Q-21 and Q-22 below display the results for comparing the CFT MTC time of the male and female GCEITF volunteers to that of the Marine Corps population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figure Q-32 through Q-35 are density plots that display the density of calculated values in a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps population.

Table Q-21. Significant Male CFT MTC Time Comparison by MOS

Male													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0341	18	3.08	0.27	2.7	3.0	3.8	1582	2.85	0.29	1.8	2.8	4.4	<0.01*
0811	35	3.02	0.30	2.0	3.1	3.6	1187	2.91	0.30	2.0	2.9	4.0	<0.01*
PI&PMG	54	2.85	0.23	2.3	2.9	3.5	57863	2.94	0.30	1.8	2.9	6.6	0.03*

Table Q-22. GCEITF Female CFT MTC Time Compared to Female Population CFT MTC Time

Female													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	103	3.43	0.30	2.8	3.42	4.3	6497	3.63	0.39	2.0	3.62	9.4	<0.01*

Figure Q-32. CFT MTC by Population for Males MOS 0341

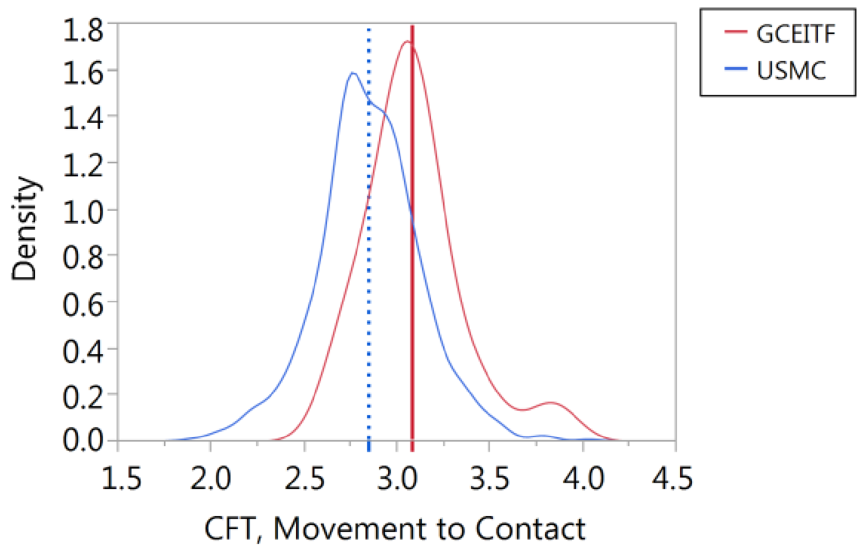


Figure Q-33. CFT MTC by Population for Males MOS 0811

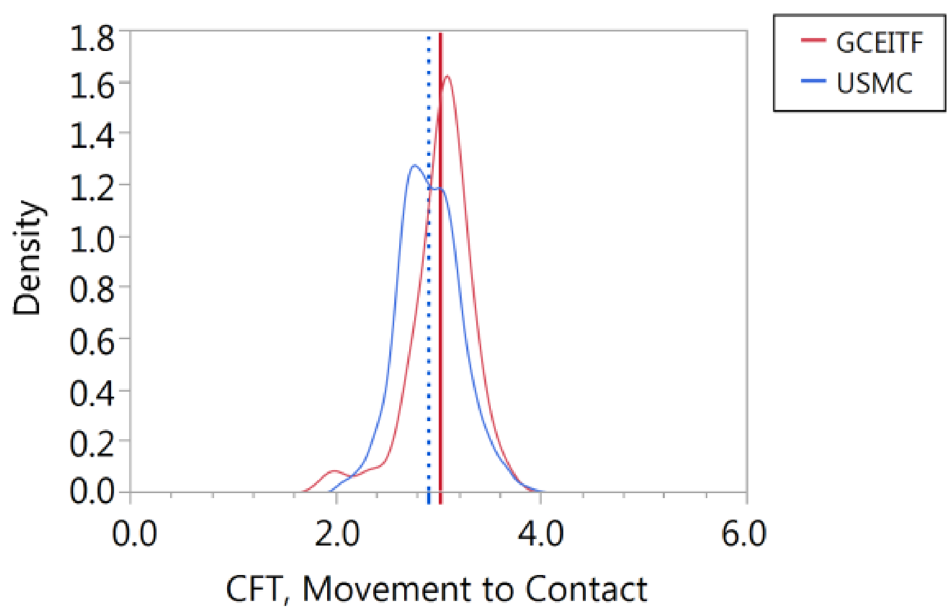


Figure Q-34. CFT MTC by Population for Males PI & PMG

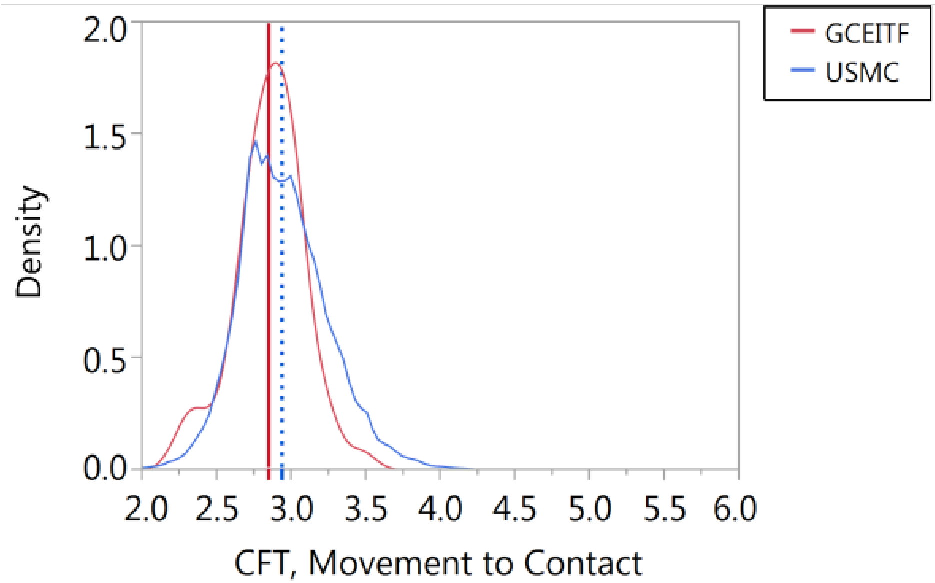
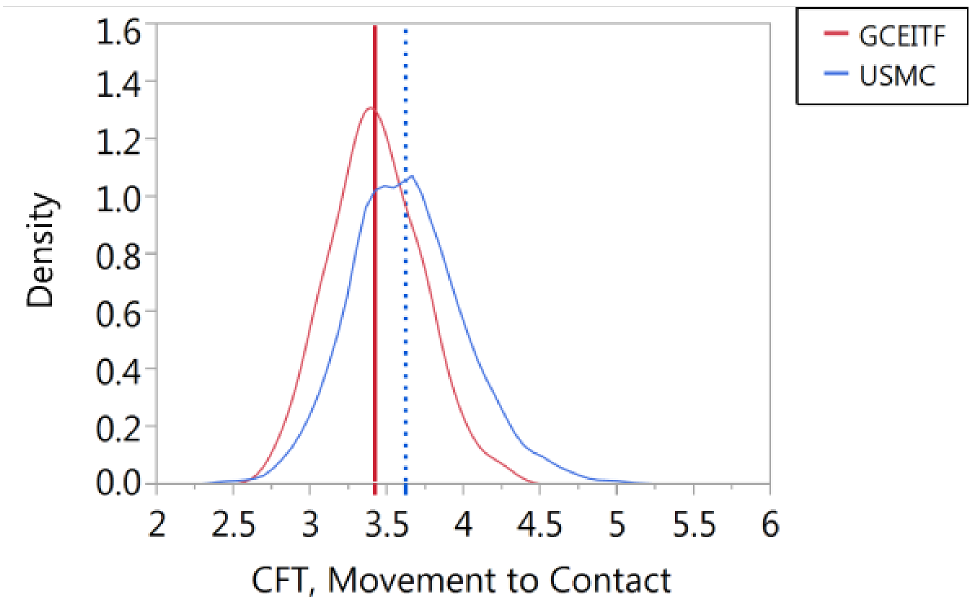


Figure Q-35. CFT MTC by Population for Females



There is no statistically significant difference in CFT MTC times between male Marines in the GCEITF when compared to the males in the Marine Corps population. The following MOSs had statistically significant differences:

- 0341 GCEITF males' average CFT MTC time was 3.08 minutes and the average CFT MTC time for males in the Marine Corps 0341 population was 2.85 minutes. The GCEITF males' average CFT MTC time was statistically significantly slower by approximately 25 seconds.

- 0811 GCEITF males' average CFT MTC time was 3.02 minutes and the average CFT MTC time for males in the Marine Corps 0811 population was 2.91 minutes. The GCEITF males' average CFT MTC time was statistically significantly slower by approximately 7 seconds.
- PI & PMG GCEITF males' average CFT MTC time was 2.85 minutes and the average CFT MTC time for males in the Marine Corps PI & PMG population was 2.94 minutes. The GCEITF males' average CFT MTC time was statistically significantly faster by approximately 5 seconds.

The average CFT MTC time of the GCEITF female was 3.43 minutes and the average CFT MTC time of a female in the Marine Corps Population was 3.63 minutes. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF females' CFT MTC times were statistically significantly faster than for females in the Marine Corps Population, by approximately 12 seconds. There are no statistically significant differences in any of the MOS-to-MOS comparisons.

Q.4.2.12 Rifle Range Score

Table Q-23 and Q-24 below display the results from the comparison of the Rifle Range score of the male GCEITF Volunteers to that of the total Marine Corps Population given the same parameters. The tables display non-parametric KST results and p-values suggesting statistical significance. Figure Q-36 through Q-40 are density plots that display the density of calculated values in a shape. If p-values are less than the a-priori determined significance level of 0.10, we conclude that there is statistical evidence that the response for the GCEITF group is different from that in the total Marine Corps population.

Table Q-23. Significant Male Rifle Range Comparisons by MOS

Male													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	65	303.55	23.27	193.0	311.0	334.0	9759	307.19	24.01	85.0	311.0	348.0	0.04*
0341	18	291.89	35.34	183.0	303.0	323.0	1591	304.92	26.41	116.0	309.0	345.0	0.07*
0811	35	300.17	14.43	250.0	299.0	324.0	1195	304.33	26.38	135.0	309.0	345.0	0.02*
PI&PMG	55	293.38	23.61	179.0	296.0	333.0	58352	306.71	21.66	47.0	310.0	347.0	< 0.01*

Table Q-24. GCEITF Male Rifle Range Compared to Male Population Rifle Range

Male													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
DOR incl	278	299.95	25.10	162.00	303.00	342.00	79189	306.83	22.31	47.00	310.00	349.00	< 0.01*

Figure Q-36. Rifle Range by Population for Males

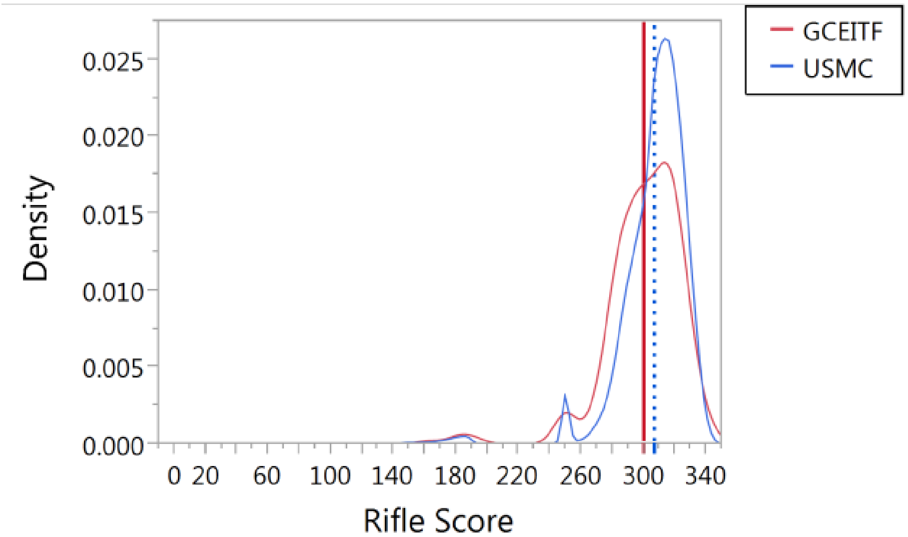


Figure Q-37. Rifle Range by Population for Males MOS 0311

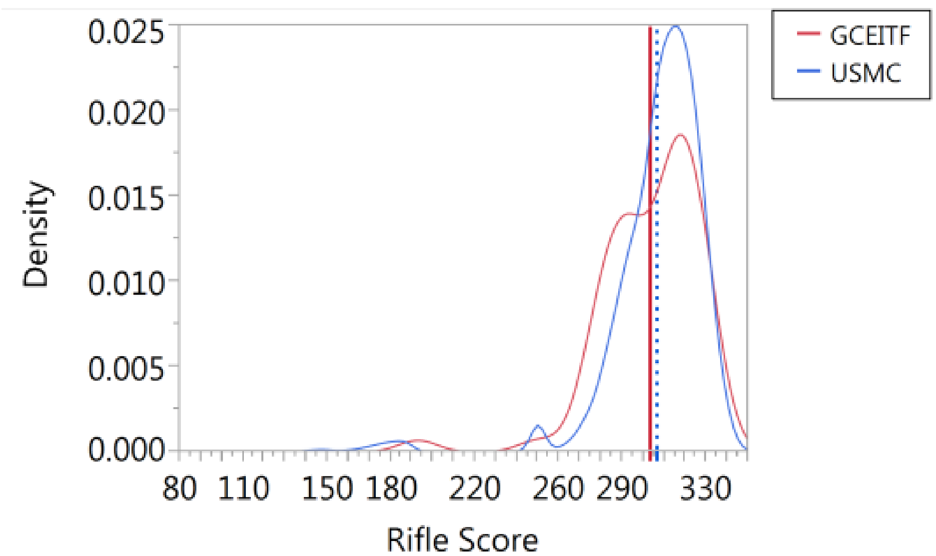


Figure Q-38. Rifle Range by Population for Males MOS 0341

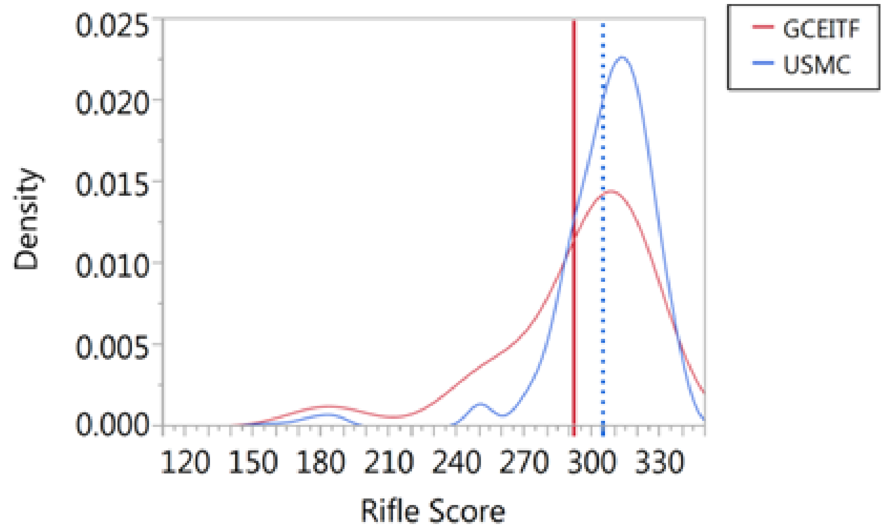


Figure Q-39. Rifle Range by Population for Males MOS 0811

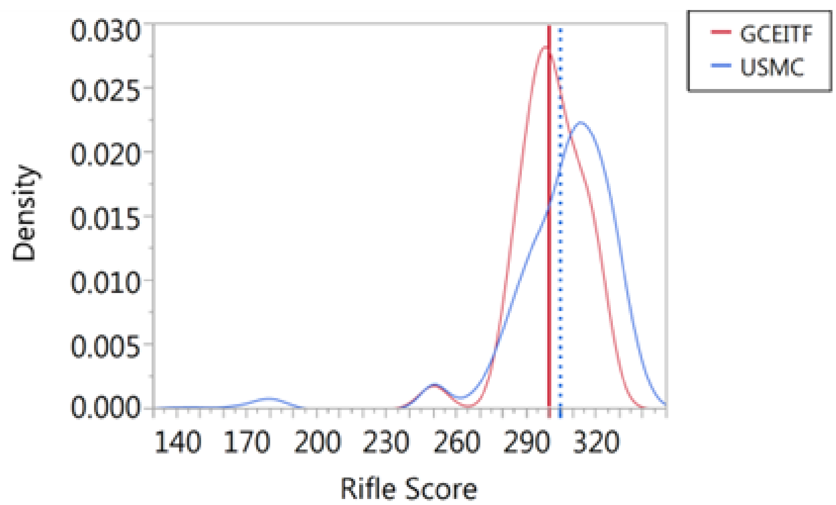
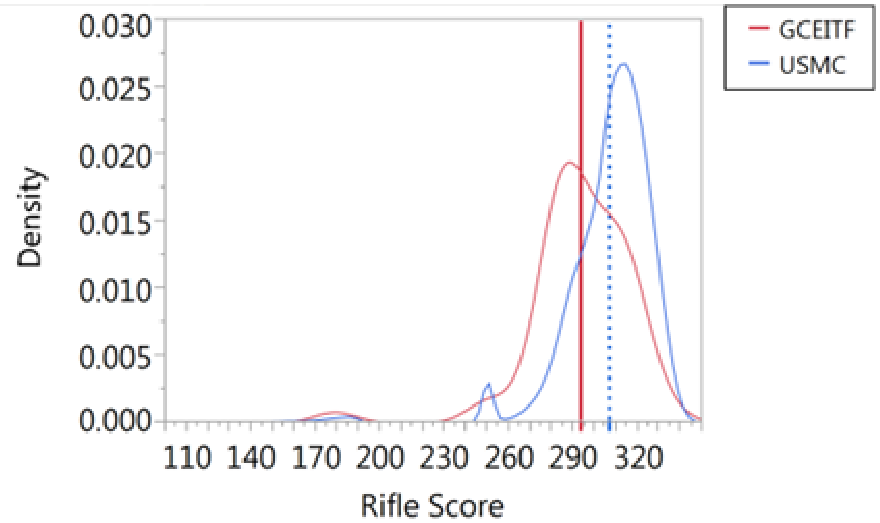


Figure Q-40. Rifle Range by Population for Males MOS PI & PMG



The average Rifle Range score for the GCEITF male was 299.95 and the average Rifle Range score for a male in the Marine Corps Population was 306.83. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF males' Rifle Range scores were statistically significantly less than the total Marine Corps population, by approximately 7 points. The following MOSs had statistically significant differences:

- 0311 GCEITF males' average Rifle Range score was 303.55 and the average Rifle Range score for males in the Marine Corps 0311 population was 307.19. The GCEITF males' average Rifle Range score was statistically significantly lower, by approximately 4 points.
- 0341 GCEITF males' average Rifle Range score was 291.89 and the average Rifle Range score for males in the Marine Corps 0341population was 304.92. The GCEITF males' average Rifle Range score was statistically significantly lower, by approximately 13 points.
- 0811 GCEITF males' average Rifle Range score was 300.17 and the average Rifle Range score for males in the Marine Corps 0811 population was 304.33. The GCEITF males' average Rifle Range score was statistically significantly lower, by approximately 4 points.
- PI and PIMG GCEITF males' average Rifle Range score was 293.38 and the average Rifle Range score for males in the Marine Corps open MOS population was 306.71. The GCEITF males' average Rifle Range score was statistically significantly lower, by approximately 13 points.

There is no statistically significant difference in Rifle Range scores between females in the GCEITF when compared to females in the Marine Corps Population. There are no statistically significant differences in any of the female MOS-to-MOS comparisons.

Q.5 Insights

- Female Marines inclined to volunteer for closed MOSs perform better, on average, when compared to the female Marine Corps population on the following physical tests: overall PFT score, PFT crunches, Overall CFT score, and MUF time.
- Females Combat Engineers (MOS 1371) inclined to volunteer for assignment to closed units perform better, on average, when compared to the female Marine Corps population on the following cognitive and physical tests: AFQT score, overall PFT score, and PFT crunches.

Appendix to Annex Q
GCEITF Population

This appendix describes the population of the GCEITF volunteers when compared to today's Marine Corps Active Component who possess the same parameters as the GCEITF population, i.e. E-1 to E-5s, full duty medical status. The appendix presents the full Descriptive Statistics and interpretive results.

Q.1 GCEITF Population Parameters and Personnel Variables

The ability to generalize the results of the GCEITF experiment to the wider Marine Corps is, in part, dependent upon how representative the GCEITF participants were of the total population of Marines.

The comparisons made between the GCEITF volunteer sample and the Marine Corps total active component population were done using the same parameters to include:

- The total Marine Corp's population was defined using the Active Component whereas the GCEITF sample contains both Active and Reserve Component Marines. The decision to use the Marine Corp's Active Component as the basis for comparison was a conservative one that embodies the idea that Reserve Marines are equal to Active Marines. This assumption is reasonable because Reserve Marines are required to maintain Marine Corps standards and could be activated at any time to serve on a continuous basis with their Active Duty counterparts. This approach is conservative due to the most likely direction of error in the assumption. If reserve Marines were believed to be inferior to active Marines, then the basis of comparison (the active Marine Corps) would be superior to the GCEITF (source from the active and reserve components). Thus, if this analysis concludes that the sample (GCEITF) was representative of the population (the active component), then the sample will have attained the higher standard to which the experiment's results could be generalized.
- Pay grade of E-5 and below
- Full-duty status (i.e., not currently on limited/light duty, or pending physical evaluation board)
- For MOS to MOS comparisons, Marines must hold the same Primary MOS (PMOS). MOS-MOS comparisons were done by gender (ie. Female 1371 Combat Engineers within the GCEITF were compared to female Marines who also possess the 1371 PMOS within the total Marine Corps given the same parameters; Male PI and PIMG Marines were compared to male open MOS Marines within the total Marine Corps given the same parameters)

From this point on, this annex will refer to the Marine Corps' total Active Component population used for comparison as the total Marine Corps Population.

The GCEITF population encompassed all the volunteers who actively participated in the experiment. While the number of volunteers was in constant flux, the maximum active participation by volunteers at any one time was 382 (278 Males/ 104 females).

Furthermore, the comparisons between the GCEITF volunteer sample and the Marine Corps total active component population were based on variables that are common among all Marines regardless of gender. The tables and graphs present the mean of each variable. While the PFT/CFT score used in the comparisons are gender-normed, the component scores of PFT and CFT events that are continuous and/or gender neutral (i.e. run times (continues time captured), crunch score (where 100 is the maximum score regardless of gender) were used to give greater insight into the individual personnel variables of the Marine within the two groups. The 12 personnel variables used include:

- Age
- Height
- Weight
- AFQT
- GT
- PFT, Total Score
- PFT, Crunches
- PFT, Run Time
- CFT, Total Score
- CFT, Movement to Contact
- CFT, Maneuver under Fire
- Rifle Range Score

All comparisons given for the below personnel variables are done so between the GCEITF population and the Marine Corps population in the aggregate and by PMOS. All GCEITF Marines who maintain open MOSs and performed the Provisional Infantry (PI) and Provisional Machine Gun (PMG) were compared to the total Marine Corps Population of currently Open MOSs. For each of the 12 personnel variables, a total of 15 comparisons were made. Males were compared to males in the aggregate (GCEITF to Total Marine Corps Population) and by the 9 closed MOSs, by MOS 1371, and by PI & PMG to open MOSs. Females were compared to females in the aggregate (GCEITF to Total Marine Corps Population), by MOS 1371, and then by PI & PMG to open MOSs. A total of 180 comparisons were made.

Q.1.1 Age

This personnel variable defines the age, specified in years, of the population.

Q.1.1.1 Age (Male Marines)

Table Q-1 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-1. Male Age Comparisons

Male Age													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	278	22.45	2.45	18.80	21.85	31.60	79212	23.03	3.03	17.40	22.30	47.10	0.02*

Male Age													
MOS (w/ DOR)	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						131	22.84	3.27	17.40	22.70	32.50	
0311	65	22.78	2.80	18.80	22.10	31.60	9766	22.30	2.77	18.10	21.70	47.10	0.15
0313	14	23.29	2.05	20.60	22.55	27.60	447	22.96	2.74	18.80	22.40	31.60	0.54
0321	0						318	23.86	2.95	18.70	23.60	36.00	
0331	13	22.03	2.59	19.50	21.00	28.40	1550	22.83	3.09	18.50	22.10	36.70	0.62
0341	18	22.77	2.27	20.40	22.10	27.80	1591	22.60	3.09	18.40	21.80	38.70	0.13
0351	4	23.80	2.74	19.80	24.70	26.00	713	22.67	3.11	18.60	21.80	37.90	0.26
0352	16	21.71	1.61	20.00	21.25	25.60	790	23.35	3.24	18.50	22.60	36.30	0.07*
0372	0						138	27.32	2.51	22.80	27.00	36.80	
0800	0						60	21.93	2.84	18.50	21.80	30.80	
0811	35	22.63	2.62	18.90	21.90	28.80	1195	22.69	2.93	18.50	22.00	35.50	0.96
0842	0						168	21.99	2.38	18.80	21.25	29.80	
0844	0						324	22.66	2.92	18.40	22.00	38.50	
0847	0						121	22.53	2.38	19.20	22.20	31.70	
0861	0						250	22.88	2.55	18.60	22.45	30.70	
1371	19	22.78	3.29	19.10	21.90	30.00	1890	22.61	3.13	18.50	21.75	40.30	0.99
1800	0						24	20.27	1.98	18.50	19.50	25.70	
1812	19	23.12	2.08	19.80	22.90	27.20	379	22.87	2.78	18.70	22.00	34.20	0.05*
1833	20	21.99	2.00	19.10	21.60	26.30	906	23.10	3.18	18.40	22.20	39.80	0.30
8011	0						18	20.43	2.35	18.60	19.60	28.20	
8972	0						65	22.11	2.35	18.60	21.50	31.80	
PI&PMG	55	21.69	2.01	19.40	20.90	28.00	58368	23.18	3.04	17.70	22.50	44.40	< 0.01*

Q.1.1.1.1 Age (Male Marines) Results

The average age of the GCEITF male Marines was 22.45 years old compared to the average age of male Marines within the total Marine Corps Population which is 23.03 years old. Based on the p-value of the K-S test being less than .10, this age difference is statistically significant. The GCEITF male Marines were statistically significantly younger than the USMC total population by approximately six months. Additionally, the following MOSs had statistically significant differences:

- 0352 GCEITF male Marines average age was 21.71 years old compared to the average Marine Corps 0352 population which is 23.35 years old. The GCEITF Marines were statistically significantly younger, by approximately twenty months.
- 1812 GCEITF male Marines average age was 23.12 years old compared to the average Marine Corps 1812 population which is 22.87 years old. The GCEITF Marines were statistically significantly older, by approximately 3 months.
- PI & PIMG GCEITF male Marines average age was 21.69 years old compared to the average Marine Corps open MOS population which is 23.18 years old. The GCEITF Marines were statistically significantly younger, by approximately eighteen months.

Q.1.1.2 Age (Female Marines)

Table Q-2 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table Q-2 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population

Table Q-2. Female Age Comparisons

All MOS	Female Age												KS Test
	GCEITF						USMC						
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	104	22.69	2.93	18.4	21.95	33.9	6631	22.77	2.87	18.1	22.10	38.9	0.77

MOS	Female Age												KS Test
	GCEITF						USMC						
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	21.28	2.38	18.9	20.40	25.8	0						
0313	7	19.57	0.93	18.4	19.50	20.8	0						
0331	9	23.19	2.82	20.1	22.70	28.4	0						
0341	8	23.38	4.36	21.2	21.60	33.9	0						
0351	3	21.47	1.56	19.8	21.70	22.9	0						
0352	4	24.60	2.72	21.8	24.25	28.1	0						

Female Age													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0811	15	23.99	2.86	20.0	24.10	28.7	0						
1371	9	22.32	2.53	19.1	22.20	26.6	124	22.06	2.57	18.5	21.40	29.2	0.53
1812	4	23.70	1.53	21.9	23.85	25.2	0						
1833	12	22.93	2.72	19.6	22.80	27.9	0						
8000	0						1	22.40		22.4	22.40	22.4	
8011	0						5	20.52	3.41	18.7	19.20	26.6	
8972	0						8	21.83	3.18	18.5	20.50	27.2	
PI&PMG	21	22.79	3.11	19.0	21.60	30.1	6491	22.79	2.88	18.1	22.10	38.9	0.98

Q.1.1.2.1 Age (Female Marines) Results

The average age of the GCEITF female Marines was 22.69 years old compared to the average age of female Marines within the total Marine Corps Population which is 22.77 years old. Based on the p-value of the K-S test being greater than .10, there is no statistically significant difference between the two groups. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.2 Height

This personnel variable defines the height, specified in inches, of the population.

Q.1.2.1 Height (Male Marines)

Table Q-3 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-3. Male Height Comparisons

Male Height													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	278	69.58	2.56	64.00	69.00	77.00	79197	69.49	2.72	58.00	69.00	84.00	0.98

Male Height													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						131	69.59	2.70	62	70.0	77	
0311	65	69.15	2.62	64	69.0	75	9765	69.47	2.69	60	69.0	80	0.99
0313	14	70.07	2.87	65	69.5	76	447	69.50	2.70	63	69.0	77	1.00
0321	0						318	70.53	2.35	62	70.0	78	
0331	13	71.69	2.56	68	71.0	76	1550	70.42	2.67	59	71.0	80	0.74
0341	18	68.83	2.20	65	69.5	73	1589	69.49	2.57	59	69.0	79	0.24

All MOS	Male Height												KS Test
	GCEITF						USMC						
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0351	4	69.50	3.32	67	68.5	74	713	69.90	2.69	62	70.0	80	0.84
0352	16	69.31	2.02	66	69.5	73	789	69.98	2.58	63	70.0	79	0.87
0372	0						136	71.10	2.40	65	71.0	77	
0800	0						60	69.07	2.97	62	69.0	76	
0811	35	69.89	2.69	65	70.0	77	1194	69.21	2.72	60	69.0	78	0.86
0842	0						168	69.84	2.69	65	70.0	78	
0844	0						324	69.65	2.87	59	69.0	77	
0847	0						121	69.75	2.84	63	70.0	76	
0861	0						250	69.76	2.44	63	70.0	78	
1371	19	69.68	2.26	65	69.0	74	1890	69.53	2.74	61	70.0	79	0.51
1800	0						24	69.17	2.78	63	70.0	74	
1812	19	69.42	2.80	64	70.0	76	379	69.37	2.81	62	69.0	77	1.00
1833	20	69.75	2.38	65	70.0	73	906	69.52	2.81	61	69.0	79	0.96
8011	0						18	68.33	2.00	64	68.5	71	
8972	0						65	70.06	2.57	65	70.0	75	
PI&PMG	55	69.55	2.51	64	69.0	75	58360	69.45	2.73	58	69.0	84	0.94

Q.1.2.1.1 Height (Male Marines) Results

The average height of the GCEITF male Marines was 69.58 inches tall compared to the average height of male Marines within the total Marine Corps Population which is 69.49 inches tall. Based on the p-value of the K-S test being greater than .10, there is no statistically significant difference between the two groups. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.2.2 Height (Female Marines)

Table Q-4 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 4 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-4. Female Height Comparisons

All MOS	Female Height												KS Test
	GCEITF						USMC						
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	104	64.40	2.31	59.00	64.00	71.00	6628	64.19	2.56	56.00	64.00	75.00	0.60

Female Height													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	65.67	2.46	62	65.5	71	0						
0313	7	64.14	1.46	63	64.0	67	0						
0331	9	65.44	2.24	63	66.0	69	0						
0341	8	64.13	3.76	59	64.0	71	0						
0351	3	63.00	1.00	62	63.0	64	0						
0352	4	63.25	1.50	62	63.0	65	0						
0811	15	64.27	2.15	59	64.0	67	0						
1371	9	64.11	1.76	60	65.0	66	124	64.17	2.21	59	64.0	70	0.98
1812	4	64.75	2.22	62	65.0	67	0						
1833	12	64.92	2.54	62	64.5	70	0						
8000	0						1	62.00		62	62.0	62	
8011	0						5	64.00	3.00	61	63.0	69	
8972	0						8	64.63	1.60	63	64.5	67	
PI&PMG	21	63.71	2.12	59	64.0	67	6488	64.19	2.57	56	64.0	75	0.67

Q.1.2.2.1 Height (Female Marines) Results

The average height of the GCEITF female Marines was 64.40 inches compared to the average height of female Marines within the total Marine Corps Population, which is 64.19 inches. Based on the p-value of the K-S test being greater than 0.10, there is no statistically significant difference between the two groups. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.3 Weight

This personnel variable defines the weight, specified in pounds, of the population.

Q.1.3.1 Weight (Male Marines)

Table Q-5 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-5. Male Weight Comparisons

Male Weight													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	278	169.79	22.92	117.00	170.00	246.00	79198	172.36	23.46	92.00	172.00	295.00	0.08*

Male Weight													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						131	170.99	23.02	113.0	170.0	228.0	
0311	65	168.06	23.69	129.0	166.0	220.0	9765	170.09	22.97	92.0	169.0	280.0	0.99
0313	14	170.86	22.19	132.0	173.0	222.0	447	171.51	22.51	110.0	170.0	244.0	1.00
0321	0						318	179.09	19.23	124.0	180.0	235.0	
0331	13	184.62	31.64	142.0	185.0	246.0	1550	182.17	22.92	112.0	182.0	281.0	0.74
0341	18	165.94	22.35	124.0	168.0	211.0	1590	170.66	22.55	104.0	170.0	265.0	0.27
0351	4	156.00	20.22	129.0	161.5	172.0	713	172.27	23.00	110.0	171.0	250.0	0.84
0352	16	169.50	29.92	127.0	165.0	239.0	789	174.19	22.46	120.0	175.0	265.0	0.87
0372	0						136	188.74	18.59	148.0	189.0	246.0	
0800	0						60	170.12	24.13	112.0	177.5	208.0	
0811	35	173.49	21.50	117.0	171.0	222.0	1194	170.77	23.59	108.0	171.0	271.0	0.86
0842	0						168	169.76	24.31	107.0	166.5	229.0	
0844	0						324	171.12	23.96	113.0	170.0	286.0	
0847	0						121	175.02	27.08	109.0	175.0	270.0	
0861	0						250	173.58	23.59	104.0	174.0	271.0	
1371	19	170.05	25.65	123.0	172.0	222.0	1890	171.75	24.54	102.0	171.0	288.0	0.51
1800	0						24	157.96	24.72	123.0	155.5	219.0	
1812	19	168.53	19.34	141.0	163.0	204.0	379	170.43	23.22	100.0	170.0	259.0	1.00
1833	20	166.25	14.96	128.0	167.5	196.0	906	171.58	23.82	106.0	171.0	277.0	0.96
8011	0						18	161.89	19.75	138.0	158.0	209.0	
8972	0						65	172.98	24.12	129.0	174.0	223.0	
PI&PMG	55	169.67	21.36	131.0	171.0	211.0	58360	172.53	23.48	97.0	172.0	295.0	0.94

Q.1.3.1.1 Weight (Male Marines) Results

The average weight of the GCEITF male Marines was 169.79 lb compared to the average weight of male Marines within the total Marine Corps Population which is 172.36 lb. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF male Marines were statistically significantly lighter than the USMC total population, by approximately 2.5lbs. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.3.2 Weight (Female Marines)

Table Q-6 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 6 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-6. Female Weight Comparisons

All MOS	Female Weight												KS Test
	GCEITF						USMC						
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	104	135.33	14.39	105.0	135.00	174.0	6628	136.02	16.33	90.0	135.00	249.0	0.74

MOS	Female Weight												KS Test
	GCEITF						USMC						
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	142.17	17.21	112	147.0	173	0						
0313	7	137	7.59	121	140.0	143	0						
0331	9	145.22	13.42	126	139.0	165	0						
0341	8	130.88	20.77	110	124.0	174	0						
0351	3	132.67	13.65	118	135.0	145	0						
0352	4	133.25	1.89	132	132.5	136	0						
0811	15	137.8	16.17	106	142.0	166	0						
1371	9	130.11	17.81	110	132.0	157	124	135.60	16.44	90	134.5	186	0.98
1812	4	134.75	8.02	127	133.5	145	0						
1833	12	132.08	11.66	110	134.0	155	0						
8000	0						1	109.00		109	109.0	109	
8011	0						5	128.20	8.32	119	129.0	139	
8972	0						8	133.25	13.17	104	134.5	146	
PI&PMG	21	131.52	11.29	105	131.0	149	6488	136.05	16.33	90	135.0	249	0.67

Q.1.3.2.1 Weight (Female Marines) Results

The average weight of the GCEITF female Marines was 135.33 lb compared to the average weight of female Marines within the total Marine Corps Population, which is 136.02 lb. Based on the p-value of the K-S test being greater than 0.10, there is no statistically significant difference between the two groups. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.4 Armed Forces Qualification Test (AFQT)

This personnel variable defines AFQT as a score specified as a numerical value.

The Armed Forces Qualification Test, or AFQT, consists of the following four sections from the Armed Services Vocational Aptitude Battery (ASVAB): Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mathematics Knowledge. The scores from these four sections make up the AFQT.

Q.1.4.1 AFQT (Male Marines)

Table Q-7 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-7. Male AFQT Comparisons

Male AFQT													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	277	60.56	19.19	22.00	59.00	99.00	79199	62.63	17.79	4.00	62.00	99.00	0.03*

Male AFQT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD		Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						131	65.58	19.01	31.0	63.0	99.0	
0311	65	58.78	18.89	31.0	55.0	96.0	9766	59.69	17.73	4.0	59.0	99.0	0.69
0313	14	70.64	20.59	38.0	77.0	95.0	447	59.72	17.74	23.0	56.0	99.0	0.05*
0321	0						318	75.32	14.09	36.0	77.0	99.0	
0331	12	69.08	20.43	32.0	75.5	90.0	1549	61.25	17.33	22.0	61.0	99.0	0.13
0341	18	58.11	21.03	27.0	52.0	92.0	1591	59.62	17.54	16.0	57.0	99.0	0.66
0351	4	74.50	19.05	46.0	83.0	86.0	713	73.16	16.46	31.0	74.0	99.0	0.52
0352	16	74.06	14.27	51.0	77.5	95.0	790	73.06	16.07	23.0	74.0	99.0	1.00
0372	0						138	74.46	13.69	39.0	75.0	99.0	
0800	0						60	57.07	15.10	35.0	54.5	93.0	
0811	35	57.77	18.91	31.0	59.0	99.0	1195	54.05	16.82	16.0	51.0	99.0	0.33
0842	0						168	69.82	13.92	36.0	70.0	97.0	
0844	0						324	69.35	14.31	31.0	69.0	99.0	
0847	0						121	70.76	14.56	36.0	72.0	99.0	
0861	0						250	66.53	13.95	33.0	67.0	99.0	
1371	19	62.26	18.73	35.0	62.0	89.0	1890	59.86	16.67	21.0	59.0	99.0	0.66
1800	0						24	50.33	15.31	32.0	48.5	90.0	
1812	19	51.42	18.88	22.0	50.0	94.0	379	53.83	17.80	10.0	51.0	97.0	0.89
1833	20	61.90	18.27	31.0	69.0	89.0	906	54.93	16.82	20.0	52.0	98.0	0.04*
8011	0						18	64.39	16.37	36.0	66.5	87.0	
8972	0						65	77.32	12.13	46.0	81.0	97.0	
PI&PMG	55	57.96	17.86	28.0	54.0	95.0	58356	63.24	17.73	9.0	63.0	99.0	0.04*

Q.1.4.1.1 AFQT (Male Marines) Results

The average AFQT score of the GCEITF male Marines was 60.56 compared to the average AFQT score of male Marines within the total Marine Corps population which is 62.63. Based on the p-value of the K-S test being less than .10, this difference is statistically significant. The GCEITF male Marines had lower scores that were statistically significant when compared to the USMC total population, by approximately 2 points. Additionally, the following MOSs had statistically significant differences:

- 0313 GCEITF male Marines average AFQT score was 70.64 compared to the average Marine Corps 0313 population which had average scores of 59.72. The GCEITF Marine' AFQT scores were statistically significantly greater by approximately 11 points.
- 1833 GCEITF male Marines average AFQT score was 61.90 compared to the average Marine Corps 1833 population which had average scores of 54.93 The GCEITF Marines' AFQT were statistically significantly greater, by approximately 7 points.
- PI & PIMG GCEITF male Marines average AFQT score was 57.96 compared to the average Marine Corps open MOS population which had average scores of 63.24. The GCEITF Marines' AFQT scores were is statistically significantly less, by approximately five points.

Q.1.4.2 AFQT (Female Marines)

Table Q-8 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 8 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-8. Female AFQT Comparisons

Female AFQT													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	104	61.66	19.22	10	63	99.0	6630	59.00	17.12	7	57	99	0.13

Female AFQT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	59.67	24.82	10	65.5	92	0						
0313	7	54.00	20.61	35	51	96	0						
0331	9	58.89	20.05	22	64	81	0						
0341	8	61.88	16.59	35	61	93	0						
0351	3	69.67	6.66	64	68	77	0						
0352	4	78.00	10.42	70	74.5	93	0						
0811	15	58.67	22.25	31	54	99	0						
1371	9	74.00	17.73	39	75	92	124	60.56	16.86	33	59.5	95	0.04*
1812	4	71.75	10.24	57	75	80	0						

1833	12	63.75	16.24	38	62	93	0						
8000	0						1	33.00		33	33	33	
8011	0						5	60.40	18.43	43	55	85	
8972	0						8	79.38	4.34	71	81.5	82	
PI&PMG	21	55.95	17.39	32	53	91	6490	58.95	17.12	7	57	99	0.98

Q.1.4.2.1 AFQT (Female Marines) Results

The average AFQT score of the GCEITF female Marines was 61.66 compared to the average AFQT score of female Marines within the total Marine Corps Population which is 59.00. Based on the p-value of the K-S test being greater than .10, there is no statistically significant difference between the two groups. However, the following MOS had statistically significant difference:

- 1371 GCEITF female Marines average AFQT score was 74.00 compared to the average female Marine Corps 1371 population which is 60.56. The GCEITF Marines' AFQT scores were statistically significantly greater by approximately 13 points.

Q.1.5 General Classification Test (GT)

This personnel variable defines GT as a score specified as a numerical value.

Q.1.5.1 GT (Male Marines)

Table Q-9 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-9. Male GT Comparisons

Male GT													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	272	106.79	12.17	80.00	106.00	136.00	79207	108.99	11.73	12	108	159	< 0.01*

Male GT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						131	111.80	12.86	86	111	151	
0311	62	105.74	12.18	80	103.0	133	9766	106.91	11.67	58	106	148	0.44
0313	14	113.71	12.37	95	113.0	133	447	107.56	11.27	81	106	148	0.30
0321	0						318	118.53	9.48	90	118	148	
0331	13	110.62	13.99	83	114.0	129	1550	108.32	11.36	80	108	150	0.31
0341	18	104.78	13.09	80	107.5	127	1591	106.90	11.46	80	106	154	0.62
0351	4	113.00	9.31	100	115.0	122	713	116.18	11.50	83	116	148	0.88

0352	16	115.63	7.03	104	115.0	129	790	116.31	10.57	89	116	148	0.74
0372	0						138	118.37	9.56	102	117	151	
0800	0						60	104.68	9.19	89	103	131	
0811	34	103.82	11.33	88	103.5	132	1195	103.47	10.56	77	101	148	0.64
0842	0						168	113.60	8.15	100	112	138	
0844	0						324	113.69	8.13	90	111	142	
0847	0						121	115.84	9.28	102	114	150	
0861	0						250	111.68	8.21	97	110	137	
1371	18	110.17	9.76	93	109.0	126	1890	108.13	10.03	83	107	145	0.96
1800	0						24	100.67	9.12	89	98	126	
1812	19	102.11	10.34	91	100.0	126	379	103.16	11.72	64	101	147	1.00
1833	20	109.05	12.72	89	107.0	130	906	104.26	10.64	82	102	139	0.18
8011	0						18	108.17	10.13	93	110	125	
8972	0						65	119.82	7.67	104	120	139	
PI&PMG	54	104.44	12.47	81	102.0	136	58363	109.33	11.73	12	108	159	0.01*

Q.1.5.1.1 GT (Male Marines) Results

The average GT of the GCEITF male Marines was 106.79 compared to the average GT of male Marines within the total Marine Corps population which is 108.99. Based on the p-value of the K-S test being less than .10, this difference is statistically significant. The GCEITF male Marines scored lower and the difference is statistically significant compared to the USMC total population, by approximately 2.2 points. Additionally, the following MOS had a statistically significant difference:

- PI & PIMG GCEITF male Marines average GT score was 104.44 compared to the average Marine Corps open MOS population which is 109.33. The GCEITF Marines' GT scores were is statistically significantly lower by approximately five points.

Q.1.5.2 GT (Female Marines)

Table Q-10 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 10 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-10. Female GT Comparisons

Female GT													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	97	104.16	12.58	74.0	104.00	134.0	6631	102.58	10.84	68	102	146	0.20

Female GT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	9	94.00	13.48	74	92	120	0						
0313	7	99.57	13.34	90	95	126	0						
0331	7	105.57	11.33	83	106	119	0						
0341	8	104.00	10.84	89	104.5	127	0						
0351	3	103.33	4.04	101	101	108	0						
0352	4	113.25	5.38	108	112.5	120	0						
0811	15	103.60	13.52	90	98	134	0						
1371	8	113.50	10.73	95	113	127	124	105.85	9.45	90	105	133	0.19
1812	4	114.75	5.12	108	115.5	120	0						
1833	12	108.83	11.98	92	105.5	131	0						
8000	0						1	92.00		92	92	92	
8011	0						5	100.20	7.95	92	98	110	
8972	0						8	117.38	5.58	111	116	127	
PI&PMG	20	100.00	12.05	83	102	120	6491	102.50	10.85	68	102	146	0.41

Q.1.5.2.1 GT (Female Marines) Results

The average GT of the GCEITF female Marines was 104.16 compared to the average GT score of female Marines within the total Marine Corps Population which is 102.58. Based on the p-value of the K-S test being greater than .10, there is no statistically significant difference between the two groups. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.6 Physical Fitness Test (PFT) Score

This personnel variable defines PFT as a score specified as a numerical value.

The PFT score is a composite score of three events: pull-ups (pull-ups or flexed-arm hang for females), abdominal crunches, and a 3-mile run. The score is gender-normed with a maximum of 300 points awarded to both male and female Marines.

Q.1.6.1 PFT (Male Marines) Score

Table Q-11 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-11. Male PFT Comparisons

Male PFT													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	273	251.79	29.27	136.00	255.00	300.00	77991	253.55	29.67	116.00	259.00	300.00	0.29

Male PFT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						131	265.91	28.26	175.0	276.0	300.0	
0311	63	258.94	27.59	173.0	266.0	297.0	9670	258.91	26.77	136.0	264.0	300.0	1.00
0313	14	246.57	44.89	136.0	263.5	292.0	441	256.87	27.34	170.0	263.0	300.0	0.65
0321	0						316	279.42	17.34	203.0	285.0	300.0	
0331	13	237.00	27.61	181.0	243.0	281.0	1532	263.84	25.69	159.0	271.0	300.0	< 0.01*
0341	17	256.41	24.73	209.0	258.0	290.0	1569	259.59	26.62	151.0	264.0	300.0	0.91
0351	3	263.33	17.56	245.0	265.0	280.0	708	258.50	28.01	139.0	265.0	300.0	0.99
0352	16	267.50	18.10	240.0	264.5	300.0	784	259.08	26.07	155.0	264.0	300.0	0.48
0372	0						136	282.21	12.54	228.0	284.0	300.0	
0800	0						60	254.55	24.10	186.0	256.0	292.0	
0811	35	239.06	34.67	165.0	246.0	294.0	1179	254.45	27.38	136.0	259.0	300.0	< 0.01*
0842	0						166	247.72	30.66	145.0	250.0	298.0	
0844	0						322	250.18	32.13	139.0	256.0	300.0	
0847	0						120	245.72	29.53	172.0	251.0	294.0	
0861	0						244	254.86	29.39	157.0	261.0	300.0	
1371	19	244.58	34.09	194.0	247.0	300.0	1876	255.41	28.03	138.0	261.0	300.0	0.09*
1800	0						24	258.29	29.14	195.0	262.5	297.0	
1812	18	242.17	29.05	181.0	243.5	286.0	375	252.00	27.33	169.0	256.0	297.0	0.32
1833	20	258.15	21.32	223.0	257.0	293.0	898	254.74	29.72	137.0	260.0	300.0	0.84
8011	0						18	270.67	19.40	232.0	278.0	293.0	
8972	0						65	249.97	27.25	170.0	248.0	292.0	
PI&PMG	55	253.24	23.58	189.0	254.0	299.0	57357	251.76	30.26	116.0	257.0	300.0	0.46

Q.1.6.1.1 PFT Score (Male Marines) Results

The average PFT score of the GCEITF male Marines was 251.79 compared to the average PFT score of male Marines within the total Marine Corps Population which is 253.55. Based on the p-value of the K-S test being greater than .10, this difference is not statistically significant. However, the following MOSs had statistically significant differences:

- 0331 GCEITF male Marines average PFT score was 237 compared to the average Marine Corps 0331 population which is 263.84. The GCEITF Marines' average PFT scores were statistically significantly lower by approximately 27 points.
- 0811 GCEITF male Marines average PFT score was 239.06 compared to the average Marine Corps 0811 population which is 254.45. The GCEITF Marines'

average PFT scores were statistically significantly lower by approximately 15 points.

- 1371 ITF male Marines average PFT score was 244.58 compared to the average male Marine Corps 1371 population which is 255.41. The GCEITF Marines' average PFT scores were statistically significantly lower by approximately 11 points.

Q.1.6.2 PFT (Female Marines) Score

Table Q-12 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 12 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-12. Female PFT Comparisons

Female PFT													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	104	275.72	20.03	211.00	280.00	300.00	6419	255.51	31.36	125	263	300	< 0.01*

Female PFT													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	285.67	13.79	254	288	300	0						
0313	7	273.86	23.37	225	281	293	0						
0331	9	284.78	8.50	272	285	300	0						
0341	8	285.13	8.68	271	287.5	295	0						
0351	3	288.33	12.58	275	290	300	0						
0352	4	286.50	12.92	274	286	300	0						
0811	15	270.47	20.00	238	274	300	0						
1371	9	281.89	14.59	253	285	300	123	259.95	31.18	160	266	300	0.08*
1812	4	281.50	18.27	259	283.5	300	0						
1833	12	273.67	18.66	238	276.5	300	0						
8000	0						1	205.00		205	205	205	
8011	0						5	252.00	27.40	204	264	269	
8972	0						8	266.13	11.37	254	266	287	
PI&PMG	21	260.52	25.01	211	264	300	6280	255.42	31.38	125	262	300	0.83

Q.1.6.2.1 PFT Score (Female Marines) Results

The average PFT score of the GCEITF female Marines was 275.72 compared to the average PFT score of female Marines within the total Marine Corps Population which is

255.51. Based on the p-value of the K-S test being greater than .10, this difference is statistically significant. The GCEITF female Marines' PFT scores were statistically significantly higher than the USMC total population by approximately 20 points. Additionally, the following MOS had a statistically significant difference:

- 1371 GCEITF female Marines' average PFT score was 281.89 compared to the average female Marine Corps 1371 population which is 259.95. The GCEITF Marines' average PFT scores were statistically significantly higher by approximately 22 points.

Q.1.7 Physical Fitness Test (PFT) Crunches

This personnel variable defines PFT crunches as a total count (quantity of crunches) to execute as many crunches as able within a 2 minute time limit during an annual physical fitness test. The PFT crunch event during the PFT is scored the same for male and female Marines. Each crunch is awarded one point with a total score of 100.

Q.1.7.1 PFT (Male Marines) Crunches

Table Q-13 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-13. Male PFT Crunches Comparisons

Male PFT Crunches													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	257	98.63	4.76	65	100	100	77075	98.59	5.39	17.00	100.00	100.00	1.00

Male PFT Crunches													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						130	98.15	7.00	52	100	100	
0311	53	99.89	0.82	94	100	100	9241	99.10	4.30	38	100	100	1.00
0313	14	97.50	9.35	65	100	100	440	99.18	3.74	60	100	100	1.00
0321	0						312	99.48	2.98	63	100	100	
0331	9	98.11	5.67	83	100	100	1458	99.18	3.89	50	100	100	1.00
0341	16	99.25	1.88	94	100	100	1517	99.10	4.61	50	100	100	0.97
0351	3	93.00	12.12	79	100	100	679	99.23	3.73	66	100	100	0.91
0352	16	98.19	5.34	80	100	100	753	99.24	3.96	51	100	100	1.00
0372	0						136	100.00	0.00	100	100	100	
0800	0						58	98.21	4.61	83	100	100	
0811	34	99.09	3.78	80	100	100	1174	99.12	4.14	54	100	100	1.00
0842	0						167	98.41	4.68	74	100	100	
0844	0						317	98.80	4.75	64	100	100	

0847	0						121	97.96	6.33	64	100	100	
0861	0						233	98.46	5.88	54	100	100	
1371	18	98.11	4.83	85	100	100	1796	99.00	4.32	56	100	100	0.86
1800	0						24	97.04	7.15	76	100	100	
1812	19	97.47	6.59	75	100	100	370	98.41	5.92	50	100	100	0.93
1833	20	99.25	3.35	85	100	100	873	98.60	5.69	50	100	100	1.00
8011	0						18	100.00	0.00	100	100	100	
8972	0						65	99.57	1.50	93	100	100	
PI&PMG	55	98.09	5.19	74	100	100	57193	98.43	5.69	17	100	100	0.96

Q.1.7.1.1 PFT Crunches (Male Marines) Results

The average PFT crunches of the GCEITF male Marines was 98.63 compared to the average PFT crunches of male Marines within the total Marine Corps Population which is 98.59. Based on the p-value of the K-S test being greater than .10, there is no statistically significant difference between the two groups. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.7.2 PFT (Female Marines) Crunches

Table Q-14 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 14 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-14. Female PFT Crunches Comparisons

Female PFT Crunches													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	103	96.83	7.49	60.0	100.00	100.0	6434	94.62	10.93	35.0	100.00	100.0	0.44

Female PFT Crunches													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	98.08	6.04	79	100	100	0						
0313	7	94.43	9.73	77	100	100	0						
0331	9	99.89	0.33	99	100	100	0						
0341	8	98.50	3.51	90	100	100	0						
0351	3	100.00	0.00	100	100	100	0						
0352	4	97.75	4.50	91	100	100	0						
0811	14	97.14	6.01	82	100	100	0						
1371	9	97.67	7.00	79	100	100	123	96.49593496	9.804125916	51	100	100	1.00

1812	4	100.00	0.00	100	100	100	0						
1833	12	95.42	11.96	60	100	100	0						
8000	0						1	65		65	65	65	
8011	0						5	93.8	13.31164903	70	100	100	
8972	0						8	98.375	4.596194078	87	100	100	
PI&PMG	21	94.00	9.30	72	100	100	6295	94.59126291	10.94367908	35	100	100	0.95

Q.1.7.2.1 PFT Crunches (Female Marines) Results

The average PFT crunches of the GCEITF female Marines was 96.83 compared to the average PFT crunches of female Marines within the total Marine Corps Population which is 94.62. Based on the p-value of the K-S test being greater than .10, there is no statistically significant difference between the two groups. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.8 Physical Fitness Test (PFT) Run Time

This personnel variable defines PFT run time as total time required to run a 3 mile distance during an annual physical fitness test specified as in minutes.

Q.1.8.1 PFT (Male Marines) Run Time

Table Q-15 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-15. Male PFT Run Time Comparisons

Male PFT Run Time (minutes)													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	257	22.30	2.14	17.48	22.17	30.48	76832	22.56	2.39	13.97	22.37	40.98	0.14

Male PFT Run Time (minutes)													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300							130	21.12	1.82	17.63	20.75	27.10	
0311	53	21.79	1.83	18.37	21.80	26.40	9231	21.99	2.19	15.83	21.80	39.20	0.63
0313	14	22.40	2.73	19.27	21.79	27.83	440	22.39	2.35	16.83	22.13	36.87	0.88
0321							312	20.33	1.72	16.95	20.08	26.23	
0331	9	21.93	1.70	19.92	21.67	24.80	1457	21.63	2.20	13.97	21.38	36.38	0.86
0341	16	22.62	2.69	19.18	21.81	28.30	1516	21.83	2.07	16.38	21.67	29.83	0.59
0351	3	24.03	5.61	20.35	21.25	30.48	678	21.91	2.18	16.53	21.73	36.38	0.90
0352	16	21.51	2.26	18.00	21.27	27.17	752	22.28	2.23	16.37	22.25	27.97	0.10
0372							134	20.64	1.75	16.33	20.50	26.50	
0800							58	22.05	1.75	18.18	21.76	26.92	

0811	34	22.48	1.93	18.97	22.71	25.97	1168	22.35	2.14	16.83	22.20	35.58	0.35
0842							167	22.35	2.24	18.32	22.05	29.72	
0844							314	22.30	2.26	17.33	22.21	27.60	
0847							121	22.92	2.22	18.68	22.68	28.73	
0861							230	22.28	2.25	16.28	22.18	28.47	
1371	18	22.46	2.19	17.48	22.94	26.83	1794	22.35	2.16	16.20	22.25	31.48	0.73
1800							24	21.22	1.65	18.40	21.07	24.52	
1812	19	23.91	1.93	20.20	24.00	29.00	367	22.48	2.03	17.22	22.35	30.25	<0.01*
1833	20	22.13	1.81	19.13	22.08	25.00	871	22.13	2.11	16.58	22.00	29.90	0.94
8011							18	20.70	1.46	18.80	20.96	23.13	
8972							65	22.85	2.24	18.47	22.70	27.52	
PI&PIMG	55	22.23	1.97	18.00	22.20	27.57	56878	22.75	2.42	14.27	22.57	40.98	0.28

Q.1.8.1.1 PFT Run time (Male Marines) Results

The average PFT run time of the GCEITF male Marines was 22.30 minutes compared to the average PFT run time of male Marines within the total Marine Corps Population which is 22.56 minutes. Based on the p-value of the K-S test being greater than .10, there is no statistically significant difference between the two groups. However, the following MOSs had statistically significant differences:

- 1812 GCEITF male Marines average PFT run time was 23.91 minutes compared to the average Marine Corps 1812 population which is 22.48 minutes. The GCEITF Marines' average PFT run times were statistically significantly slower, by approximately 1:25.

Q.1.8.2 PFT (Female Marines) Run Time

Table Q-16 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table Q-16 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-16. Female PFT Run Time Comparisons

Female PFT Run Time (minutes)													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	103	23.79	2.20	19.3	23.67	30.7	6396	25.76	2.55	17.55	25.70	40.98	<.01*

Female PFT Run Time (minutes)													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value

0311	12	22.88	1.92	19.68	22.97	25.57							
0313	7	23.18	1.12	21.52	23.2	24.62							
0331	9	23.27	1.80	19.33	23.37	25.58							
0341	8	23.07	1.04	21.72	22.87	24.92							
0351	3	22.79	2.25	20.62	22.62	25.12							
0352	4	22.84	1.78	20.97	22.79	24.83							
0811	14	24.54	2.53	20.87	24.49	29.62							
1371	9	22.99	1.34	20.53	23.43	24.67	121	25.27	2.55	19.85	25.25	32.67	<0.01*
1812	4	23.43	3.35	19.38	23.68	27							
1833	12	24.20	1.95	20.07	24.28	27.92							
8000							1	26.33		26.33	26.33	26.33	
8011							5	27.36	1.44	26.02	26.93	29.25	
8972							8	25.27	1.72	23.15	25.14	28.55	
PI&PIMG	21	25.00	2.70	20.1	25.5	30.7	6260	25.77	2.55	17.55	25.72	40.98	0.64

Q.1.8.2.1 PFT Run time (Female Marines) Results

The average PFT run time of the GCEITF female Marines was 23.79 minutes compared to the average PFT run time of female Marines within the total Marine Corps Population which is 25.76 minutes. Based on the p-value of the K-S test being greater than .10, the difference is statistically significant difference and faster, on average, by approximately two minutes. Additionally, the following MOSs had statistically significant differences:

- 1371GCEITF female Marines average PFT run time was 22.99 minutes compared to the average Marine Corps 1371 population which is 25.72. The GCEITF Marines' average PFT run times were statistically significantly faster.

Q.1.9 Combat Fitness Test (CFT) Score

This personnel variable defines CFT as a score specified as a numerical value.

The CFT score is a composite score of three events: Movement to Contact Run (880 yards), Ammunition Can lifts, Maneuver under Fire. The score is gender-normed with a maximum of 300 points awarded to both male and female Marines.

Q.1.9.1 CFT (Male Marines) Score

Table Q-17 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-17. Male CFT Score Comparisons

Male CFT Score													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	275	287.44	11.19	236.00	290.00	300.00	78567	288.29	11.83	148.00	291.00	300.00	0.04*

Male CFT Score													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						127	290.17	12.39	221.0	294.0	300.0	
0311	63	287.41	11.47	258.0	291.0	300.0	9732	290.93	9.99	214.0	294.0	300.0	0.04*
0313	14	285.71	17.18	236.0	289.0	300.0	446	291.07	10.12	233.0	294.0	300.0	0.44
0321	0						316	297.59	5.33	254.0	300.0	300.0	
0331	13	287.46	12.56	256.0	288.0	300.0	1539	293.32	8.29	242.0	296.0	300.0	0.16
0341	18	284.56	10.33	261.0	287.5	297.0	1580	291.42	10.27	157.0	294.0	300.0	0.01*
0351	4	290.25	6.65	282.0	291.0	297.0	706	291.89	9.37	234.0	295.0	300.0	0.65
0352	16	290.06	8.14	273.0	290.5	300.0	784	291.39	9.14	239.0	294.0	300.0	0.84
0372	0						136	298.40	3.25	284.0	300.0	300.0	
0800	0						60	286.25	13.61	249.0	289.5	300.0	
0811	35	285.49	9.69	262.0	286.0	300.0	1185	288.59	10.76	226.0	291.0	300.0	0.06*
0842	0						167	290.54	8.84	257.0	292.0	300.0	
0844	0						322	288.39	10.57	250.0	291.0	300.0	
0847	0						121	286.93	11.99	251.0	289.0	300.0	
0861	0						247	289.29	9.90	253.0	291.0	300.0	
1371	19	283.68	11.35	265.0	283.0	300.0	1872	289.20	10.71	222.0	292.0	300.0	0.02*
1800	0						24	289.33	7.58	273.0	290.0	300.0	
1812	19	287.11	10.68	264.0	291.0	300.0	378	285.71	12.06	234.0	288.0	300.0	0.95
1833	20	292.20	9.18	272.0	295.0	300.0	904	290.71	10.40	234.0	293.0	300.0	0.59
8011	0						18	290.67	9.90	271.0	294.0	300.0	
8972	0						65	290.00	9.83	245.0	292.0	300.0	
PI&PMG	54	288.81	11.48	243.0	293.0	300.0	57838	287.38	12.26	148.0	290.0	300.0	0.19

Q.1.9.1.1 CFT (Male Marines) Score Results

The average CFT score of the GCEITF male Marines was 287.44 compared to the average CFT score of male Marines within the total Marine Corps Population which is 288.29. Based on the p-value of the K-S test being less than .10, the difference is statistically significant. The GCEITF male Marines' CFT scores were statistically lower than the USMC total population by approximately 1 point. Additionally, the following MOSs had statistically significant differences:

- 0311 GCEITF male Marines average CFT score was 287.41compared to the average Marine Corps 0311 population, which is 290.93. The GCEITF Marines' average CFT scores were statistically significantly lower by approximately 3.5 points.

- 0341 GCEITF male Marines average CFT score was 284.56 compared to the average Marine Corps 0341 population which is 291.42. The GCEITF Marines' average CFT scores were statistically significantly lower, by approximately 6.7 points.
- 0811 GCEITF male Marines average CFT score was 285.49 compared to the average Marine Corps 0811 population which is 288.59. The GCEITF Marines' average CFT scores were statistically significantly lower, by approximately 3 points.
- 1371 GCEITF male Marines average CFT score was 283.68 compared to the average Marine Corps 1371 population which is 289.20. The GCEITF Marines' average CFT scores were statistically significantly lower, by approximately 6 points.

Q.1.9.2 CFT (Female Marines) Score

Table Q-18 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 18 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps Population.

Table Q-18. Female CFT Score Comparisons

Female CFT Score													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	103	292.98	8.37	266	297	300	6476	286.83	13.17	143	290	300	< 0.01*

Female CFT Score													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	292.92	8.74	276	296.5	300	0						
0313	7	291.14	5.96	282	291	299	0						
0331	9	296.33	7.12	279	300	300	0						
0341	8	296.75	4.23	290	299	300	0						
0351	3	299.33	1.15	298	300	300	0						
0352	4	298.50	1.73	297	298.5	300	0						
0811	15	289.73	9.97	272	294	300	0						
1371	9	292.56	10.62	274	300	300	123	288.17	12.82	236	293	300	0.30
1812	4	296.25	4.50	290	297.5	300	0						
1833	12	292.67	7.95	276	295	300	0						
8000	0						1	269.00		269	269	269	
8011	0						5	262.60	13.05	249	260	279	

8972	0						8	286.00	11.45	271	287.5	300	
PI&PMG	20	290.75	9.51	266	293.5	300	6337	286.83	13.17	143	290	300	0.30

Q.1.9.2.1 CFT (Female Marines) Score Results

The average CFT of the GCEITF female Marines was 292.98 compared to the average CFT score of female Marines within the total Marine Corps Population which is 286.83 Based on the p-value of the K-S test being greater than .10, there is a statistically significant difference between the two groups. The GCEITF female Marines' CFT scores are higher, on average, by approximately 6 points. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.10 Combat Fitness Test (CFT) MTC

This personnel variable defines CFT Movement to Contact (MTC) run time as total time required to run an 880 yard distance during an annual combat fitness test specified as in minutes.

Q.1.10.1 CFT (Male Marines) MTC

Table Q-19 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-19. Male CFT MTC Time Comparisons

Male CFT MTC Time (minutes)													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	275	2.93	0.26	1.97	2.92	3.83	78719	2.93	0.30	1.75	2.90	8.00	0.18

Male CFT MTC Time (minutes)													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300							128	2.85	0.23	2.3	2.8	3.6	
0311	63	2.92	0.26	2.5	2.9	3.7	9740	2.88	0.28	1.8	2.9	8.0	0.78
0313	14	2.88	0.27	2.6	2.9	3.6	447	2.86	0.27	2.0	2.8	4.2	1.00
0321	0						316	2.66	0.20	2.1	2.7	3.4	
0331	13	2.93	0.27	2.3	2.9	3.3	1540	2.84	0.25	1.7	2.8	4.0	0.29
0341	18	3.08	0.27	2.7	3.0	3.8	1582	2.85	0.29	1.8	2.8	4.4	<0.01*
0351	4	2.91	0.14	2.7	2.9	3.1	708	2.88	0.26	2.0	2.9	4.3	0.73
0352	16	2.93	0.20	2.4	3.0	3.3	785	2.86	0.28	2.0	2.8	4.2	0.34
0372							136	2.70	0.18	2.2	2.7	3.2	
0800							60	2.94	0.22	2.5	2.9	3.5	
0811	35	3.02	0.30	2.0	3.1	3.6	1187	2.91	0.30	2.0	2.9	4.0	<0.01*
0842							168	2.86	0.27	2.3	2.8	3.9	

0844							322	2.92	0.32	2.2	2.9	3.8	
0847							121	2.95	0.28	2.3	2.9	3.9	
0861							247	2.92	0.28	2.0	2.9	3.8	
1371	19	2.99	0.29	2.4	3.0	3.5	1875	2.93	0.27	1.8	2.9	4.3	0.40
1800							24	2.89	0.16	2.6	2.9	3.2	
1812	19	2.98	0.20	2.7	3.0	3.3	378	3.02	0.30	2.0	3.0	4.2	0.79
1833	20	2.84	0.21	2.3	2.8	3.2	904	2.85	0.30	2.0	2.8	4.1	0.67
8011							18	2.85	0.23	2.4	2.8	3.4	
8972							65	2.75	0.24	2.3	2.8	3.3	
PI&PMG	54	2.85	0.23	2.3	2.9	3.5	57863	2.94	0.30	1.8	2.9	6.6	0.03*

Q.1.10.1.1 CFT (Male Marines) MTC Score Results

The average CFT MTC time for GCEITF male Marines was 2.93 minutes compared to the average CFT MTC time of male Marines within the total Marine Corps Population which is 2.93. Based on the p-value of the K-S test being greater than .10, the difference is not statistically significant. The following MOSs had statistically significant differences:

0341 GCEITF male Marines average CFT MTC time was 3.08 compared to the average Marine Corps 0341 population which is 2.85. The GCEITF Marines' average CFT MTC time scores were statistically significantly slower by approximately 25 seconds.

0811 GCEITF male Marines average CFT MTC time was 3.02 compared to the average Marine Corps 0811 population which is 2.91. The GCEITF Marines' average CFT MTC time scores were statistically significantly slower by approximately 7 seconds.

PI & PMG GCEITF male Marines average CFT MTC time was 2.85 compared to the average Marine Corps PI & PMG population which is 2.94. The GCEITF Marines' average CFT MTC time scores were statistically significantly faster by approximately 5 seconds.

Q.1.10.2 CFT (Female Marines) MTC

Table Q-20 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 20 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-20. Female CFT MTC Time Comparisons

All MOS	GCEITF						USMC						KS Test
	Female CFT MTC Time (minutes)												
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	103	3.43	0.30	2.8	3.42	4.3	6497	3.63	0.39	2.0	3.62	9.4	<0.01*

MOS	GCEITF						USMC						KS Test
	Female CFT MTC Time (minutes)												
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	3.27	0.24	2.93	3.2583	3.78							
0313	7	3.35	0.18	3.08	3.35	3.63							
0331	9	3.32	0.35	2.88	3.3	4							
0341	8	3.31	0.32	2.83	3.3917	3.73							
0351	3	3.14	0.29	2.88	3.0833	3.45							
0352	4	3.31	0.17	3.15	3.3	3.5							
0811	15	3.60	0.30	3.15	3.55	4.25							
1371	9	3.31	0.27	2.77	3.3667	3.73	123	3.58	0.42	1.97	3.5833	4.98	0.06
1812	4	3.41	0.29	3.07	3.4167	3.75							
1833	12	3.48	0.27	3.08	3.475	3.92							
8000							1	3.72		3.72	3.7167	3.72	
8011							5	3.88	0.40	3.53	3.7	4.53	
8972							8	3.63	0.35	3.33	3.4667	4.23	
PI&PMG	20	3.60	0.26	3.1	3.5667	4.17	6358	3.63	0.39	2.27	3.6167	9.35	0.90

Q.1.10.2.1 CFT (Female Marines) MTC Results

The average MTC time of the GCEITF female Marines was 3.43 minutes compared to the average MTC time of female Marines within the total Marine Corps Population which is 3.63 minutes. Based on the p-value of the K-S test being less than .10, this difference is statistically significant. The GCEITF female Marines CFT MTC times were statistically significantly faster than the USMC total population, by approximately 12 seconds. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.11 Combat Fitness Test (CFT) MUF

This personnel variable defines CFT Maneuver under Fire (MUF) time as total time required to conduct the MUF event during an annual combat fitness test specified as in minutes.

Q.1.11.1 CFT (Male Marines) MUF

Table Q-21 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-21. Male CFT MUF Time Comparisons

All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	275	2.44	0.31	1.80	2.40	3.75	78688	2.40	0.32	1.08	2.37	8.00	0.06*

Male CFT MUF Time (minutes)

MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300							128	2.35	0.30	1.6	2.3	3.2	
0311	63	2.44	0.34	1.8	2.4	3.8	9738	2.32	0.29	1.5	2.3	5.3	<0.01*
0313	14	2.50	0.47	1.8	2.5	3.4	447	2.32	0.28	1.6	2.3	3.4	0.21
0321							316	2.15	0.23	1.6	2.1	3.4	
0331	13	2.48	0.36	2.1	2.5	3.5	1540	2.26	0.26	1.5	2.2	3.5	0.06
0341	18	2.48	0.23	2.0	2.5	2.9	1582	2.32	0.31	1.6	2.3	8.0	<0.01*
0351	4	2.42	0.15	2.3	2.4	2.6	708	2.30	0.26	1.6	2.3	3.6	0.37
0352	16	2.27	0.18	2.0	2.3	2.6	785	2.34	0.26	1.6	2.3	3.8	0.37
0372							136	2.04	0.26	1.5	2.1	2.8	
0800							60	2.44	0.28	1.9	2.4	3.0	
0811	35	2.48	0.30	2.0	2.5	3.3	1187	2.41	0.29	1.5	2.4	3.6	0.28
0842							167	2.36	0.24	1.8	2.3	3.2	
0844							322	2.42	0.28	1.8	2.4	3.5	
0847							121	2.45	0.31	1.9	2.5	3.4	
0861							247	2.39	0.27	1.8	2.4	3.4	
1371	19	2.58	0.30	1.9	2.6	3.0	1876	2.37	0.28	1.7	2.3	3.8	<0.01*
1800							24	2.41	0.17	2.1	2.5	2.7	
1812	19	2.44	0.32	2.1	2.3	3.2	378	2.44	0.29	1.7	2.4	3.4	0.55
1833	20	2.27	0.27	2.0	2.2	2.9	904	2.34	0.31	1.5	2.3	3.9	0.41
8011							18	2.39	0.29	2.0	2.3	2.9	
8972							65	2.45	0.32	1.7	2.5	3.3	
PI&PIMG	54	2.44	0.28	1.9	2.4	3.1	57829	2.43	0.32	1.1	2.4	4.9	0.80

Q.1.11.1.1 CFT (Male Marines) MUF Results

The average CFT MUF times of the GCEITF male Marines was 2.44 minutes compared to the average CFT MUF times of male Marines within the total Marine Corps Population which is 2.40 minutes. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF male Marines' CFT MUF times were statistically significantly slower than the USMC total population, by approximately 2 seconds. Additionally, the following MOSs had statistically significant differences:

- 0311 GCEITF male Marines average CFT MUF times was 2.44 minutes compared to the average Marine Corps 0311 population which had an average of 2.32 minutes. The GCEITF Marines' average CFT MUF times were statistically significantly slower, by approximately seven seconds.
- 0341 GCEITF male Marines average CFT MUF times was 2.48 minutes compared to the average Marine Corps 0341population which had an average of 2.32 minutes. The GCEITF Marines' average CFT MUF statistically significantly slower, by approximately nine seconds.
- 1371 GCEITF male Marines average CFT MUF times was 2.58 minutes compared to the average male Marine Corps 1371population which had an average of 2.37 minutes. The GCEITF Marines' average CFT MUF statistically significantly slower, by approximately 12 seconds.

Q.1.11.2 CFT (Female Marines) MUF

Table Q-22 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 22 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-22. Female CFT MUF Time Comparisons

Female CFT MUF Time (minutes)													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	103	3.09	0.42	2.2	3.08	4.1	6489	3.27	0.47	2	3	7	<0.01*

Female CFT MUF Time (minutes)													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	3.12	0.40	2.5	3	3.8							
0313	7	3.33	0.35	3	3.2	4							
0331	9	2.90	0.43	2.3	2.9	3.7							
0341	8	2.93	0.39	2.4	2.9	3.5							
0351	3	2.49	0.42	2.2	2.4	3							
0352	4	2.89	0.26	2.6	2.9	3.2							
0811	15	3.24	0.41	2.5	3.2	3.9							
1371	9	3.18	0.47	2.7	3	4	123	3.20	0.47	2.417	3.167	5.833	0.67
1812	4	2.91	0.30	2.5	3	3.2							
1833	12	3.14	0.45	2.5	3.1	4.1							
8000							1	3.88		3.883	3.883	3.883	
8011							5	4.14	0.53	3.583	4.2	4.917	

8972							8	3.48	0.34	2.867	3.5	3.867	
PI&PIMG	20	3.15	0.42	2.6	3.2	4	6350	3.27	0.47	2	3.23	7.17	0.44

Q.1.11.2.1 CFT (Male Marines) MUF Results

The average CFT MUF times of the GCEITF female Marines was 3.09 minutes compared to the average CFT MUF times of female Marines within the total Marine Corps Population which is 3.27 minutes. Based on the p-value of the K-S test being less than 0.10, this difference is statistically significant. The GCEITF female Marines' CFT MUF times were statistically significantly faster than the USMC total population, by approximately 11 seconds. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

Q.1.12 Rifle Range Score

This personnel variable defines Rifle Range as a score specified as a numerical value.

Q.1.12.1 Rifle Range Score (Male Marines)

Table Q-23 compares male Marines within the GCEITF to all male Marines within the total Marine Corps Population and by MOS.

Table Q-23. Male Rifle Range Score Comparisons

Male Rifle Range Score													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	278	299.95	25.10	162.00	303.00	342.00	79189	306.83	22.31	47.00	310.00	349.00	< 0.01*

Male Rifle Range Score													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0300	0						131	313.66	15.06	250.0	315.0	343.0	
0311	65	303.55	23.27	193.0	311.0	334.0	9759	307.19	24.01	85.0	311.0	348.0	0.04*
0313	14	295.57	38.80	187.0	300.0	340.0	447	306.95	22.77	147.0	311.0	342.0	0.32
0321	0						318	318.84	15.01	250.0	320.0	347.0	
0331	13	301.31	18.63	271.0	298.0	339.0	1550	308.51	22.37	137.0	312.0	347.0	0.00
0341	18	291.89	35.34	183.0	303.0	323.0	1591	304.92	26.41	116.0	309.0	345.0	0.07*
0351	4	304.00	13.74	289.0	303.5	320.0	713	307.68	22.57	133.0	312.0	342.0	0.91
0352	16	312.94	11.65	293.0	313.0	342.0	790	309.74	23.23	163.0	313.0	341.0	0.93
0372	0						138	322.48	11.42	285.0	322.5	349.0	
0800	0						60	303.12	19.36	250.0	306.5	336.0	
0811	35	300.17	14.43	250.0	299.0	324.0	1195	304.33	26.38	135.0	309.0	345.0	0.02*
0842	0						168	309.90	20.79	131.0	313.0	341.0	
0844	0						324	306.93	23.96	139.0	312.0	341.0	

0847	0						121	310.15	14.55	260.0	312.0	339.0	
0861	0						250	307.90	21.83	146.0	311.0	341.0	
1371	19	310.00	18.01	271.0	311.0	342.0	1890	306.34	23.42	139.0	310.0	346.0	0.90
1800	0						24	308.33	14.67	281.0	311.0	332.0	
1812	19	288.37	39.31	162.0	299.0	329.0	379	293.99	36.09	124.0	303.0	336.0	0.37
1833	20	305.60	21.57	250.0	309.5	330.0	906	308.20	21.49	171.0	312.0	347.0	0.55
8011	0						18	302.44	20.65	250.0	308.0	327.0	
8972	0						65	310.71	13.29	273.0	311.0	335.0	
PI&PMG	55	293.38	23.61	179.0	296.0	333.0	58352	306.71	21.66	47.0	310.0	347.0	< 0.01*

Q.1.12.1.1 Rifle Range Score (Male Marines) Results

The average Rifle Range score of the GCEITF male Marines was 299.95 compared to the average Rifle Range score of male Marines within the total Marine Corps Population which is 306.83. Based on the p-value of the K-S test being less than .10, this difference is statistically significant. The GCEITF male Marines' Rifle Range scores were statistically significantly less than the USMC total population, by approximately 7 points. Additionally, the following MOSs had statistically significant differences:

- 0311 GCEITF male Marines average Rifle Range score was 303.55 compared to the average Marine Corps 0311 population which had an average score of 307.19. The GCEITF Marines' average Rifle Range scores were statistically significantly lower, by approximately four points.
- 0341 GCEITF male Marines average Rifle Range score was 291.89 compared to the average Marine Corps 0341population which had an average score of 304.92. The GCEITF Marines' average Rifle Range scores were statistically significantly lower, by approximately thirteen points.
- 0811 GCEITF male Marines average Rifle Range score was 300.17 compared to the average Marine Corps 0811 population which had an average score of 304.33. The GCEITF Marines' average Rifle Range scores were statistically significantly lower, by approximately four points.
- PI and PIMG GCEITF male Marines average Rifle Range score was 293.38 compared to the average Marine Corps open MOS population which had an average score of 306.71. The GCEITF Marines' average Rifle Range scores were statistically significantly lower, by approximately thirteen points.

Q.1.12.2 Rifle Range Score (Female Marines)

Table Q-24 compares female Marines within the GCEITF to all female Marines within the total Marine Corps Population. Table 24 compares the 1371 Combat Engineer female Marines within the GCEITF to female 1371 Combat Engineers within the total

Marine Corps Population and the PI and PMG GCEITF Marines to the total Marine Corps' open MOS female Population.

Table Q-24. Female Rifle Range Score Comparisons

Female Rifle Range Score													
All MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
	103	296.99	23.61	168.0	302.00	337.0	6629	295.98	26.62	109	300	344	0.94

Female Rifle Range Score													
MOS	GCEITF						USMC						KS Test
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max	P-value
0311	12	297.67	17.64	270	293.5	330	0						
0313	7	286.71	21.76	250	287	313	0						
0331	9	307.78	14.02	291	309	330	0						
0341	8	304.13	19.03	265	308.5	328	0						
0351	3	305.67	6.51	299	306	312	0						
0352	4	295.00	30.30	250	307	316	0						
0811	14	283.86	36.48	168	291	316	0						
1371	9	298.78	15.00	279	296	319	124	296.4677419	26.34277966	141	299	333	1.00
1812	4	320.50	14.80	304	320.5	337	0						
1833	12	289.42	21.61	250	296.5	310	0						
8000	0						1	293		293	293	293	
8011	0						5	287.6	17.03819239	262	290	309	
8972	0						8	306.75	17.20257456	282	310	325	
PI&PMG	21	299.67	23.28	250	306	327	6489	295.9631684	26.64596459	109	300	344	0.37

Q.1.12.2.1 Rifle Range Score (Female Marines) Results

The average Rifle Range score of the GCEITF female Marines was 296.99 compared to the average Rifle Range score of female Marines within the total Marine Corps Population, which is 295.98. Based on the p-value of the K-S test being greater than 0.10, there is no statistically significant difference between the two groups. Additionally, there are no statistically significant differences in any of the MOS-MOS comparisons.

THIS PAGE INTENTIONALLY LEFT BLANK

Annex R.
Methodology

This annex details the methodology portion of the Ground Combat Element Integrated Task Force (GCEITF) experiment. The sections outline the Volunteer Timeline, Design of Experiment, Descriptive and Basic Inferential Statistics, Statistical Models, and Limitations.

R.1 Volunteer Timeline

R.1.1 Recruitment and Selection

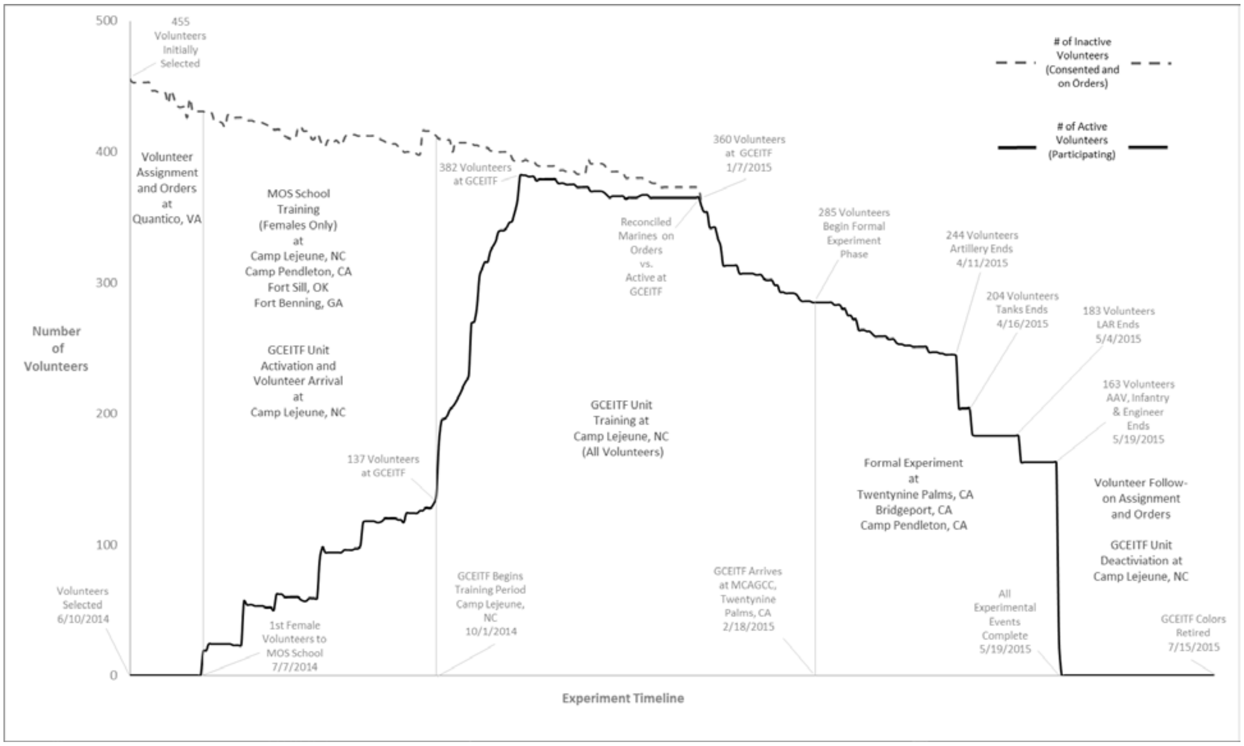
On 10 June 2014, MCOTEA made the initial selections for volunteer participation from the pool of applicants who had provided informed consent. The initial selection consisted of 455 volunteers (287 Male/168 Female). MCOTEA registered all female volunteers who needed to attend Formal Learning Centers (FLCs) for Military Occupational Specialty (MOS) training with Training and Education Command (TECOM).

R.1.2 School Assignments and Orders

All female volunteers attending FLCs were issued Temporary Additional Duty orders. From that point, MCOTEA began the process of assignment and orders through Manpower and Reserve Affairs (M&RA), Headquarters Marine Corps. Male and female volunteers who were not required to attend an FLC were issued Permanent Change of Station (PCS) or Permanent Change of Assignment (PCA) orders directly to the GCEITF at Camp Lejeune, NC, with reporting dates to coincide with their peers in their respective MOSs. Females who successfully completed the course of instruction at the FLCs were issued PCS/PCA orders just prior to graduation.

Not all volunteers wanted, or were physically able, to complete the experiment. From initial notification in June 2014 through May 2015, SOME volunteers voluntarily withdrEw or were involuntarily withdrawn from the experiment. Initially, MCOTEA sought alternates to replace the volunteers. Volunteers were replaced on a case-by-case basis from the initial selections in June 2014 through November 2014 using the same procedures outlined in the Experimental Assessment Plan. Replacements were no longer sought after November, because the loss of training time would negatively impact the volunteer's participation in the experiment. Figure R-1 illustrates the quantity of volunteers who were on orders, and those actively participating in an experimental activity.

Figure R-1. Volunteer Timeline from Notification through Deactivation



In Figure R-1, the vertical axis is the number of volunteers in the experiment; the horizontal axis is time. The timeline depicted begins on 10 June 2014 with the initial selection of volunteers, and ends on 15 July 2015, when the GCEITF deactivated. Thin gray vertical lines indicate times when significant experimental activities began along the timeline. The heavy dashed line represents the number of volunteers who were on orders but had not begun participating in any experimental activities. The solid black line represents the number of volunteers who actively participated in some experimental activity, including attending MOS training, GCEITF unit training, and formal experimentation. The dashed line shows a general downward trend as volunteers withdrew from the experiment. The upward spikes in the dashed line show where alternate selections took place, with the last significant selection event occurring in November 2014.

R.1.3 FLC Training and GCEITF Unit Training

The solid black line in Figure R-1 trends upward on 7 July 2014, when the first of the female volunteers began training at MOS schools. The line continues to trend upward in steps as more females begin classes at subsequent MOS schools. The small downward dips in the lines—from the start of the first females attending MOS schools until volunteer training begins on 1 October 2014—represent volunteers who withdrew from the experiment. Most volunteers arrived from 1 October 2014 through 1 November 2014, indicated by the steady incline of the solid black line. From 1 November 2014 through 7 January 2015, the number of volunteers remained relatively stable, with small numbers of volunteers withdrawing from the experiment, as indicated by the modest downward trend in the number of active volunteers. On 7 January 2015, MCOTEA reconciled with the GCEITF the number of volunteers who were on orders with those who were actually present at the unit. This is the point in the figure where the dashed and solid lines converge. Shortly after 7 January 2015, there is a large decrease in the volunteer population, as many male Marines voluntarily withdrew from the experiment after the holiday leave period.

R.1.4 Experiment Phase

On 18 February 2015, the GCEITF training period officially ended as the GCEITF arrived with 285 volunteers at MCAGCC, Twentynine Palms, CA. From arrival at MCAGCC to 7 March 2015, when experimental trials officially began, the unit prepared equipment, acclimatized, and practiced experimental tasks. By the time the first experimental trials began, the number of volunteers had dropped to 264.

From 7 March to 11 April 2015, all elements of the experiment engaged in experimental trials with a modest decrease in the number of volunteers, due mostly to injuries. Beginning on 11 April 2015, the solid black line declines in a stepwise manner, as each of the MOSs concludes experimental activities and large numbers of volunteers conclude active participation. Artillery was the first to complete all experimental trials, followed by Tanks on 16 April 2015. By 28 April 2015, the Assault Amphibious Vehicle (AAV), all Infantry MOSs (including Provisional Infantry), and Engineers were done with their experimental trials and moved on from MCAGCC to Camp Pendleton (AAVs) and Bridgeport (all Infantry MOSs, Provisionals, and Engineers). The next to conclude all experimental trials was the Light Armored Vehicle (LAV) crews, on 4 May 2015. The final elements of the experiment to conclude activities were all Infantry MOSs and Engineers, on 18 May, with AAVs concluding on 19 May 2015. From 20 May 2015 until the GCEITF shutdown on 15 July 2015, the volunteers were in an administrative status while awaiting orders to their follow-on units or released from active duty for Reservists.

R.1.5 Volunteer Summary

MCOTEA recruited and obtained informed consent from a total of 598 Marines, although the maximum active participation by volunteers at any one time was 382. Of the 598 volunteers, 422 volunteers actually participated in the experiment in some fashion, either by attending a formal school, training with the unit, and/or participating in the formal experiment.

R.2 Analysis Limitations

R.2.1 Independence assumption

Some of the statistical analysis techniques used assume that the outcomes of each trial are statistically independent of one another. For many of the tasks, this assumption is justified because the number of volunteers were sufficient and their roles within tasks had little relationship to one another. Thus, if the same individual is sampled multiple times for the task but is performing different functions or is binned with different people, the data may be sufficiently independent for the purposes of the experiment.

Where possible, we implemented mixed-model analysis with random effects for individuals to explicitly adjust for the fact that our trials were not, strictly speaking, independent. However, there are some tasks for which a mixed model was not possible to run, and the results we present in the descriptive statistics sections also do not account for the lack of independence. For these types of tasks, it is possible for lack of independence to result in higher power of the test to detect differences between groups, and increased Type I error rates.

R.2.2 Sample selection

As detailed in the EAP Section 7, the GCEITF experiment was likely to suffer from selection bias—the notion that the Marines' decision to participate (volunteer) in the experiment may be correlated with the specific traits that affect the study itself. This could have manifested in the volunteers' motivation, performance, survey answers, and other aspects of the experiment. To the extent that we could, we tried to ensure the sample was representative of the population of Marines, with respect to observable physical characteristics that the USMC collects and records. This analysis is reported in the Population Annex. Based on that analysis, our conclusion is that with respect to these characteristics, the GCEITF sample was representative of the Marine Corps at large, with a few small and substantively unimportant exceptions. However, the possibility remains that the sample was different from the population in unobserved ways.

R.2.3 Sample size

The details of how we originally determined the size of the experiment are in the EAP Section 7. However, as the number of volunteers decreased, so did the number of possible trials that could be run per day. As a result, many tasks saw a reduction in sample size, some only running one-third of the planned trials. Whenever reasonable, a comparison group (usually, the LD group) was eliminated from consideration by the research team, splitting the smaller sample size among two different concentrations—control (C) and high-density (HD)—rather than three, to maintain better experimental power to detect differences. Still, with variation in some tasks, particularly high among integrated groups, a larger sample size would have resulted in more-precise estimates and more confidence in stating observed differences.

R.2.4 Modeling group outcomes using individual data

The experiment was designed to detect differences in performance between gender-integrated and all-male units. It is reasonable to perform such analyses using ANOVA or mixed-modeling techniques, particularly because this was a genuine experiment, with random assignment of volunteers to groups of various integration levels. However, in additional analyses, we also tried to determine how *individual* characteristics contributed to group outcome in an attempt to inform gender-neutral standards. We did this using data from each individual volunteer in the group, sometimes combining variables, when sample size permitted. Ultimately, our models assume a specific functional form: we assume that the group outcome depended on some linear combination of all participants' personnel variables. Intuitively, this need not be the case, and the poor fit of many of our models confirms that we gain little explanatory power by including personnel variables in the models. This is not to say that the variables do not have an impact on the outcome; more complex modeling may be needed to uncover these relationships.

R.3 Design of Experiment

For each task identified, sample size of the unit needed to perform a task (a 12-person squad or a 4-person crew) were drawn repeatedly from the relevant participant population to represent various concentration groups.

R.3.1 Integration Levels

The experimental plan specified that for each MOS, one experimental group would be entirely male (control group), one would contain one female (low-density or LD), and the third would contain two female participants (high-density or HD). As volunteers left the experiment, some concentration levels had to be changed or dropped from evaluation altogether. For each MOS, trials were run for at least the control group and one of the integrated concentrations. The result of each trial for each of these units constitutes one observation for the collective tasks analysis. The final male-female concentrations per MOS are displayed below, in Table R-1.

Table R-1. Integration Levels – by MOS

MOS	Number of volunteers in the group	Number of females in LD group	Number of females in HD group
0311	12	NA	2
0313	3	1	2
0331	3	1	2
0341	4	NA	2
035X	4	NA	2,3,4

MOS	Number of volunteers in the group	Number of females in LD group	Number of females in HD group
PI	12	2	4,5
PIMG	3	NA	3
0811	6	1	2
1812	2	1	NA
1833	3	1	2
1371	8	2	4
Mountaineering Closed MOS	12	NA	6
Mountaineering Open MOS	12	NA	6

R.3.2 Sample Size and Statistical Significance

The experiment was sized to detect at least a 30% difference in task outcomes (see EAP, section 7 for relevant design parameters, as well as sample sizes and effect sizes). Throughout the course of the experiment, sample sizes for some MOSs and tasks dropped—substantially in some cases—mainly due to the loss of volunteers and to environmental and experimental constraints. Each MOS-specific annex details the number of planned and executed trials for each task.

Ultimately, many differences lower than 30% were detected to be statistically significant. The analysis of each task includes a section with comments on potential operational implications of all statistically significant and nearly statistically significant differences.

R.3.3 Random assignment

For each MOS, prior to the beginning of each trial cycle, a run roster was drawn to randomly select Marines from the pool of volunteers to fill the rotating positions in each control and integrated group. In cases where the Marines were unavailable due to injury or from leaving the experiment, volunteers of the same gender were substituted for the empty spots at random. The ordering of integrated and control squads each day was randomized throughout the experiment to maintain balance between groups being exposed to various environmental factors.

R.3.4 Pilot Trial Cycles

To test data collection devices and protocol, and to ensure consistency between trials, pilot trial cycles consisting of two complete trial cycles were run prior to the start of experimental data collection. Because procedures used in executing the experiment and collecting data ultimately were not the same during pilot trial cycles as during the rest of the experiment, pilot data were not incorporated into the datasets analyzed for each MOS.

R.3.5 Outliers and Potentially Influential Points

Prior to data analysis, we removed points from the data that were clear outliers. By outliers, we mean points that were actually invalid data. Invalid data include data that were wrong because of human (data collector) error, results that were affected by trial interruptions and other test incidents, and points that were clearly erroneous due to data collection equipment.

For many tasks, we identified a small number of points that were valid data but were so unusual in their values that they could potentially drive the results. Because they were valid data that we could not exclude from analysis for good reason, we did not go through the formal process of computing Cook's distance for these points to detect if they affected our analysis in a statistically significant way. Rather, we simply present results with and without these points to illustrate how our conclusions might change in their absence.

R.4 Descriptive and Basic Inferential Statistics

This section details the analysis techniques used to describe the data and make conclusions about the difference in performance between control and integrated units, without adjusting for other variables. The statistical procedures described below enable us to take the sample data from the experiment and make statements about the differences in performance that we would expect to observe in the broader USMC population.

R.4.1 Critical Billets

For LAV, Tanks, AAV, and Artillery MOSs, there were metrics that involved only one or two volunteers in specific billets (we term them *critical billets*)—rather than the entire group—performing physically demanding tasks. In these cases, it no longer made sense to analyze the data just by integration level: if a male was performing a task, we would expect the result to be the same, whether he was on an integrated or control team. Hence, we provide analyses of such tasks by the gender of the person or people in those billets. The analysis techniques remain the same as when we compare gender-integrated groups to control groups.

R.4.2 Welch's t-test

The Welch's t-test is a statistical test that can be used to compare differences in two means across groups to the variation among and between groups, allowing unequal variances in two groups. We use Welch's t-test to describe whether differences between average responses for two integration groups are statistically significant or unlikely due to chance. The test is used for most of our descriptive analyses where there are only two groups to consider, and the outcomes are, or can be, assumed to be continuous (such as elapsed time or target hit probabilities). For each result, we provide a p-value, which we compare to the a-priori determined significance level of 10% (0.10). P-values less than 10% are considered statistically significant and are an indication that the observed differences between group means cannot be attributed to chance.

Welch's t-test is based on several assumptions. The first is that the observations are independent and identically distributed. Strictly speaking, the independence assumption may be violated in some of our analyses, since sampling was done with replacement. This is less problematic for the MOSs, where we had a large number of volunteers (0311s, for instance). In all cases, Marines were sampled to rotate through different positions on their respective squads and crews, and work with different Marines in other positions. All Marines within each MOS went through MOS school and, subsequently, trained together as a unit. For this reason, the violations of the independence assumption are unlikely to be strong. In section R.5.3.2, we try to address this concern whenever possible. The second assumption is that the response variable is normally distributed. We evaluate this assumption using the Shapiro-Wilk test for our task results, except where sample size is large enough (greater than 30) to use the Central Limit Theorem to make this assumption less important. We use a significance level of 1% for the Shapiro-Wilk test. In cases where at least one of the groups' results is not normally distributed, we use a non-parametric equivalent of the t-test, described in section R.4.4.

R.4.3 Analysis of Variance (ANOVA)

ANOVA provides a statistical test that can be used to compare differences in two or more means across groups to the variation among and between groups. We use ANOVA to describe whether differences between average responses for the different integration groups are statistically significant, or unlikely due to chance. ANOVA is used for most of our descriptive analyses where outcomes are, or can be, assumed to be continuous, such as elapsed time or target hit probabilities. For each result, we provide a test statistic (denoted by F in relevant tables), as well as a p-value, which we compare to the a-priori determined significance level of 10% (0.10). P-values less than 10% are considered statistically significant and are an indication that the observed differences between group means cannot be attributed to chance.

ANOVA tests all the means simultaneously—that is, in the case of comparing three means, a significant result shows only that the three means in consideration are not all the same and that further testing needs to be done to describe the size and direction of the differences between them. For more details on this type of analysis, see section R.4.5.

ANOVA is based on a few assumptions. The first two are identical to those for Welch's t-test listed above. The third assumption is that the variances of the result are similar across groups. This assumption is considered violated when the biggest variance under consideration is at least twice the smallest variance. ANOVA is sensitive to the violation of this assumption. For that reason, when variances across groups appear unequal, we use Robust ANOVA based on the Huber M-estimation method, which is a more robust fit procedure.

In the sections on descriptive statistics, we present ANOVA results for comparing means of two and three groups, although Welch's t-test, described in section R.4.2, is also used for comparing two groups. For two groups, the difference between Welch's t-test and ANOVA lies in the assumption of equal variances between groups (ANOVA assumes it, Welch's t-test does not), so the results are usually similar but not always exactly the same.

R.4.4 Nonparametric tests

The t-test and ANOVA assume that the data are normally distributed for each group under consideration. For some MOSs and tasks, that assumption failed and the sample sizes in each group were not sufficient to use the Central Limit Theorem. In those cases, we use non-parametric techniques to test for differences between groups.

The non-parametric equivalent of a t-test is the Mann-Whitney U-test, also called the Wilcoxon rank-sum test. It is designed to test whether two samples come from the same population or whether one sample comes from a population with higher values than the other. Rather than comparing means for the two samples, the Mann-Whitney U-test compares ranks of observations between the two groups, and indicates whether there is a pattern in the magnitude of observations from each population (i.e., if higher observations tend to fall more under the control or the integrated group).

The non-parametric equivalent of an ANOVA is the Kruskal-Wallis test. It is designed to test whether at least two samples come from the same distribution and is an extension of the Mann-Whitney U-test to three or more groups. Just like the Mann-Whitney U-test, the Kruskal-Wallis test looks for patterns in ranks. Significant results from the Kruskal-Wallis test are an indication that the comparison groups are stochastically different.

In tables of descriptive statistics, where necessary, we simply replace the value of the t-statistic, or the F-statistics and relevant p-values, with their non-parametric equivalents, indicating where the change was made.

R.4.5 Multiple Comparisons

In cases where ANOVA or Kruskal-Wallis test identified that the groups compared were not all the same, we compared each pair of groups to identify how they were different. Using multiple t-tests or Mann-Whitney tests for this purpose may result in inflated familywise Type I error rates: each test on its own has a 10% such rate, and if we are comparing three groups, we perform three such comparisons. Thus for post-hoc analyses of tasks that were identified significantly different across groups using ANOVA, we further analyzed the differences using the Tukey procedure in place of the regular t-test. Such a procedure does not have a clear non-parametric equivalent, so where a Kruskal-Wallis test gave significant results, we proceeded with Mann-Whitney tests for each pair of interest. However, to adjust for the inflated familywise Type I error rate, instead of comparing our p-values to 10%, we use the Bonferroni adjustment (known to be conservative for a large number of comparisons) and compare our p-values to 10% divided by the number of comparisons we make.

R.4.6 Additional Descriptive Analyses

Differences between groups may meet statistical significance criteria, but we also seek to provide operation context for their interpretation. For many of the tasks in the MOS-specific annexes, we include additional descriptive—often, qualitative analyses—based on the Marine Corps doctrine and subject matter expertise. These analyses fall into three possible categories: Contextual Comments, Additional Insights, and Subjective Comments. In the Contextual Comments sections, we try to put the observed differences into operation context, comparing results with Marine Corps standards, where they exist, and giving operationally realistic examples. In the Additional Insights sections, we provide some subject matter expert observations and judgments on the differences, as well as commentary based on literature outside the USMC, where relevant. Subjective Comments, which are primarily located in the MOS-specific appendices, provide trial-specific comments and observations of the GCEITF staff, broken into broad categories.

For those tasks that did not produce a statistically significant result, we still provide some interpretation if the differences were close to being statistically significant, since it is unclear if there is truly a lack of difference between the populations, or if a bigger sample size would have resulted in a statistically significant finding.

R.5 Statistical Models

The previous section discussed descriptive analyses as they pertain to differences in performance due to integration level only. This section details the analysis techniques used to make conclusions about the difference in performance between control and integrated units, adjusting for other variables. The goal of modeling, as applied here, is to estimate, simultaneously, the effect of gender-integration levels and other relevant variables on group performance, and to see whether accounting for the volunteers' physical attributes affects the conclusions of descriptive analyses above. This kind of model can be helpful in identifying which physical attributes best predict group performance measure and can be helpful in establishing gender-neutral physical standards.

R.5.1 USMC Personnel Variables

Personnel data for this portion of the analysis refers to the following variables found in the Marine Corps Total Force Management System (TFMS), commonly measured for every Marine: age, height, weight, Armed Forces Qualification Test (AFQT) score, general technical (GT) score, the Combat Fitness Test (CFT) maneuver-under-fire (MANUF) time, the CFT movement-to-contact (MTC) time, the Physical Fitness Test (PFT) three-mile run time, the PFT Crunch Score (number of crunches), and the annual combined rifle score. The personnel variables used in the statistical modeling are described below.

R.5.1.1 Age

The age of the volunteer in years as reported in TFMS.

R.5.1.2 Height

The height of the volunteer in inches as reported in TFMS.

R.5.1.3 Weight

The weight of the volunteer in pounds as reported TFMS.

R.5.1.4 Armed Forces Qualification Test (AFQT)

AFQT scores are computed using the standard scores from four Armed Services Vocational Aptitude Battery (ASVAB) subtests: arithmetic reasoning (AR), mathematics knowledge (MK), paragraph comprehension (PC), and word knowledge (WK). The formula used to derive the AFQT score is 2VE + AR + MK. The verbal expression (VE) score is used to measure communicative ability and consists of a scaled score derived from adding the value of the word knowledge (WK) raw score to the paragraph comprehension (PC) raw score. The AFQT composite score, as reported in TFMS, was used in the analysis.

R.5.1.5 General Technical (GT)

The GT score is calculated by adding the verbal expression (VE) and the arithmetic reasoning (AR) scores from the ASVAB. The GT composite score as reported in TFMS was used in the analysis.

R.5.1.6 Combat Fitness Test (CFT)

The CFT is used to assess a Marine’s physical capacity in a broad spectrum of combat-related tasks. The CFT was designed to evaluate strength, stamina, agility, and coordination, as well as overall aerobic capacity. The CFT is a complement to the PFT and measures the functional elements of combat fitness through execution of a series of events that represent every Marine’s combat experience, emphasizing the Corps’ ethos of “every Marine a rifleman.” The CFT is a scored, calendar-year annual requirement for all active duty Marines, regardless of age, gender, grade, or duty assignment. It is required to be conducted between 1 July and 31 December of each year.

The CFT is a three-part combat physical-fitness evaluation that consists of a movement to contact (MTC) simulated by an 880-yard run, a maximum set of ammunition lifts (AL) using a 30-pound ammo can, and a 300-yard maneuver-under-fire (MANUF) course. Both the MTC and MANUF are timed events and are thus comparable. However, the male and female standard for the ammunition-can lift varies based on age. Therefore, the modeling included the MTC and MANUF times, but excluded the AL number score.

R.5.1.7 Physical Fitness Test (PFT)

The PFT is a collective measure of general fitness throughout the Marine Corps. The PFT was designed to test the strength and stamina of the upper body, midsection, and lower body, as well as efficiency of the cardiovascular and respiratory systems. The PFT is a scored, calendar-year annual requirement for all active duty Marines, regardless of age, gender, grade, or duty assignment. It is required to be conducted between 1 January and 30 June of each year. It consists of a three-part test: a timed 3-mile run, a maximum set of crunches performed in a time of 2 minutes, and either a flexed-arm hang (for females) or pull-ups (for males and an option for females).

The modeling considered the 3-mile run time and the number of crunches performed during the PFT, but excluded the number of pull-ups and the flexed-arm hang time. Females have the option either to do a flexed-arm hang or complete the pull-ups portion of the PFT. Males are limited to pull-ups only, no flexed-arm hang. A passing score for females consists of 3 pull-ups; the maximum pull-ups are 8. The male Marine passing score consists of 3, but the maximum is 20. Currently, female Marines who opt for pull-ups and execute more than 8 pull-ups still receive a score of 8 in the system. Because there is no direct comparison between males and females in this category, the flexed-arm hang/pull-ups portion of the PFT is excluded for analysis.

R.5.1.8 Annual Combined Rifle Score

The intent of the Marine Corps annual rifle training is to sustain, improve, and evaluate marksmanship skills, to include demonstration of proficiency in fundamental marksmanship skills and combat shooting skills. Qualification scores are based on an aggregate of Table 1A (a known distance range with the capability to fire from 100, 200, 300, and 500 meters), and Table 2 (which requires Marines to fire at stationary targets from 25 meters and at moving targets from 100 meters). The aggregate rifle range score as reported in TFMS was used in the analysis.

R.5.2 Multicollinearity

Several variables listed above measure similar dimensions of a Marine. For instance, height and weight measure a person's size; GT and AFQT scores measure mental aptitude; and PFT and CFT scores and times measure physical fitness; these variables are correlated. Within the GCEITF, height and weight were correlated with the coefficient of 0.63 for females and 0.58 for males. For GCEITF females, the correlation coefficient between GT and AFQT scores is 0.92; it is 0.90 for males. We present the correlations for the physical fitness variables in Table R-2.

Table R-2. Correlations between Physical Fitness Measures for GCEITF Females

	CFT, Movement to contact	CFT, Maneuver under fire	PFT Run Time (minutes)	PFT, Crunches
CFT, Movement to contact	1.00	0.54	0.69	-0.07
CFT, Maneuver under fire	0.54	1.00	0.42	-0.14
PFT Run Time (minutes)	0.69	0.42	1.00	-0.08
PFT, Crunches	-0.07	-0.14	-0.08	1.00

Some of these correlations are quite high, indicating that when used in a linear model, these variables will attempt to explain the same type of variation in the outcome. This typically results in volatility of the model coefficients but does not reduce the reliability of the model overall. Because our goal is to have models that are as complete as possible, and because we do not interpret the coefficients of these variables, we proceed with using all variables in our models, acknowledging this multicollinearity issue. However, we do use variable selection methods, discussed in section R.5.3.4, to reduce the complexity of resulting models.

R.5.3 Modeling Techniques

R.5.3.1 Ordinary Least Squares Regression

Ordinary Least Squares (OLS) regression is an analysis technique that evaluates a linear relationship between a set of independent variables—which for us are integration levels, the USMC personnel variables, and the dependent variable (or the outcome of interest). OLS can be used to evaluate the simultaneous impact of all these variables and identify which variables predict the response the best. It also enables us to comment on how each variable affects the outcome, holding other variables constant. This is especially useful as we try to evaluate whether integration levels of the units matter, given the physical attributes of all the trial participants.

OLS regression models assume that the residuals are independent and identically distributed (IID) per the normal distribution with a mean of zero. We do not check this assumption, because the poor fit of the models precludes us from recommending them as useful.

R.5.3.2 Linear Mixed Models

The traditional OLS model, described above, assumes “fixed effects” for the variables that it includes. This means that the variables take only the fixed values. For instance, gender can take values of male and female only, so it is considered a fixed effect.

A random effect, by contrast, can be evaluated for a variable for which we observed only a random sample in our dataset, or a variable that contributes to a hierarchy of observations. For instance, we could say that all the Marines in our data represent a sample of the USMC, and, furthermore, that to the extent that the same Marine is placed in the same billet more than once, the data are nested in that way. Thus, we can include a random effect for the volunteers who fill each billet. This not only models the fact that our volunteers are a sample from a bigger population but also helps us explicitly account for the fact that we sample with replacement and our trials are not, strictly speaking, independent of one another. A model that incorporates fixed and random effects is called a *mixed model*. Whenever sample size was sufficiently large to generate reasonable model estimates, we ran mixed models due to these advantages.

Mixed models assume that the residuals are IID normal *for each level of the random effect*. Because each Marine participated in only a handful of trials—thus providing a very small sample of residuals with which to test this assumption—it is not testable in our models.

R.5.3.3 Multinomial models

We use multinomial logistic regression to model Shoulder-launched Multipurpose Assault Weapon (SMAW) accuracy in Annex E of this report. Each squad got to fire four shots, so the only possible outcomes of the model are 0, 1, 2, 3, and 4 when we are counting the number of hits on target. The model assumes that using the covariates—which include integration levels and our USMC personnel variables—each trial outcome can be classified into one of the five categories of the number of hits. Then the model quantifies the relationships between these independent variables and the categories, and provides a predicted probability that each integration level has of obtaining 0, 1, 2, 3, and 4 hits per target.

R.5.3.4 Variable selection

Prior to model fitting, we selected a set of individual covariates that the Marine Corps records on each Marine that could theoretically be related to the outcomes we modeled. Each model that we fit started out with as many of these covariates as was possible, given the sample sizes for each task. After the models were fit, however, we sought to reduce their complexity by taking away the variables that were not related closely enough to the outcome. We did so using the Akaike Information Criterion (AIC), which can help select a model out of many candidate models by quantifying the tradeoff between model complexity and the goodness of fit.

Specifically, the calculation of AIC includes the number of model parameters and the maximized likelihood. The number of model parameters represents a penalty for model complexity and overfitting; the model with the lowest value of AIC is considered the best balance of fit and complexity. Note, however, that although AIC can be used to compare models, it ultimately will not indicate if all models under consideration fit poorly, which was the case for some of our results.

Whenever it was possible to estimate a model using maximum likelihood methods, we used AIC for variable selection and commented only on the reduced models. We always started out with the fullest model we could fit and reduced it by eliminating one variable at a time, a technique known as *backward elimination*.

When maximum likelihood estimations failed to converge, we could no longer compute the AIC and had to comment on individual variables' significance in the full model.

Note that we used AIC even in cases where we modeled the result as a function of integration and one variable at a time, to see whether the model needed to include integration and that variable or just one of them.

R.5.3.5 Modeling when sample sizes are too small

In many cases, we did not have sufficient sample size to run full models that include all covariates for all trial participants. For instance, a rifle squad is composed of 12 individuals, meaning that we would have to include 12 variables for each covariate of interest. The problem grows quickly with the number of covariates, and the sample size did not support running full models for many tasks. Thus the next step in our analysis was to run the models for one variable at a time and include the integration level to see if any patterns emerged. For instance, we would model the group outcome as a function of the integration level and the height of all participants to see if all or most heights were predictive of the outcome. Next, we regressed the outcome on integration level and weight of all the participants, and so on. In the modeling sections of the annexes, we comment on whether the integration level was significant in these models, as well as whether the individual variables were significantly correlated with the outcomes.

As far as deciding how to model continuous data, we followed the following logic for choosing models:

1) Attempt to run a full mixed-model with all variables for all participants.
2) If 1) fails, attempt to run a mixed model with one variable at a time.
3) If 1) and 2) fail, attempt to run an OLS regression model with all variables for all participants.
4) If 1), 2), and 3) fail, run an OLS regression model with one variable at a time.

For the extremely small sample sizes of the mountaineering portion of the experiment, modeling the outcomes did not make sense because of the large number of participants in each trial. Thus, we limited those analyses to descriptive statistics only.

R.5.3.6 Interpreting model results

As discussed in section R.2.4, the main shortcoming of OLS regression is that we are forced to assume a very specific functional form. In most cases, the models did not fit the data well, giving results that did not make much sense from a theoretical or a policy perspective. This is also likely a consequence of the multicollinearity of the data, established in 0. As is common in linear models (where independent variables are strongly correlated), many of our significant model coefficients had signs opposite of what one would expect (less fit Marines performing tasks faster), or contradicting signs (a variable having a positive impact on the group response for one of the team members, but a negative for another).

An additional complication was presented when we could not fit models with all the personnel variables for all the trial participants due to sample size issues. In these cases, we had to resort to fitting one model per variable. For instance, we would regress the outcome on the integration level of the unit and the age of each participant. The interpretation of the model results would be whether the ages of individual participants are correlated with the outcome of the unit. Then we would repeat this analysis for weight and other variables. The limitation of this type of analysis is that we never get to see how each variable affects the response while other variables are held constant. In addition, it is difficult to combine the results of such models into one final model.

With a few exceptions, our models did not identify any discernable patterns. For instance, it was typical to see that a variable mattered for one group member but not the rest. The policy implication of such a finding is unclear: an Infantryman, for instance, is expected to be able to perform any duty out of the 12 squad positions, not just one. Thus, such findings could not be used, effectively, to shape gender-neutral standards for an MOS, as opposed to a finding where a variable consistently came out as significant for all or most group members. In most of our models, no discernable patterns were obvious, limiting us to considering only integration level as a plausible, useful predictor of the response.

R.5.3.7 Lanchester models

Lanchester's laws are common analysis techniques used to model battles by comparing strengths of two military forces over time. We use Lanchester's Square Law, which has assumptions that are more in line with contemporary Infantry engagements, in which fire can come from multiple directions and multiple targets can be attacked. The models assume that each unit can kill only one equivalent unit at a time, and our simulations further assume that the enemy units are comparable to the experimental all-male units in strength and equipment. The last assumption is notably strong, but it presents a conservative scenario that is worth considering. The models ultimately use differential equations to help predict the winner of an engagement and forecast the rate of attrition and losses on both sides.

R.5.4 Readiness Data

Data on individual readiness look quite different from our group performance data, so they are analyzed differently from the rest. Injuries and related data are recorded for individual Marines, and some have multiple observations. This section details the analysis techniques used to make conclusions about differences in readiness by gender, accounting for other variables.

R.5.4.1 Modeling counts

For our models of numbers of medical visits or days missed from the experiment, the outcome of interest is a count of days. There are two distributions that are frequently used to model counts: the Poisson distribution and the Negative Binominal distribution. We use the Negative Binomial distribution in our analyses because our data are overdispersed and the variance is much bigger than the mean (the variance is equal to the mean for the Poisson distribution).

A complicating factor in these analyses is that the outcome of interest is zero for most volunteers: most Marines did not miss experimental days; many did not get injured or go to see a medic. For data like these, where the number of zeros is more than one would expect (given the Negative Binomial distribution), a model that explicitly takes this into account is appropriate. This model is called the Zero-Inflated Negative Binomial model, which models the outcome in two parts: The first part of the model simply predicts, for each person, the likelihood of a non-zero outcome. For instance, for each person, it predicts (given a set of covariates) the probability of visiting a medic at all or missing at least one experimental day. Then the model predicts the number of times the event will happen for those whose outcome is nonzero. In other words, given that this Marine will miss experimental days, based on his personal characteristics, how many days do we expect him to miss? The data did not allow for estimation of the first part of the model using all the personnel variables, so we used only gender. However, the count part of the model did include additional personnel variables.

R.5.4.2 Time to first injury or medical visit

In our analysis of who is likely to get injured sooner, we modeled the number of days to first occupational injury, as well as the number of days to first medical visit for all volunteers. This analysis was done using a Cox Proportional Hazards model—a technique in survival analysis that is frequently used to analyze time to an event. The purpose of it is to model the amount of time before an event occurs as a function of observed covariates. However, the data are censored: we stopped observing Marines when they dropped out of the experiment or when the experiment was over, whichever came first. Survival analysis models two attributes of the risk of the outcome of interest simultaneously: how it changes over time and how it varies with independent variables included in the model. The model results in coefficients indicating the effect of each variable, as well as p-values indicating the variables' statistical significance.

R.5.4.3 Frailty analyses

Over the course of the experiment, volunteers often got injured more than once. Moreover, it is likely that that those injuries are not independent of one another: once a person gets hurt, they are more likely to get hurt again than they were to get injured in the first place. This can be because they are simply more prone to injuries or because some injuries or illnesses tend to manifest themselves multiple times. To model risk of injury (per unit of time, an experimental day in our case), we use a *frailty model*.

Frailty models are similar to survival models in the sense that the outcome of interest is an instantaneous hazard rate. However, a frailty model can be used to describe not just time to the first event but also time between events. This is done by incorporating a random effect (the frailty), which has a multiplicative effect on the hazard—meaning that once someone experiences an injury, he or she is considered more prone to other injuries. Thus this model is particularly useful when we have repeated measures for some individuals. The random effect in the frailty model allows the estimation of the hazard rate for a heterogeneous sample so that individuals have different hazards of getting injured.

In a frailty model, two distributions need to be specified ahead of time: the distribution of the baseline hazard, and that of the frailty. Several choices exist for each, but it is possible to use AIC and Bayesian Information Criterion (BIC) to pick the optimal combination of both. For our readiness data, the two criteria chose the Weibull distribution for the hazard and the Gamma distribution for the frailties. The estimation of the model is based on maximizing a marginal likelihood. The results show the impact of included risk factors on a person's predicted frailty rates and include statistical significance.

R.5.5 Population Variables

We compare the distribution of personnel variables in the GCEITF to that in the USMC at large to evaluate whether our volunteer sample is representative of the population with respect to these variables. To do so, we compare GCEITF males to the USMC at large and also USMC males *within their MOS,* and GCEITF females to USMC females at large (except for 1371s, for whom we provide comparisons within MOS and to all

USMC females). The comparison group was picked out to match GCEITF on pay grade and other dimensions detailed in Annex Q.

To compare our sample to the population—since we are interested in the entire distribution rather than comparing a particular statistic, like the mean—we use the Kolmogorov-Smirnov (K-S) test. The K-S test is a nonparametric technique that can help decide whether it is likely that a sample came from a particular reference probability distribution. To do this, the test compares the empirical distribution function for each variable in our GCEITF sample to the cumulative distribution function of the comparison population. A significant p-value in this test is an indicator that the two distributions differ, and that, on the relevant variable, the sample may not represent the population well.

THIS PAGE INTENTIONALLY LEFT BLANK