NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

ENHANCING INFANTRY TRAINING EVALUATION: THE CASE FOR QUANTITATIVE COMPARISON OF UNIT PERFORMANCE

by

Kevin Benedict

June 2023

Thesis Advisor:	Thomas W. Lucas
Second Reader:	Paul Johnson (MCOTEA)

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE				*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.				
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2023	**3. REPORT TYPE AND DATES COVERED** Master's thesis		
4. TITLE AND SUBTITLE ENHANCING INFANTRY TRAINING EVALUATION: THE CASE FOR QUANTITATIVE COMPARISON OF UNIT PERFORMANCE				**5. FUNDING NUMBERS**
6. AUTHOR(S) Kevin Benedict				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000				**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A				**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.				**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Marine Corps Training & Readiness (T&R) standards assess individual and unit performance on a binary or subjective scale that can be enhanced through data analysis. Currently, commanders must rely solely on experience and competence to match the qualitative standards to their unit's perceived performance. Comparing quantitative standards can provide the necessary context to better train and assess individual and small unit performance. Data from the Ground Combat Element Integrated Task Force (GCEITF) is analyzed for T&R tasks for which units do not have the means to collect data and T&R tasks for which results can be quantitatively measured. The GCEITF data is the best collected data on individual and small-unit infantry operations that the Marine Corps has, and continued analysis can provide valuable analytical insights. The final product is an interactive dashboard that allows a user to upload training data and outputs summary statistics and figures to visualize unit performance compared to the GCEITF performance. Recommended future work is creating a data repository that stores all uploaded data to keep the comparison metrics current, and expanding the number of tasks the comparison tool supports.

14. SUBJECT TERMS			**15. NUMBER OF PAGES**
infantry, training, USMC, metrics, standards, measures, performance, U.S. Marine Corps, GCEITF, Ground Combat Element Integrated Task Force			87
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

ENHANCING INFANTRY TRAINING EVALUATION: THE CASE FOR QUANTITATIVE COMPARISON OF UNIT PERFORMANCE

Kevin Benedict
Captain, United States Marine Corps
BS, Lemoyne College, 2017

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2023

Approved by: Thomas W. Lucas
Advisor

Paul Johnson
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Marine Corps Training & Readiness (T&R) standards assess individual and unit performance on a binary or subjective scale that can be enhanced through data analysis. Currently, commanders must rely solely on experience and competence to match the qualitative standards to their unit's perceived performance. Comparing quantitative standards can provide the necessary context to better train and assess individual and small unit performance. Data from the Ground Combat Element Integrated Task Force (GCEITF) is analyzed for T&R tasks for which units do not have the means to collect data and T&R tasks for which results can be quantitatively measured. The GCEITF data is the best collected data on individual and small-unit infantry operations that the Marine Corps has, and continued analysis can provide valuable analytical insights. The final product is an interactive dashboard that allows a user to upload training data and outputs summary statistics and figures to visualize unit performance compared to the GCEITF performance. Recommended future work is creating a data repository that stores all uploaded data to keep the comparison metrics current, and expanding the number of tasks the comparison tool supports.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

CASEVAC	casualty evacuation
CMC	Commandant of the Marine Corps
CSV	comma separated values
DOD	Department of Defense
DON	United States Department of the Navy
EIM	Expectations for Infantry Marines
ETJ	electronic training jacket
FTM	functional test managers
GCEITF	Ground Combat Element Integrated Task Force
HSR	human subjects research
IAR	infantry automatic rifle
MAAWS	multi-role anti-armor anti-personnel weapon system
MAGTF	Marine Air Ground Task Force
MCAGCC	Marine Corps Air Ground Combat Center
MCOTEA	Marine Corps Operational Test and Evaluation Activity
MCTIMS	Marine Corps Training Information Management System
MET	mission essential task
MFP	mortar firing position
MOS	military occupational specialty
MOVES	modeling, virtual environments, and simulation
MSPS	MOS specific physical standards
MS-SQL	Microsoft structured query language
MWTC	Mountain Warfare Training Center
MWX	MAGTF warfighting exercise
NPS	Naval Postgraduate School
PI	provisional infantry
SAL	skill acquisition level

SAT	systems approach to training
SBF	support by fire
SMAW	shoulder-launched multipurpose assault weapon
SOI-E	School of Infantry-East
T&R	training & readiness
TECOM	Training and Education Command
TOF	time of flight
TTP	tactics, techniques, and procedures
UTM	unit training management
WPP	weapons player pack

EXECUTIVE SUMMARY

This research is focused on infantry training and performance at the individual and small-unit level with two established goals. The first goal is to use data from the Ground Combat Element Integrated Task Force (GCEITF) to find useful insights that were not covered in the GCEITF *Experimental Assessment Report*. The GCEITF was an experiment conducted by the Marine Corps that at its height had 264 Marine volunteers, lasted over one year, and incurred a total cost of $36 million (Marine Corps Operational Test and Evaluation Activity 2015). Since the GCEITF had equipment that is not widely available, like sensors on weapons and targets, and expert-level data-collection techniques tracked all data, it is unlikely that the Marine Corps will ever collect data in the same manner and on the same scale again. Therefore, the focus of this research is to find insights from the data that are not likely to be collected again. In practical terms, this means focusing on individual and crew accuracy across various weapons platforms.

The second goal is to demonstrate the benefit of enhancing training standards with quantitative components that can be measured and compared to other unit's performance. In the Marine infantry, the planning and execution of training are both done exceptionally well. However, the evaluation and assessment of that training oftentimes lacks the preparation and scrutiny given to the planning and execution phases. The standards by which all training is evaluated are given in NAVMC 3500.44D, the *Infantry Training & Readiness Manual* (Department of the Navy [DON] 2020). The standards that are quantitative in nature are written in a binary manner that does not allow for depth of understanding beyond pass/fail, and those that are qualitative in nature are too subjective to allow for a collective definition or understanding of successful execution. In implementation, findings from the GCEITF were used to show what a "Marine Corps average" metric could look like.

Adding a quantitative and objective component to training and readiness (T&R) tasks will help evaluators and leaders provide specific and relevant feedback that will ensure training value is maximized. Up until now, units have been able to rely on resident expertise to evaluate and assess proficiency in execution. However, the United States has

now entered a great power competition phase with the People's Republic of China. In recognizing the threat of conflict with China, and the warfighting potential of emerging technologies, there is drastic organizational change within the Marine Corps. Personnel management, training and education, equipment, and unit organization are all experiencing revolutionary changes that modernize the force and therein lies the problem with relying solely on resident expertise to evaluate performance: the missions that Marine units will be tasked with, the gear they will use, and the capabilities of the young Marines who will accomplish the mission will all be different than they were in the past. Thus, it is paramount that there are means available to assess performance given this new environment.

The general workflow to establish quantitative components for applicable T&R tasks began as follows: consider individual, crew, or section level T&R tasks that could benefit from a quantitative component to supplement the current standard; review the GCEITF *Experimental Assessment Report* to see if that task had already been considered; and then search through the nearly 300 gigabytes of data available from the GCEITF to see if the data supported the task of interest. After selecting the raw data, Python and R were both used for cleaning, processing, and visualization. As an example, consider the T&R task 0341-WPNS-1005, "Fire a Mortar in Handheld Mode" with the standard "to achieve effects on target" (DON 2020, p. 14-11). This standard is intentionally vague, as the desired effects of fire support are unique to the situation. However, there can still be quantitative descriptors that bring perspective to the capability of the unit and the weapon system. Using GCEITF data to serve as a benchmark for comparison, expectations can be developed to put in context what average performance for that task is. Furthermore, information specific to weapons employment, like accuracy and round adjustments, can be determined as shown in Table 1 and Figure 1.

Table 1. GCEITF Handheld Mortar Impact Accuracy

Impact to Target Distance	Number of Rounds within Threshold Distance	Percentage of Rounds within Threshold Distance
35m	103	23%
50m	164	37%
75m	260	60%
150m	425	97%
200m	440	100%

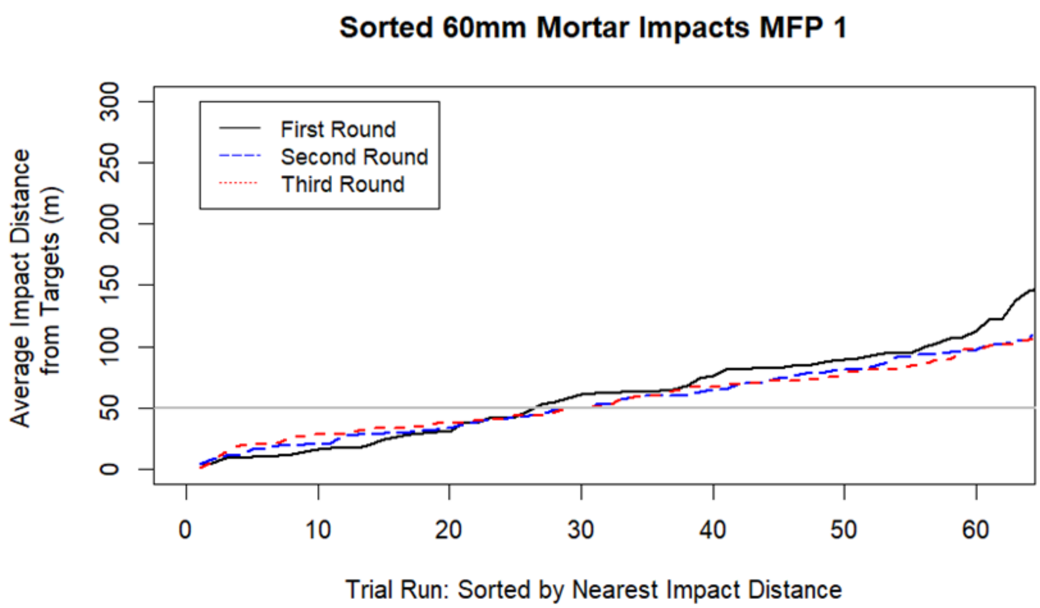


Figure 1. GCEITF Mortar Impact Distance Plot

This process was repeated through twelve iterations for tasks applicable to 0311 Rifleman, 0331 Machine Gunners, and 0341 Mortarman. Examples of findings are given here:

- For 0311 Rifleman, conducting fire and movement as part of a squad attack, 22% of rounds expended are hits and 43% are effectively fired (meaning a hit or within one meter of the target). Of the 43% of effectively fired rounds, 65%-85% are hits when the shooter is within 100 meters of the target. Farther than 100 meters, less than 50% of the effectively fired rounds are hits.

- For 0331 Machine Gunners, physical fitness is paramount to providing timely offensive medium machine gun support. Over the course of the trials, there is consistent correlation with movement times throughout the sequence of support. The Marines who began the initial movements quickly also finished the final movements quickly, as opposed to fatiguing throughout the event.
- For 0341 Mortarman, employing 60mm in handheld mode, 23% of rounds fired landed within the doctrinal measure of suppression of 35 meters. Extending the concept of effective fire to mortars, 60% landed within 75 meters of the target. Additionally, corrections were made effectively. If the initial round fell more than 50m from the target, the following rounds landed closer to the target.

The next major effort was visualization and presentation of the results. For tasks that can be easily measured and repeated, like 0341-WPNS-1005 previously discussed, the visualization was done via an interactive dashboard to maximize utility. As new technologies and concepts are fielded within the infantry, measures of performance will be most useful when compared with the performance of adjacent units. Encompassed within this idea is the concept of lateral information sharing. By hosting an application that allows for a unit to upload training data, and see how their unit compares to others, commanders will gain the required perspective to accurately evaluate their performance. The GCEITF results were used as the benchmark comparison in the current implementation of the tool, called Expectations for Infantry Marines (EIM). However, the best implementation of the tool would allow for a data repository of user-uploaded data that over time would accurately represent the current Marine Corps proficiency in any given task.

There are three recommendations for future work on this topic. First, many live-fire ranges are supported by contractors with reactive targets. Working alongside the contractors, a standard report can be developed that takes unit data at the completion of each range and provides accuracy information to the unit leaders. This maximizes the training benefit of Marines conducting the range and the money spent on contractor support. Next, recommended data collection procedures and evaluator notes can be created for a subset of T&R tasks. The purpose of these evaluator notes is to provide a quantitative benchmark for an evaluator to quickly understand the context with which the subjective standard is written in. Alongside the evaluator notes, a more robust version of EIM can be developed that has added detail in the data-collection template and allows for storing data to track progress over time. Last, the statistical information produced from the GCEITF data can be used as inputs to existing combat models to produce a more accurate output.

References

Department of the Navy (2020) Infantry training and readiness manual. NAVMC3500.44D. Washington, DC, https://www.marines.mil/Portals/1/Publications/NAVMC%203500.44D%20Infantry%20T-R%20Manual.pdf?ver=2020-05-21-104009-273.

Marine Corps Operational Test and Evaluation Activity (2015) Ground Combat Element Integrated Task Force: Experimental Assessment Report, Quantico, VA.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I owe the greatest thank you to my advisor, Dr. Tom Lucas, for his commitment to this project and his commitment to research that improves the Department of Defense. Next, thank you to Mr. Paul Johnson who accomplished the Herculean task of designing and executing the GCEITF study, which offers the best collected data on individual and small-unit infantry operations. Last, thank you to all the Marines I have worked alongside and learned from; the thoughts presented here are due to all the great guidance I've received from all of you throughout the years.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

> The dirty little secret of combat is that these young Americans will get the job done despite lackluster training and in spite of less-than-inspired leadership. The difference between that group and a group that is well trained and well led is the price of the butcher's bill in war.
>
> —McCoy (2011)

A. BACKGROUND AND IMPORTANCE

In the Marine infantry, units are either at war or preparing for war. In preparation for war, training evolutions are created that imitate wartime actions. These evolutions have three phases: planning, execution, and assessment. It is imperative that all three phases be completed as efficiently as possible, for both near-term practical and far-term consequential reasons. The primary practical reason for requiring efficient training is ensuring the responsible use of taxpayer resources, while the consequences of poor training are unnecessary casualties in combat.

Commanders are charged with assessing their units' capabilities, which in turn informs the planning process to sustain strengths or improve weaknesses. There is an inherent reliance upon the commander's experience and competence to accurately assess capabilities, provide feedback on performance, and skillfully plan training to increase capability. Doctrine should assist commanders in fulfilling this responsibility.

The guiding document for planning and assessing training is the NAVMC 3500.44D, *Infantry Training & Readiness Manual* (Department of the Navy [DON] 2020). The T&R Manual prescribes tasks with associated standards, conditions that the task must be completed under, sustainment intervals, step-by-step components, and additional references. The standards for tasks range from extremely specific ("By achieving a minimum score of 70%" for 0311-M27-1005) to broadly interpretable ("To employ demolitions in support of commander's intent" for 0351-DEMO-2001). What all standards have in common is the need for an evaluator to decide if the Marine (or unit) passed or failed according to the standard. In fact, current doctrine mandates this by stating "the training standards are the evaluation objectives" (DON 2015, p. 23). In the 0311-M27-

1005 example, an evaluator can simply calculate if the Marine being evaluated scored a 70% or higher. However, this metric lacks the context to truly assess performance beyond a binary pass or fail. For 0351-DEMO-2001, the evaluator must decide if the Marine passed or failed without any objective standard by which to compare. In all cases, the subjective and/or binary nature of the standards dilutes the quality of assessment and feedback.

The current version of the infantry T&R Manual has 851 tasks. The sheer volume of the T&R Manual makes it impossible for all commanders and evaluators to expertly assess performance and provide knowledgeable feedback. Furthermore, the role of the infantry within the broader force is changing. Across the Marine Corps, modernization efforts from both technological and force-structure perspectives are demanding a wider capability range from individual Marines and units. To meet this demand, experimental infantry battalions are being studied in an effort to mature the force. Marines at all levels will be expected to perform a broader and more difficult mission set, under less supervision, with increasingly complex technology. These changes also increase the need for standards which provide clear means of assessment for commanders.

In combat, units will fight how they were trained, whether good or bad. To that end, it is imperative that training plans are aligned to unit capabilities and limitations, and that performance assessment be relevant and specific. Furthermore, this needs to be done in a time-constrained environment. A unit will never have enough time to master all Mission Essential Tasks (METs) and T&R tasks that compose each MET given deployment schedules and personnel rotation. Therefore, identifying creative means by which unit proficiency can be quickly measured and assessed will save time and help to focus training efforts.

Incorporating quantitative performance metrics into existing standards does not take responsibility away from the commander to understand his/her unit's proficiency, nor does it suggest that assessment of infantry tasks can be done with complete objectivity. Rather, these metrics can assist commanders in appropriately assessing their unit's performance across the spectrum of operational requirements and plan training accordingly. These metrics become especially important for tasks that are not completed often, or tasks in areas which the commander does not have personal expertise in. While there is no excuse for professional incompetence, the complexity of the nature of the infantry precludes any individual from being an expert in all aspects of it. Understanding the distribution of task performance across the force can help to manage expectations and establish realistic baselines.

As the Marine Corps undergoes force modernization and embraces the realities of the modern-day threat and technological environments, there is an appetite for change to the status quo. Establishing and updating objective standards benefits adjacent units across the force. Training data is already consolidated on the Marine Corps Training Information Management System (MCTIMS), though it is currently entered as a binary metric for most tasks. Through existing platforms like MCTIMS, idea-sharing and training lessons-learned can be spread laterally to the benefit of the entire community. Instead of binary inputs being uploaded into a database for higher units to track training completion, quantifiable metrics could also be uploaded for adjacent units to view. A pass or fail line may still be established to definitively mark a unit as trained, but it could be contextualized by descriptive statistics calculated from thousands of other entries.

Without distributions showing performance at the average and extremes, there is an implicit expectation for unit leaders to apply subjective standards to their units without clearly defined measures of success. This lack of objectivity may result in unfocused training plans and overly generalized feedback. Quantifying below-average, average, and above-average performance, and making that information readily available to leaders across the force, can improve efficiency and effectiveness.

B. PROPOSED SOLUTIONS

The realization that the current system of recording training completion omits valuable information is not an original thought. All Marines have an Electronic Training Jacket (ETJ) that is maintained on MCTIMS that shows individual training completed since the Marine joined the service. This is valuable information and helps unit training managers identify which individuals in their units need to complete training according to prescribed deadlines. However, this information does not capture progression; rather, it tracks whether a Marine completed the task within the prescribed period. By capturing more information on individual and unit tasks, units can form an organizational memory by seeing trends in performance over time.

Leaders at the School of Infantry-East (SOI-E), a key service-level training command, have also recognized that the current framework for standards does not provide insight to a sufficient level of detail. The school has created a list of 39 core competencies, each of which encompasses multiple T&R tasks (Emmel et al. 2022, p.36). The core competencies are seen as critical skills that all infantry Marines must be capable of achieving. For each core competency, a rubric called a Skill Acquisition Level (SAL) is used to evaluate Marines. The SAL is a continuum that has Novice, Advanced Beginner, Competent, Proficient, and Expert categories. T&R tasks are modular in nature, progressing from individual to unit level events. Thus, by leveraging the category framework, the SAL for each core competency can assign T&R tasks to each category. A group of basic T&R tasks appears in each Novice category, then more complex T&R tasks are added to the next categories.

While the core competency framework still relies on evaluators to apply subjective standards on a by-task basis, it does enforce objective standards for determining proficiency. Furthermore, the implementation of the method has an important distinguishing benefit: it provides a feasible means for tracking performance over time. Instead of a training record that shows whether a Marine completed a task in the previous year, there is a marked line of where that Marine progressed to. Under the current system, the only means of assessing progression is by physically observing training over time. Documenting individual and unit performance history allows the commander to better analyze the impact of the training program, quickly assess the capabilities of new members joining the unit, and identify areas the unit lacks in.

The benefit of adding quantitative and objective measures to existing standards is a simple concept which is difficult to implement. Individual tasks that are time-based and accuracy-based can be collected quite easily, even for someone lacking a scientific background. The biggest challenge to overcome is data collection at the unit level. In 2015, the Marine Corps created the Ground Combat Element Integrated Task Force (GCEITF), a $36 million experiment lasting over one year from start to finish, that studied the effects of gender integration into ground combat arms (MCOTEA 2015a). The experiment was motivated by the Department of Defense mandate that required all Military Occupational Specialties (MOS) to be open to any gender. The study and following analysis focused on the performance differences between all male units and gender-integrated units. It is unlikely that a study of this scale or level of detail will ever be done again in the Marine Corps, so it is important to draw all available insights the data can provide beyond gender's effect on performance.

The work detailed in the following chapters uses data from the GCEITF to establish a quantifiable baseline performance for T&R tasks for the following MOSs: 0311 Rifleman, 0331 Machine Gunner, 0341 Mortarman, and 0351 Assaultman.

The final output of this work is an example dashboard titled Expectations for Infantry Marines. From a mathematical standpoint, the word *expectation* refers to the average of a random variable which is fitting for the descriptive statistics presented. More importantly, the dashboard is intended to provide realistic expectations for infantry leaders when assessing their unit's performance. The intent of the dashboard is to show that applying basic statistical concepts can enhance the lens through which we view our training and evaluation protocols and make better time-management decisions. The infantry leader only has two jobs: win in combat and prepare Marines for combat. Quantitative and objective standards will help with both.

THIS PAGE INTENTIONALLY LEFT BLANK

II. TRAINING METHODOLOGY

> You need to make sure that your training is so hard and varied that it removes complacency and creates muscle memory—instinctive reflexes—within a mind disciplined to identify and react to the unexpected.
>
> —Mattis (2019)

A. TRAINING

The Commandant of the Marine Corps (CMC) is charged with manning, training, and equipping the force to provide Combatant Commanders combat-ready units across the globe. When General Berger assumed the role of CMC, he implemented changes that are aimed at modernizing the force to counter the pacing threat of China in the Indo-Pacific region and summarized those changes in writing in *Force Design 2030.* On the matter of infantry training, General Berger wrote "I concur with [Integrated Planning Team] conclusions that our current entry level and advanced infantry training program and policies will not meet future demands of our infantry elements" (U.S. Marine Corps 2020, p. 11). From the CMC's perspective, the infantry force of the future will be tasked and tested at a higher level than ever before. He goes on to say, "explore ways to challenge existing models and paradigms to yield a more capable and mature infantry and reconnaissance force" (U.S. Marine Corps 2020, p. 11). This research aims to meet the CMC's intent of providing a means to yield a more capable infantry force.

The Marine Corps currently uses the Systems Approach to Training (SAT) methodology to meet training requirements. The SAT process consists of five phases: analyze, design, develop, implement, and evaluate (DON 2016, p. 3-1). The SAT process can be improved by objective training standards in the analyze and evaluate phase. Leaders are expected to "assess unit's strengths and weaknesses in support of the mission and commander's training guidance" (DON 2015, p. 2) in the analyze phase, and "review training deficiencies of subordinate units and individuals" (DON 2015, p. 2) in the evaluate phase, but often do so through a limited scope of experience. These can be immediately improved when properly benchmarked with data from across the force. In fact, this line of thinking already exists in doctrine, though not in implementation. Standards-based training is a principle which states "the use of common procedures and uniform operational methods creates a common perspective regarding unit and individual performance within the Marine Corps" (DON 2015, p. 4). This is a worthy goal, but difficult to accomplish without the aid of data analysis and a means to share that information.

To implement the SAT methodology, commanders use Unit Training Management (UTM) which is "the use of SAT and Marine Corps training principles that maximize training results and focuses unit training priorities on the wartime mission" (DON 2015, p. 1). UTM is, essentially, guidance for the development of a training plan that is tailored to each unit and their expected combat mission. Much of that guidance is focused on the planning and evaluation of training; thus, it is aligned with the analyze and evaluate phases of the SAT process. In addition to personal proficiency, current doctrine urges commanders to seek subordinate opinions and states "The Commander uses the broad experience and knowledge of key subordinates to help determine the organizations' current proficiency" (DON 2016, p. 5-5). Relying on subordinate expertise to determine unit proficiency is applicable at all levels, but likely works best in larger commands. As an effect of the law of large numbers, there is likely little variance in performance or proficiency between divisions, regiments, and even battalions. However, at the company level and below, there is going to be a larger variance on the average performance of that unit. The ideas suggested in this research are not meant to influence divisions, regiments, or battalions; rather, success is measured by the immediate insights that Sergeants leading squads and Second Lieutenants leading platoons can gain by numerically comparing their Marines' performance against the performance of other Marines across the force. Even given a small unit with the best subordinate leaders, statistical principles suggest that further insights can be gained with a larger sample size. In an age of instant information sharing, inputs from across the force can easily produce the entire community's *current* proficiency, which will give all leaders a valuable benchmark by which to compare their unit.

B. EVALUATION

Doctrinally, a standard is defined as "the level of proficiency to which the behavior must be performed" (DON 2010, p. 3-12). A standard must meet four criteria:

completeness, accuracy, time, and realistic. This definition suggests objectivity across all standards and sets the minimum performance required to be considered trained according to the standard. The existing standards that do meet this definition only capture the minimum information required to determine pass or fail performance. Whether performance is assessed against a by-task distribution, or against a group of tasks as in the SAL method, information beyond minimum passing standard is crucial to properly evaluate Marines' performance.

Although often used interchangeably, there is a useful differentiation between testing and evaluation. Doctrinally, a test is defined as something that "measures proficiency against established standards and results in a pass-fail rating" (DON 2016, p. 7-4) while an evaluation "seeks to determine, by assessing strengths and weaknesses, where in the performance scale an individual or unit is at a given time" (DON 2016, p. 7-4). This is an important distinction to understand, and this research focuses on evaluation. In practice, testing is the primary concern as there is a required entry in MCTIMS for task completion. This research seeks to aid commanders in evaluation as previously defined by creating a data-driven performance scale available for reference.

Current doctrine gives four methods for evaluating training: post-training checks, sampling, on-the-job observations, and evaluations by higher headquarters (DON 2015, p. 21). All four have in common the mandate to evaluate and assess Marines' performance with an assumption that the evaluators are subject matter experts for the task being performed, stating "they [evaluators] should be basing their evaluation on their proven abilities and recent experiences in accomplishing the same tasks" (DON 2015, p. 22). It is clear that evaluators should leverage their experience, but that is far too limited of a perspective to rely on as compared to community-wide comparison. Furthermore, the T&R task standards which form the basis of training programs are the evaluation objectives. If a standard is too vague, then the goal of common perspective and performance across units will not be reached. The most scientific method of evaluation is sampling, where an evaluator randomly selects ten percent of a unit to complete a task and applies the results as an assessment of the broader unit's proficiency. While this method is attempting to apply statistical principles, it may not be mathematically sound given the variance in individual proficiency, unit size, and task distribution.

C. GCEITF

1. GCEITF Background Information

In 1994, the Secretary of Defense signed a memorandum that stated, “service members are eligible to be assigned to all positions for which they are qualified, except that women shall be excluded from assignments to units below the brigade level whose primary mission is to engage in direct combat on the ground” (Office of the Secretary of Defense [OSD] 1994). This became known as the 1994 Direct Ground Combat Definition and Assignment Rule and was intended to “expand opportunities for women” (OSD 1994). In January 2013, the Secretary of Defense eliminated the 1994 rule, and required each service to implement this change “after the development and implementation of validated, gender-neutral occupational standards” (OSD 2013). The Marine Corps then set out to create MOS Specific Physical Standards (MSPS), beginning with ground combat arms MOSs. Through physically demanding tasks representative of the respective MOS, the MSPS seeks to identify whether individuals have the physical fitness required for that MOS. The MSPS are a great example of well-defined objective training standards that improve force lethality.

The memorandum also added that any exceptions, meaning MOSs that would remain closed to women, must be “narrowly tailored, and based on a rigorous analysis of factual data regarding the knowledge, skills, and abilities needed for the position” (OSD 2013). The Marine Corps sought to objectively identify what effect, if any, integrating women into combat arms would have on lethality. Thus, the GCEITF was formed. Mr. Paul Johnson, Scientific Advisor at Marine Corps Operational Test and Evaluation Activity (MCOTEA) was selected to design and lead the study. An experiment of this size was noteworthy in itself, but the underlying motivation for the experiment created a politically charged atmosphere that captured the focus of the Marine Corps, the Department of Defense (DOD), and the national news on several occasions.

Alongside Dr. Jane Pinelis, lead statistician for the experiment, Mr. Johnson documented the work in a book titled *The Experiment of a Lifetime.* One challenge faced for the design of the experiment was determining how to quantify the difference in performance of infantry tasks between all-male and gender-integrated units. The root of this problem is even broader; how can one quantifiably measure small-unit infantry squad performance? Contextually, there are two facets to this question: statistically significant, and practically important. Most measured data can be analyzed with a hypothesis and tested for statistical significance. However, the key is finding tests that have practical importance for the infantry force. In other words, the study sought to avoid an experiment in which statistically significant differences could have been found that had no impact on the success or failure of the infantry task. To solve this problem, a series of interviews was conducted with Marines from MOS fields being studied. Importantly, there was a wide range of answers given to the interview questions asked which highlights the subjectiveness and variation that occurs across the force, thus further justifying the requirement to bring objectivity to training standards. The researchers' goal was to find a numerical value to match to below-average, average, and above-average performance for a given infantry task. At the end of the interviews, it was found that a 30 percent difference in performance was an acceptable heuristic to deem the result practically important. Ms. Pinelis wrote "a 30 percent difference stands out because it could be easily noticeable by Marines: it was when they instinctively knew, without any instrumentation, that the task was taking too long" (Johnson et al. 2019, p. 49). Essentially, this 30 percent difference serves as the lower bound for accepted performance across tasks. While the analysis of this work does not exclusively use the 30 percent difference heuristic, it is an important point to highlight for two reasons: a lot of effort went into discovering it, and the simplicity is such that nearly any unit leader can apply it.

In addition to determining how to measure differences in task performance, a decision had to be made as to which tasks would be tested. To achieve this, Marines representing their MOS community titled Functional Test Managers (FTM) were assigned to provide expert opinions throughout the experiment. Mr. Johnson gave guidance to FTMs to ensure all selected tasks met the following criteria: applicable to all missions across the range of military operations, critical to success of Marines in the occupation, physically demanding, and able to be completed without relying on years of occupational experience. Thus, with this guidance and conversations within their communities, FTMs selected those tasks which served as the foundation for proficiency. These tasks are analyzed later in further chapters holistically, to be used as an example for objective measures that can accompany current T&R tasks, as well as by parts to draw additional insights relevant for infantry leaders.

The conceptual idea of the experiment now needed to become a reality. Since the experiment was categorized as Human Subjects Research (HSR), all Marine participants were assigned to the GCEITF voluntarily, and could quit at any time. From a scientific perspective, voluntary participation raises the question of whether the sample is representative of the population. These concerns were largely dismissed when volunteers were sought from across the force, all with different motivations for joining the study and unique backgrounds that mitigated the effects of a biased sample. A similar consideration was sample size requirements for each task. Dr. Pinelis, as the lead statistician, relied on past data to create initial sample size requirements. Since the goal of the study was determining whether there are performance differences between all-male and gender-integrated units, the null hypothesis for each test is that there are no differences between the groups. This is important statistically as there must be sufficient evidence in favor of the alternative hypothesis, in this case performance differences between the two groups, for the null hypothesis to be rejected. Designing the test in this manner mitigated bias and helped to defend from any criticism that suggested a reluctance to allow females in combat arms. For the experimental analysis, a significance level of 0.10 was selected as were varying effect sizes of 30%, 38%, and 46%, representing small, medium, and large effect sizes, respectively. Using previous data from MCOTEA testing allowed for the estimations of standard deviations to influence the effect size estimates. Finally, a power level of 0.80 was chosen, which provides the probability of correctly rejecting the null hypothesis given the null hypothesis is false by a specified amount (MCOTEA 2015a). The resulting minimum sample sizes to start the experiment are given in Table 1.

Table 1. Minimum Sample Sizes Required. Adapted from MCOTEA (2015a)

	Rifleman 0311	Machine Gunner 0331	Mortarman 0341	Anti-Tank Missileman 0352	Combat Engineer 1371
Female	6	6	6	4	6
Male	66	12	18	13	18

These details are important to highlight the mathematical validity of the sample size of the study and give caution to the deviations presented in further chapters. The sample sizes were selected such that performance differences between genders could be detected; however, this research largely ignores gender differences. Therefore, sample size concerns are voiced when necessary. Table 2 shows the number of volunteers at the start and end of the trial runs.

Table 2. Number of Volunteers by MOS. Adapted from MCOTEA (2015a)

	Rifleman 0311	Machine Gunner 0331	Mortarman 0341	Anti-Tank Missileman/ Assaultman 035X	Combat Engineer 1371
Start	45	12	13	12	26
End	34	11	11	9	21

For some MOSs, the number of participants is below the minimum threshold due to volunteers being injured or quitting. However, the design of the experiment focused on randomization of participants. That is, each fire team or squad completing a task was formed by randomly selecting individuals based on gender. By repeating trials with this random selection process, and relying on the Central Limit Theorem, individual differences between Marines were averaged out. For the infantry, each trial consisted of a two-day cycle. Day one was offensive operations, conducting a rifle squad-attack supported by machine guns and mortars and ending with a casualty evacuation. Day two was defensive operations, conducting a 7km hike, digging defensive positions, and heavy weapons employment. The combat engineers also had a two-day cycle; day one was offensive operations consisting explosive breaching, demolitions raid, and obstacle emplacement. The second day was defensive operations, with Marines performing a 7km hike, route clearance operations, and cache reduction. For both the infantry and combat engineers, at the end of day two, new groups were randomly assigned. Individual Marines participated in a cycle that was four days on and one day off. Trials began on 7 March 2015 and ended 26 April 2015. Additional testing was conducted in May of 2015 aboard Mountain Warfare Training Center (MWTC) in Bridgeport, California. MWTC is composed of rugged mountainous terrain high in the Sierra Nevada mountain-range, which makes the conduct of infantry operations exceedingly difficult. The trial cycle tested a squad's ability to conduct a logistical resupply. Each squad conducted a 4.6km hike, crossed a gorge using single-rope bridges, climbed a 40-foot rock wall, and made a return 5km hike. Each individual participant completed a one-day cycle and then rested for one day.

At the end of the study, conclusions were documented in an "Experimental Assessment Report" (MCOTEA 2015a). Some conclusions found statistically significant differences in performance between all-male and gender-integrated units. Ultimately, there was no service exemption from the Secretary of Defense and today there are female infantry Marines serving across the force.

2. Previous Work with Dataset

Major Martin Meehan, a 2020 graduate of the Naval Postgraduate School (NPS) Operations Research program, began efforts to study GCEITF data in his thesis titled "Moving the Corps into the Information Age: Data-Driven Training Standards and Analytics to Support Evaluation." The data he received was formatted as Microsoft Structured Query Language (MS-SQL) tables, and much of his work was converting the data into Comma Separated Values (CSV) files for ease of analysis. An important output of his work was the creation of a data dictionary with the goal being "to provide a clear and concise guide identifying each of the parameters and their data type" (Meehan 2020, p. 31). In addition to the data dictionary, he presented several ideas which this research supports. First, he identified the ability to quantify training standards to inform training and proficiency evaluation which this research directly supports. Second, he suggested a reevaluation of current training standards in cases that data suggests a T&R task is incongruent with actuality. Last, he recommends developing range-specific evaluation. As mentioned previously, this research falls under a larger theme of lateral information sharing. For the data being analyzed, there is not a clear comparison between a squad attack on a specific range at Marine Corps Air Ground Combat Center (MCAGCC) Twentynine Palms California and a squad attack on a range in a different environment with different terrain. However, identifying metrics which should be measured and tracked on designed scenarios at ranges across the Marine Corps will allow for a higher fidelity of evaluation for units using those facilities. Major Meehan's efforts resulted in a ready-for-use dataset which can undoubtedly provide insights to infantry leaders.

THIS PAGE INTENTIONALLY LEFT BLANK

III. DATA ANALYSIS

> What we need in battle is more and better fire. What we need to seek in training are any and all means by which we can increase the ratio of effective fire when we have to go to war.
>
> —Marshall (1947)

A. INTRODUCTION

The analysis in this chapter was conducted solely with data from the GCEITF experiment. Sections are organized by MOS for 0311 Rifleman, 0331 Machine Gunners, 0341 Mortarman, and 0351 Assaultman. The goal of each section is to find an insight that is relevant to an infantry unit leader without simply redoing analysis that is already covered in the GCEITF *Experimental Assessment Report*. When applicable, the analysis is aligned to one or more T&R tasks.

B. 0311 RIFLEMAN

1. Marksmanship Methodology

Accuracy with an assigned weapon system is the hallmark trait of a Marine infantryman; ability in all other areas is null if, in the definitive moments of a fight, a Marine cannot accurately engage a target. The T&R Manual has "proficient marksman" as a core capability for all enlisted infantry billet descriptions and the Marine Corps requires all Marines, regardless of MOS, to qualify annually on a graded rifle range. For 0311 Rifleman, the T&R manual provides the following tasks directly related to accuracy with a rifle:

The five tasks in Table 3 exemplify that doctrinal standards can be either specific and objective, as seen in 0311-M27-1005 and 0311-M27-2002, or broad and subjective, as seen in the remaining three tasks. For the selected tasks with objective standards, the standard is to achieve 70% accuracy on a predetermined course of fire from a static position. These two tasks are intended to practice marksmanship before applying the skill in a more realistic setting. The next three tasks relate to marksmanship in a realistic setting, and the standards for each have only subjective grading criteria. The implication is that evaluators must decide, without any established baseline, if the individual or unit completed the task in a satisfactory manner. The reason the standards are so vague is because of the dynamic nature of infantry attacks. There cannot be one objective standard for how well a unit establishes a base of fire, or conducts fire and movement, that would encapsulate all the variables of the offensive mission, the enemy situation, and the terrain the unit is operating in. However, through data collection and analysis, expectations can be established, and comparisons can be drawn. Evaluators will still need to use discretion when comparing unit performance against expectations and baselines, but will ultimately be better informed. In addition to using data to compare performance, data can identify insights that may be unnoticed. The analysis within this section seeks to establish baselines for individual marksmanship while operating under realistic conditions to provide a baseline for expected performance and identify factors which contribute to marksmanship in the attack. Before explaining the results, it is necessary to cover the details of a live-fire squad attack.

Table 3. 0311 Rifleman Marksmanship T&R Tasks. Adapted from DON (2020).

Task: Description	Standard
0311-M27-1005: Execute the Infantry Automatic Rifle (IAR) Transition Course of Fire	By achieving a minimum score of 70%.
0311-M27-2002: Execute the Infantry Automatic Rifle (IAR) Unknown Distance Course of Fire	By achieving a minimum score of 70%.
INF-MAN-3014: Establish a base of fire	To suppress an enemy position to support the movement of friendly forces.
0300-RFL-1004: Engage Targets with the Service Rifle	To eliminate the threat(s).
0311-OFF-1001: Conduct Fire and Movement	To neutralize the enemy threat.

To replicate combat actions in training, live fire and maneuver ranges are created that mimic combat actions. One popular example is the squad attack range, where a squad of 12 Marines attacks a notional enemy position. A common tactic is to close the distance between the Marines' assault position and the enemy defense, and the most effective way to do this is by fire and movement. Fire and movement consists of two distinct yet mutually supporting elements: the "fire" element suppressing enemy forces and the "movement" element advancing towards the enemy forces. When the "movement" element reaches a covered position, that element now assumes the duties of the "fire" element, allowing the other to move. This exchanging sequence continues until reaching a limit of advance. Upon reaching the limit of advance, additional targets may be presented to replicate an enemy force counterattacking. These targets are engaged from a stationary position. Figure 1 depicts Marines conducting fire and movement in Camp Lejeune, North Carolina.

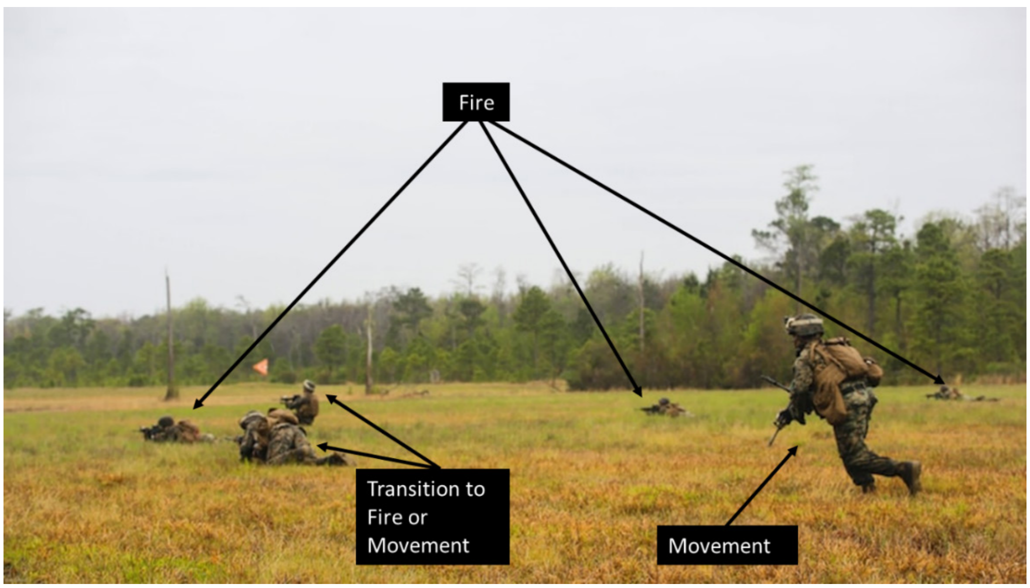


Figure 1. Marines Conducting a Squad Attack Using Fire and Movement. Source: marines.mil (2016).

At the lowest level, fire and movement is accomplished with two individuals, so for the squad attack consider six pairs of individuals each conducting fire and movement while maintaining alignment with the other pairs around them. The environment created during a squad attack, as shown in Figure 1, becomes complex even during routine training. Therefore, it is of interest to determine individual accuracy within this environment that more closely resembles combat.

As part of the GCEITF experiment, a squad attack scenario was created where Marines completed fire and movement at the squad level at Range 107 aboard Marine Corps Air Ground Combat Center (MCAGCC), Twentynine Palms, California. All squad attack trials consisted of 12 Marines and 18 targets, with each target simulating an individual enemy fighter, as shown in Figure 2 (MCOTEA 2015a).

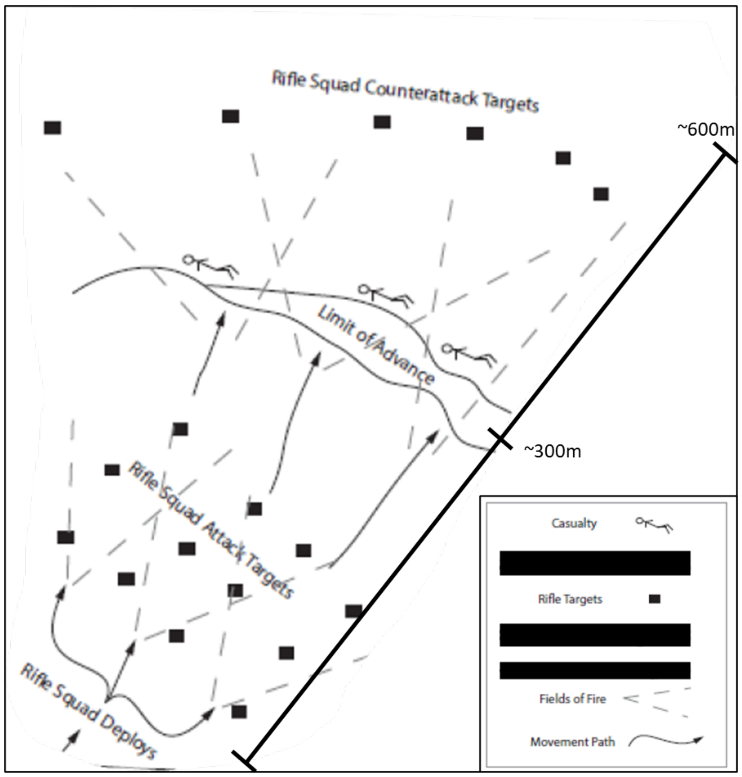


Figure 2. Range 107 Diagram. Adapted from MCOTEA (2015a).

A unique aspect of the GCEITF experiment is that all Marines had a Weapons Player Pack (WPP) fitted to their rifles so that each time the rifle was fired a measure was recorded that documented the location and time of firing. Additionally, all targets had sensors that recorded hits, near misses, and time of impact. A near miss is defined as a round passing within one meter of the target. By comparing the data from the WPPs and the targets, it is possible to link target impacts to the shooter who fired the round, and then collect statistics on the results. All data from the targets will be referred to as impact data, while all data from the WPP will be referred to as shot data.

The first attempt at linking shots to impacts was done with a time-based metric on 141,943 shots recorded by WPPs and 167,813 impacts recorded by targets. While location data was available that could provide distance once an impact had been assigned to a shot, there was no bearing recorded from the WPP at the time of firing. This meant that shots could not be matched to impacts geospatially, but rather based on time alone. All observations were recorded with 1/10 second precision, which meant multiple shots and impacts were being recorded as the exact same time, and the variance associated with ballistics prevents precise time-of-flight measurements. This method proved sensitive to small changes in the time-of-flight interval and therefore the results will not be presented.

Next, consolidated data consisting of 29,544 impacts linked to shots by analysts at MCOTEA was examined. This data had a bearing associated with the shooter and thus is considered valid data. This data still required some preparation before it was ready to be analyzed. First, there was no initial loadout recording how many rounds of ammunition the Marine began the range with and no recorded amount of ammunition turned in upon completion of the range. Ammunition data was available in separate files recorded for each day of the range. Since there were 82 squad attack trials completed (counting the control group and the integrated group as separate trials), there were approximately 164 individual CSVs that had to be referenced to get initial loadout and rounds turned in by each individual. Next, the original shot data containing all WPP observations was checked to find all shots recorded by the WPP that did not have a corresponding impact. By subtracting rounds remaining at the end of the range from the initial loadout, and checking the number of WPP observations, there were now two metrics to determine how many rounds an individual fired on the range. Since there are potential data recording issues with the rounds being given out or turned in, and potential malfunction of the WPP, it is useful to find instances where the two metrics differ. An issued magazine holds 30 rounds of ammunition, and it is common practice to give or receive a magazine of ammunition while conducting a squad attack range. Therefore, all instances where the difference between the two total shot metrics differed by more than 30 were discarded. Among the 649 individual observations, 24 were discarded for an inability to determine total rounds fired.

2. Marksmanship Results

a. Individual 0311 Rifleman Accuracy Across all Data

The first marksmanship metric of interest is individual mean accuracy in the attack. Hits are defined as rounds that impact the target, so the hit percentage is calculated as the number of hits divided by total rounds fired. Near misses are defined as rounds that impact within one meter of the target. The summation of hits and near misses is referred to as effective fire because hits incapacitate the enemy while near misses suppress the enemy. Thus, the proportions of effective fire are the number of hits plus the number of near misses divided by the total number of rounds fired. Figure 3 shows individual accuracy across all squad attack trials. In total, 34 individuals who participated in a squad attack trial from 1 to 14 times are represented in the 0311 Rifleman accuracy data.

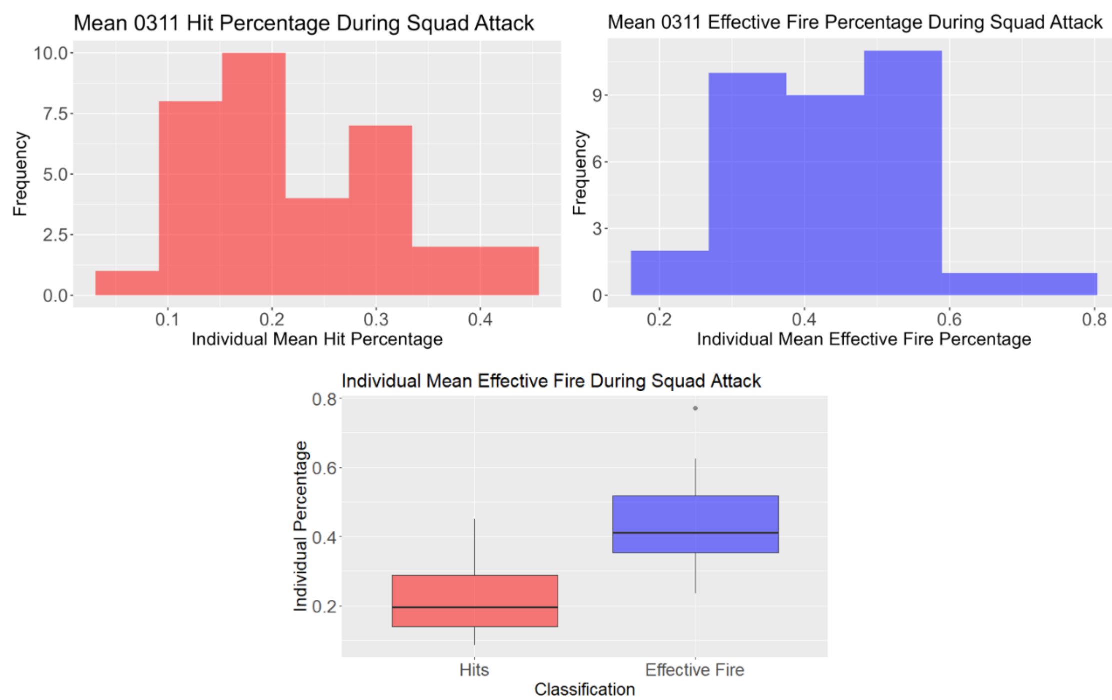


Figure 3. Individual Accuracy

Recall that the passing standard for the individual accuracy tasks in the T&R Manual is by achieving 70% accuracy or higher. Those tasks attempt to replicate realistic conditions by having the Marine engage from different shooting positions while wearing all gear. However, Figure 3 shows that in an attack scenario, the single best average effective fire percentage was 77%, while the single best average hit percentage was 45%. The interpretation is not that the T&R task standard is set unrealistically high, but rather, that the T&R standard simply does not translate to realistic scenarios. Those tasks must be accomplished with the goal of bettering marksman abilities but should not be confused for an individual's performance in a more chaotic environment. A realistic expectation is that the average entry-level Marine will hit a target with 22% of rounds expended, and effectively fire 43% of rounds expended.

b. Individual Accuracy vs. Distance

With the expected value of effective fire being 43% in mind, it is important to see the effect of distance on the proportion of hits within the effective fire sample. Figure 4 shows all effective fire recorded during the attack portion of the squad attack trials, meaning that the rounds were shot while executing fire and movement.

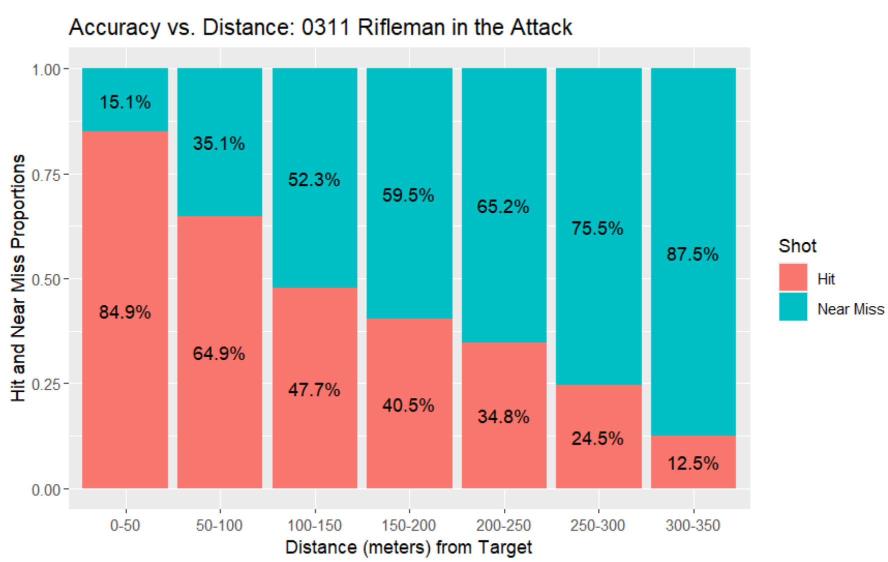


Figure 4. Accuracy vs. Distance

It is expected that as distance from the target increases, accuracy decreases. However, quantifying the rate of decrease, specifically in the attack, is of critical importance to the infantry leader. For attacks above a squad size, it is common to have a separate element suppressing the enemy to allow the maneuver element to get within 100 meters of the enemy position. Figure 4 shows that at 100 meters, the distance where fire and movement is the only method of suppression, 65% of effectively fired rounds are impacts. With further research, it is possible to link shot observations that do not have an impact to a distance category based on location of the shooter at time of firing. This would provide the denominator needed to know the proportion of hits by total rounds fired at this crucial distance.

c. *Effects of Rifle-Mounted Grenade Launcher on Accuracy*

At the time of the data collection, the only means of employing the fire team organic 40mm grenade launcher was rifle mounted. Today, there is an option to employ the 40mm grenade launcher as a stand-alone weapon system. Opponents of the rifle-mounted employment option often state that the rifle becomes unwieldy, and as a result, accuracy suffers. Figure 5 separates shooters who had employed the M4 rifle with and without the M203 40mm grenade launcher and plots hits and effective fire.

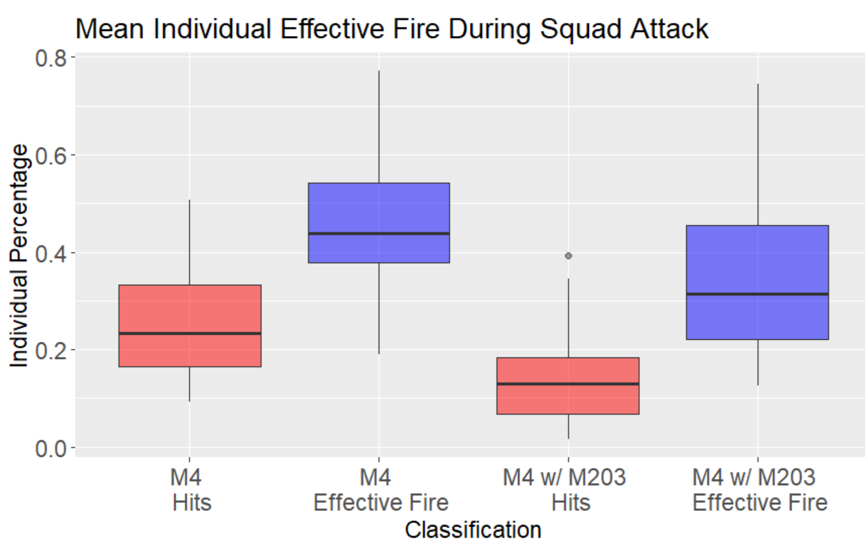


Figure 5. M4 vs. M4 with M203 Accuracy

The data show that rifle shots from M4s equipped with M203s are, on average, less accurate than shots from M4s alone. The proportion of effective fire for the M4 group is normally distributed, evidenced by a Shapiro-Wilk Normality test with a p-value of 0.64 (Shapiro et al. 1965). However, the proportion of effective fire for the M4 with M203 group is not normally distributed, with a Shapiro-Wilk Normality test p-value of 0.02 (Shapiro et al. 1965). Both groups, however, have a distribution that is reasonably symmetric about the median. Therefore, a Wilcoxon Signed Rank Test can be used to test the null hypothesis that the mean difference between the M4 and the M4 with M203 groups equals 0, against the alternative hypothesis that the mean difference between the M4 and M4 with M203 groups is less than zero (Wilcoxon 1945). The p-value for this test is 0.01, with a 95% upper confidence bound of -0.05, thus rejecting the null hypothesis and concluding that equipping an M4 with an M203 results in a lower proportion of effective fire. For context, a Marine with an M203 should be focused on employing the M203, and thus the rifle shots recorded by Marines with M203s are likely more rushed than Marines employing M4s alone. While the difference between rifle fire accuracy is statistically significant, it may not be practically important if the M203 is being usefully employed. Each trial run had three Marines (out of 12 in the squad) armed with M203s, and each of those three Marine had six 40mm grenades issued at the start of the trial run. Of those 18 rounds, the mean number of effectively fired grenades is 6.5, with a standard deviation of 3.4. This implies that M203 gunners were not solely focused on employing the M203 and leads to the practical conclusion that rifle accuracy suffers as a result of mounting the M203 grenade launcher on the rifle.

d. 0311 Rifleman vs. Provisional Infantry Accuracy

One concern of the GCEITF was that possible differences found between male 0311 Rifleman and female 0311 Rifleman could be attributed to experience (MCOTEA 2015a). The female 0311s had just recently completed Rifleman training at the School of Infantry. To ensure this effect was not impacting the results, male Provisional Infantry (PI) Marines also participated in the experiment. These PI Marines were not 0311 Rifleman, and thus were equally or more inexperienced than the female 0311s. It is of interest to test the difference in accuracy between all 0311 Rifleman and all PI Marines. Figure 6 shows the boxplots of hits and effective fire for 0311 Rifleman and PI Marines.

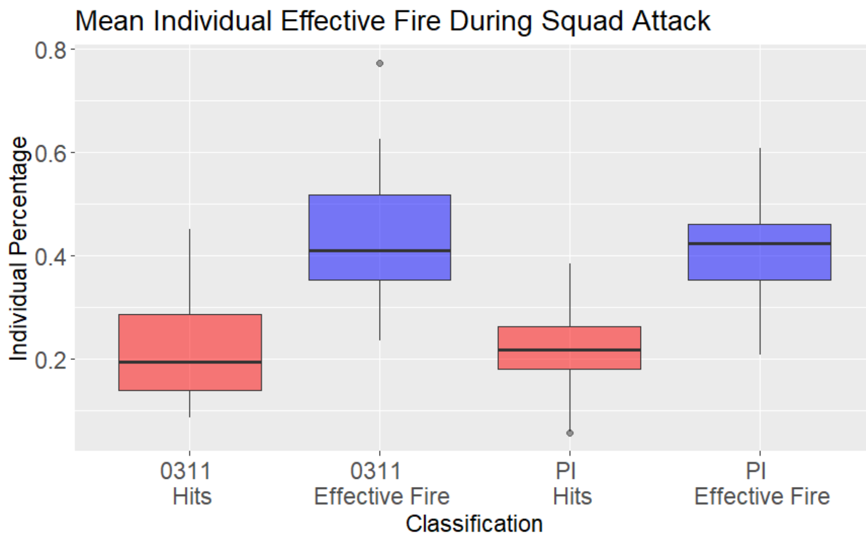


Figure 6. Rifleman vs. Provisional Infantry

At first look, it appears that the PI Marines have a slightly higher median accuracy than the 0311 Rifleman. There are 34 individual mean accuracy observations for the 0311 Rifleman, and 37 individual mean accuracy observations for the PI Marines. A sample of size 30 was taken from each group, and a Shapiro-Wilk normality test confirmed that both groups are approximately normally distributed (Shapiro et al. 1965). An F-test confirmed that there is no significant difference between the sample variances. With these assumptions met, an unpaired two-sample t-test was conducted with the null hypothesis being the mean difference between the two groups is equal to zero, and the alternative hypothesis being the mean difference between the two groups is not equal to zero. For hits, the p-value of the test is 0.62 with a 95% confidence interval on the difference in mean of (-0.03, 0.05). For effective fire the p-value is 0.38, with a 95% confidence interval on the difference in mean of (-0.03, 0.08). In both cases, there is a failure to reject the null hypothesis at the 0.05 significance level, thus concluding that there is no statistically significant difference between the two groups.

3. 0311 Accuracy Concluding Remarks

Infantry leaders should recognize that accuracy in the attack is likely lower than expected and place an increased focus on marksmanship during live fire and maneuver. Additionally, the use of targets that record impacts, such as Marathon targets, offers an opportunity for further analysis. The data from the targets is available at the conclusion of live fire training but is not formally utilized. There is value in quantitatively describing the effects of fire in different scenarios, and there should be a repository with access to the raw target data and a summary provided to units at the conclusion of training.

The first recommendation for future work includes taking the matched shot data provided by MCOTEA and comparing it to the separate impact and shot data using the time interval approach described. There may be an identifiable trend that shows accurate matches using the time interval method, which would allow for more detailed analysis. Specifically, if there was a recorded shot and impact linked with a recorded distance, a time delta close to that time of impact could be selected to find shots fired that did not impact. It could be surmised that these shots were fired at approximately the same distance as the impact, verified by the status of the targets at the given time. This would give an accurate denominator for proportions of effective fire out of total rounds fired at varying distances.

Recent changes to the Marine Corps Annual Rifle Qualification credit lethal hits, as opposed to hits anywhere on the target. Since coordinate data for impacts are recorded, further analysis could further classify accuracy categorized by lethal hits, non-lethal hits, and near misses. By pairing this data with the GPS data, movement times can be compared to times of effective or ineffective fire. This metric could provide the level of detail required to determine the proficiency of Marines conducting fire and movement.

C. 0331 MACHINE GUNNER

1. Offensive Medium Machine Gun Support

In offensive situations, medium machine guns are typically employed as a supporting asset for the riflemen in the maneuver force. The role of the machine gun unit is to provide suppressive fires from a Support by Fire (SBF) position that allows the maneuver force to close the distance to the enemy position with minimal enemy resistance.

A general sequence for offensive machine gun support is movement to the initial SBF position, engaging targets, and then moving to an alternate SBF position to prepare for an enemy counterattack (DON 2019). This sequence, and every stage within, are subject to the tactical situation and change frequently. Therefore, there is no established T&R task to objectively measure proficiency for offensive machine gun support. T&R task 0331-EMPL-2004 “Direct the Employment of a Machinegun Unit in the Offense” has the standard “to provide effective machinegun fires in support of the scheme of maneuver.” The lack of specificity within this task and standard cannot be overcome through a rewrite; however, it can still benefit from adding a quantitative component. The GCEITF use of offensive machine gun support did not directly support the riflemen maneuver force, but rather shot separate targets. Therefore, the art of machine gun employment, such as the coordination of shifting and ceasing fires based on the position of the maneuver element, cannot be assessed. However, measures of the science of machine gun employment can be discovered by referencing the sequence used in the GCEITF attack and shown in Figure 7. With a quantitative baseline established for the physical and technical aspects of offensive machine gun support, evaluators can better focus on the artful employment of the machine gun unit.

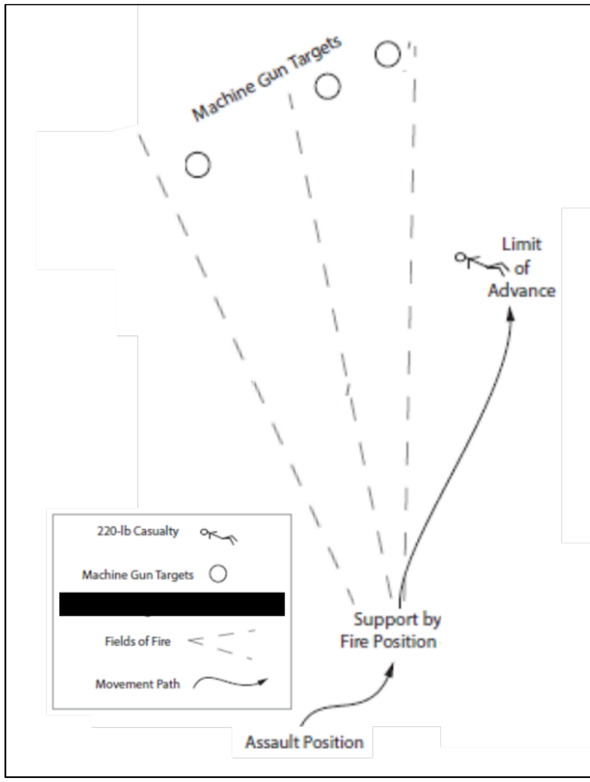


Figure 7. Offensive Machine Gun Scheme of Maneuver. Adapted from MCOTEA (2015a).

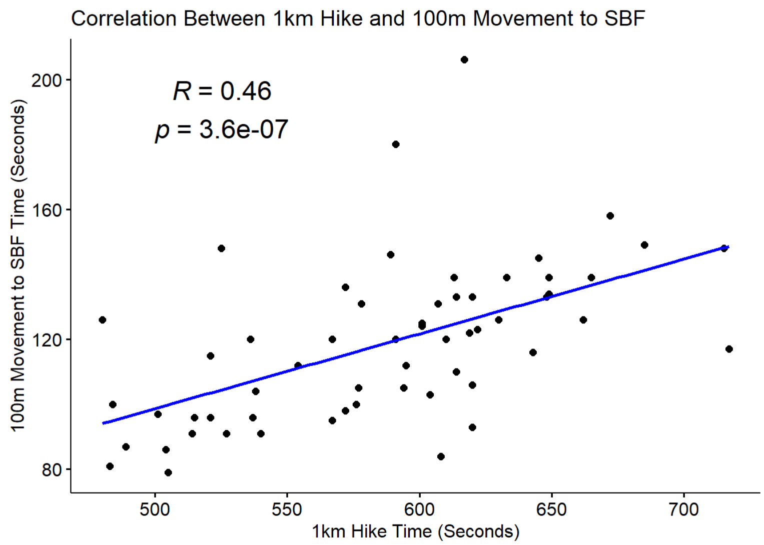
During the GCEITF, a single machine gun team conducted an approximately one-kilometer movement to an assault position, a 100-meter movement from the assault position to the SBF position, emplaced the M240B on a tripod, and conducted a 400-round course of fire. Upon completion of that engagement, the team then displaced 300 meters to a second SBF position and concluded with a 200-round course of fire and a casualty evacuation (CASEVAC) of a 220-lb dummy (MCOTEA 2015a). Summary statistics of task completion are shown in Table 4 for 58 trial runs.

Table 4. Mean Completion Times (MM:SS)

Metric/ Statistic	~1km Hike	100m Movement to SBF	Mount and Load M240B	Engage Targets (400 Rounds)	Displace 300m to LOA	Mount M240B and Engage Targets	Total
Mean	09:47	01:59	00:57	06:14	04:21	01:51	25:08
Standard Deviation	00:59	00:25	00:20	01:32	00:44	00:57	04:57
Minimum	08:00	01:19	00:05	03:09	02:41	00:05	15:19
Maximum	11:57	03:26	02:06	09:00	06:22	05:00	37:51

There was little variation in the routes taken or the positions used for all trial runs. However, there were opportunities for each machine gun team to perform the task in a different manner. For example, the minimum time to mount and load the M240B at the initial SBF position and the secondary SBF position are both five, seconds as shown in Table 4. The time for mounting and loading the M240B is calculated as the time from the tripod being emplaced to the time that "Gun Up" was announced. In most situations, the tripod would have been emplaced before the Marine carrying the M240B arrived to the SBF position to mount and load the weapon. However, in some cases, the weapon arrived before the tripod. Therefore, the weapon could be loaded and waiting for the arrival of the tripod, which makes the "mount and load time" as low as five seconds.

Being a machine gunner requires superior physical fitness to complete the mission from start to finish. Thus, it is worthwhile to examine all physical events associated with the GCEITF offensive support trial runs not as independent events but rather successively. If all teams had similar levels of physical fitness, then there should not be a correlation between individual movements. For example, if a team completed the one-kilometer hike quickly, then it should be expected that subsequent movements were done slowly as an effect of fatigue. However, this is not the case. Figure 8 shows the correlation between the one-kilometer hike and the 100-meter movement from the assault position to the SBF position, as well as the correlation between the one-kilometer hike and the 300-meter movement from the SBF position to the LOA.



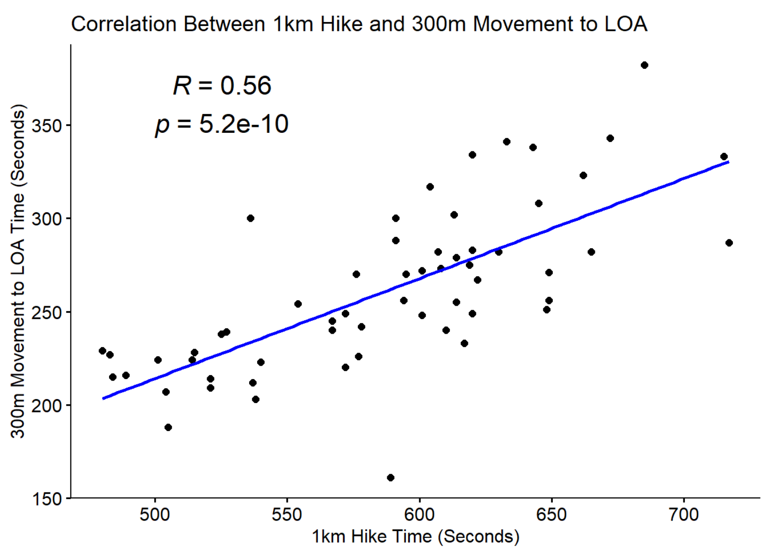


Figure 8. Correlation Between Offensive Machine Gun Movement Times

In both cases, it is readily apparent that movement times are positively correlated. Thus, the practical conclusion is that faster hike times means a higher level of physical fitness within that machine gun unit. Since the data are not normally distributed, Kendall's rank correlation test is used to determine the correlation coefficient (Smeeton et al. 2007, ch. 10). The null hypothesis for this test is that the true tau is equal to zero, meaning the data are uncorrelated, against the alternative hypothesis that the data are correlated. The low p-values for both instances presented in Figure 8 lead to a rejection of the null hypothesis and a conclusion that the movement times are correlated. The implication of this is very important: physical fitness is paramount to the speed at which a machine gun unit can operate. This is a simple observation, but the implications for unit effectiveness are important. First, this effect is not due to additional recovery time or rest time. The time between concluding the one-kilometer hike to the assault position, and leaving the assault position towards the SBF position, was less than one minute for every group. Additionally, time spent firing at the assault position was actually lower for those groups that hiked faster, as shown in Figure 9.

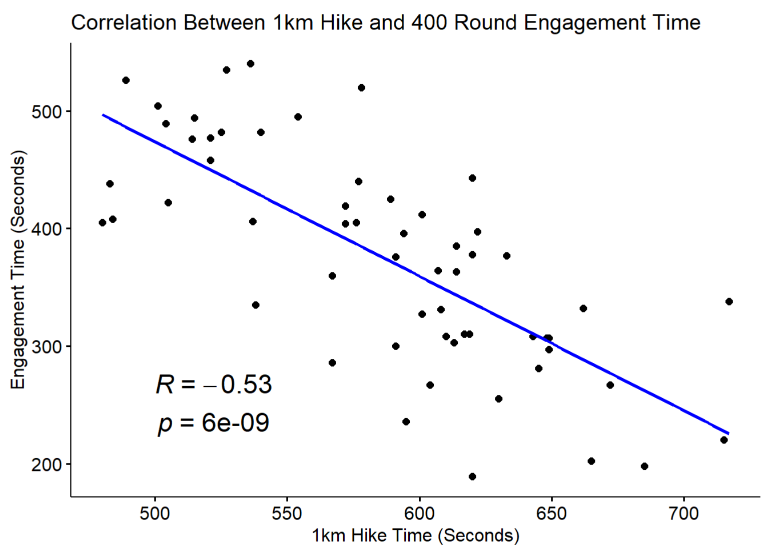


Figure 9. Correlation Between Hike and Engagement Times

There is no readily available practical explanation for this. The concern that the groups moving and firing quickly were not as effective, meaning not as accurate, as those groups with slower movement and engagement times is not supported either. Figure 10 shows that there is not a statistically significant effect on the number of impacts recorded during each engagement and the length of the engagement.

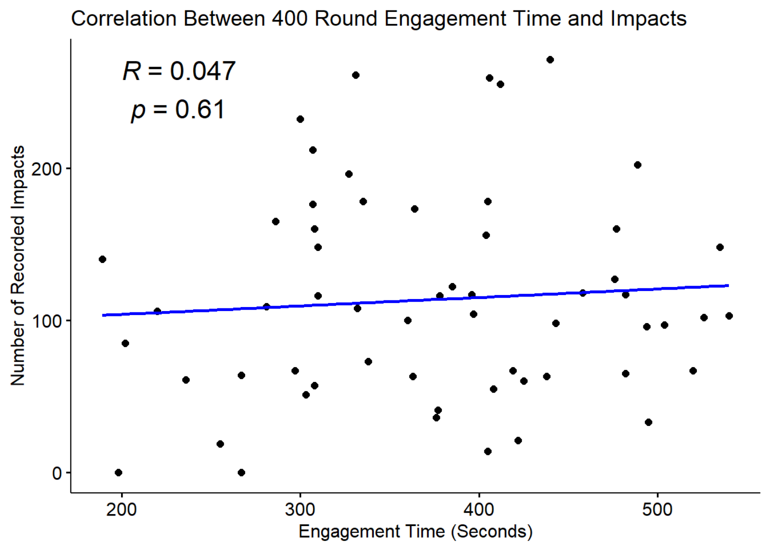


Figure 10. Correlation between Engagement Time and Impacts

It is important to mention that expenditure data is not readily available, which means all engagements may not have fired the same number of rounds. That is why the metric used is number of impacts, instead of a proportion of impacts to shots fired. Given that there is no correlation between the engagement time and number of recorded impacts, it is likely that the variance of number of rounds fired is low across all trials. Time of day considerations were also checked, and there is no relation between movement times or accuracy based on time of day. Therefore, there is no concern that morning trials were faster than afternoon trials due to cooler weather.

The next important implication is that physical fitness does not directly transfer to the technical skills required of machine gunners. This seems counterintuitive based on Figure 9 showing a negative correlation between physical fitness and length of engagement times, and Figure 10 showing that there is no correlation between the number of impacts and length of engagement times. The conclusion is that the Marines who were more physically fit engaged the targets faster without a decline in accuracy, thus implying a greater technical proficiency. However, there is no correlation between physical fitness and the "Gun Up" time as shown in Figure 11. "Gun Up" refers to the process of preparing the M240B to be fired and consists of emplacing a tripod, attaching a pintle mount to the

M240B and placing the pintle mount into the tripod, and attaching a Traverse & Elevation mechanism to the M240B and placing the mechanism onto the tripod.

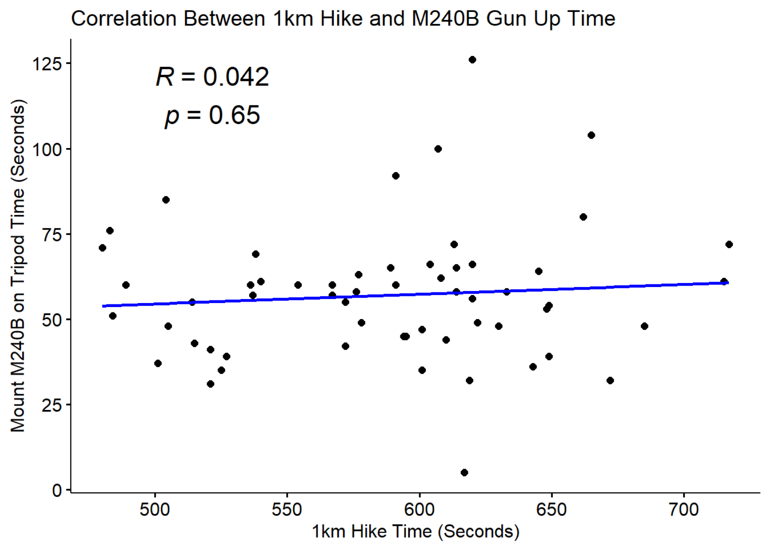


Figure 11. Correlation between 1km Hike and M240B Gun Up Time

Given the lack of correlation shown in Figure 11, it is reasonable to conclude that physical fitness and technical requirements of machine gunners are not directly related. The conclusion for the machine gun unit leader is that physical fitness is paramount to providing timely offensive machine gun support. Additionally, technical skills must be drilled thoroughly to prevent a gap in providing timely support.

2. Heavy Machine Guns

a. Preparing the M2 .50 Caliber Machine Gun for Employment

The analysis presented for offensive machine gun support is focused largely on the physicality involved with being a machine gunner. The theme of physicality is continued when considering heavy machine guns, as movement with the weapon system is limited due to the weapon's size and weight. The M2 .50 caliber machine gun, which is the weapon system discussed throughout this section, weighs 128 pounds when assembled on a tripod, and 100 rounds of .50 caliber ammunition weighs 35 pounds (DON 1996). One of the tasks completed in the GCEITF was mounting the M2 .50 caliber machine gun on a vehicle, which is a T&R task. The Task, Standard, and Event Components are shown in Table 5.

Table 5. Associated T&R for Heavy Machine Gun Mounting. Adapted from DON (2020).

Task	INF-MGUN-3002: Mount a heavy machine gun on a tactical vehicle.
Standard	To prepare for mounted operations.
Component 1	Inspect the heavy machine gun and associated components.
Component 2	Attach the mounting adapter to the tactical vehicle.
Component 3	Attach the cradle and mount to the mounting adapter.
Component 4	Attach the traverse and elevation mechanism to the heavy machine gun.
Component 5	Attach the ammunition brackets to the cradle and mount.
Component 6	Attach the heavy machine gun receiver to the cradle and mount.
Component 7	Attach the barrel to the heavy machine gun receiver.

The standard for INF-MGUN-3002 has no quantitative component associated with it. In fairness, this is not a skill that will generally need to be completed in a time constrained environment. However, it is an easily measurable task that can show the basic technical proficiency of a heavy machine gun vehicle crew. Figure 12 shows the distribution of 60 timed trials completed during the GCEITF, during which a team of three Marines moved an M2 .50 caliber machine gun from the ground to a vehicle (mount) and from the vehicle to the ground (dismount). Of note, all trials required a headspace and timing check to be completed, which is no longer required for the M2A1 .50 caliber machine gun in use today. Therefore, the times presented in Figure 12 are likely slower than would be expected with the M2A1.

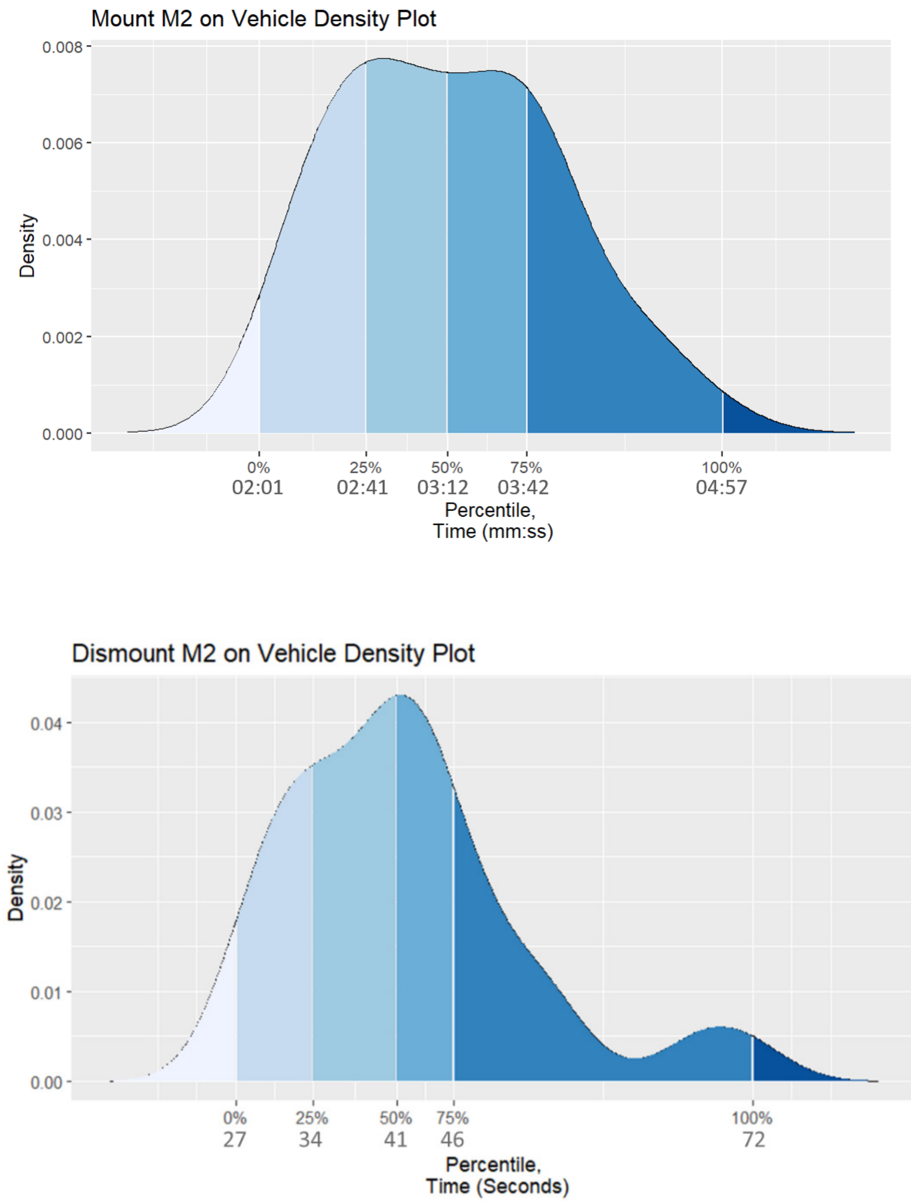


Figure 12. Mount and Dismount M2 on Vehicle Density Plots

Despite slight differences in the M2 and M2A1, which would produce slightly different trial times, there is still immense value to the infantry leader from a plot like Figure 12. In previous chapters, the range of experience among unit leaders was discussed. Leaders at the Platoon Sergeant and higher level may not have ever been in a unit that employs vehicle-mounted heavy machine guns. This task is easy to complete while in a garrison setting, and the results compared to a plot like that of Figure 12 can provide those unit leaders an immediate sense of the technical proficiency of all their vehicle crews. Then, this task could be completed after several days in a field setting, or at night with the use of night-vision-devices, and the results could be compared to the garrison baseline. Through low-cost and low-effort data collection, and the appropriate summary statistics returned from that data, the infantry force can make critical insights that have otherwise gone unnoticed.

b. Employing the M2 .50 Caliber Machine Gun

During the GCEITF, a crew of three Marines fired the M2 .50 Caliber Machine Gun mounted on a tripod on the ground. The T&R Manual has task 0331-M2-1001 "Engage Targets with a M2A1 Heavy Machinegun" with the standard "to engage target(s) in accordance with fire command(s)" (DON 2020, p. 13-9). Accuracy with the M2 was not a metric used by the GCEITF, but accuracy information was recorded and discussed as 0331 Machine Gunner supplemental information (MCOTEA 2015a). As with all analysis in the Experimental Assessment Report, accuracy was considered based on the level of gender integration within machine gun crews. In this section, total accuracy is considered independent of gender integration. For the course of fire, there were three targets at ranges of 450m, 600m, and 800m. The maximum effective firing range of the M2 is 1,830m, implying that the GCEITF firing ranges were between 25% and 44% of the weapon's capability (DON 1996). Accuracy data was stored in 114 separate Excel files, and after reading those files into Python and cleaning the data, there were 113 usable trials. One trial was dropped due to an erroneous data entry. Figure 13 shows the distribution of hits for each target from the 113 trials.

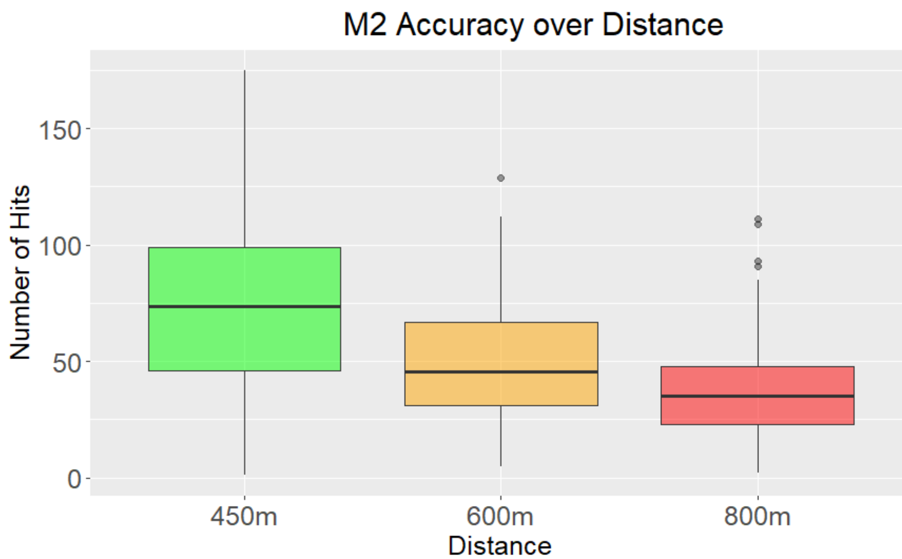


Figure 13. M2 Hits over Distance

Details for the 400 round course of fire were not provided, and it cannot be assumed that an equal number of rounds were fired at each target. This means an accuracy number in the form of a proportion cannot be calculated. However, there is data that records when each target was visible or concealed to the Marines on the firing line. By calculating the total time each target was visible, known as the exposure time, then one can intuit the amount of opportunity the Marines had to engage each target. Figure 14 shows the distributions of exposure time for each target.

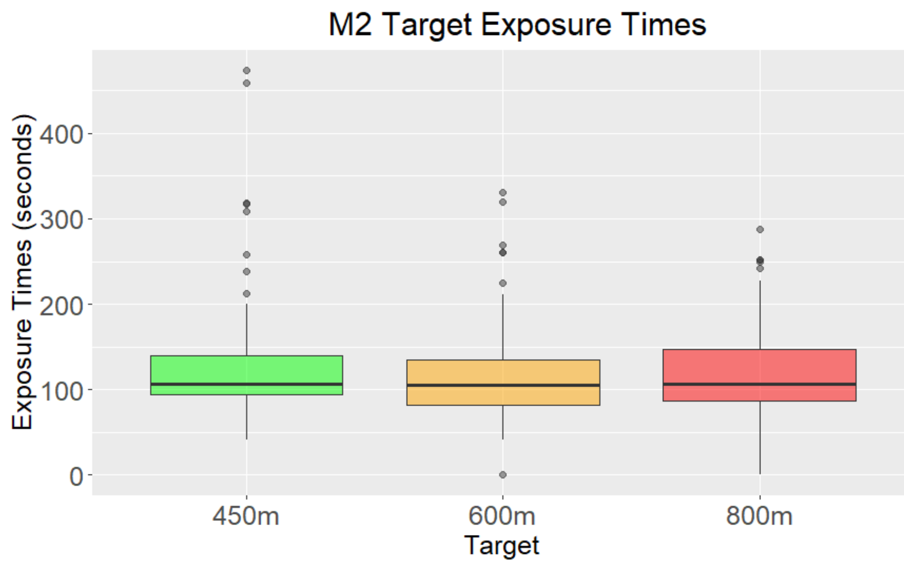


Figure 14. M2 Target Exposure Times

Figure 14 reveals that the distributions of exposure times are similar for the three targets. An ANOVA test, with the null hypothesis that the mean of each distribution is the same, and alternative hypothesis that at least one mean is different, results in a p-value of 0.45. Therefore, the null hypothesis is not rejected, and we thus conclude that there is no evidence that suggests the mean target exposure times are different. Therefore, the decreasing number of hits shown in Figure 13 is suggestive of a decrease in accuracy with the M2 as the distance to the target increased.

3. 0331 Machine Gunner Concluding Remarks

For further analysis on offensive machine gun support, the force needs to collect data for live fire that requires close coordination between the maneuver element and the machine gun unit providing support. Assuming operating forces do not have access to equipment which records weapon shot data but do have access to targets which record impact data, the following metrics could quantitively measure the effectiveness of machine gun support. First, times between target impacts or near misses can be measured that indicate gaps in suppression. The sound of machine gun fire does not suppress the enemy, but effectively fired rounds do. Second, times of overlapping fire between the maneuver element and the machine gun unit providing support can be collected by evaluators. By filtering the target impact data beginning with the first rifle round fired and ending with the last machine gun burst fired, the total time and impacts recorded during that time interval can quantitatively describe the execution of the handover of fires.

With evaluators tracking the times of the commands given by the machine gun unit leader, the by-gun effectiveness of a squad or section can also be quantified. For example, a machine gun squad consisting of two M240Bs can be tasked with "talking guns" at the rapid rate of fire, where the first gun fires a burst, then after a two-to-three second pause the second gun fires a burst, both aiming at the same target. By collecting the number of impacts over time, a line plot can show by-gun effectiveness. Figure 15 was created using notional data of the number of recorded impacts for 20 bursts fired with "talking guns" to show the by-gun contribution.

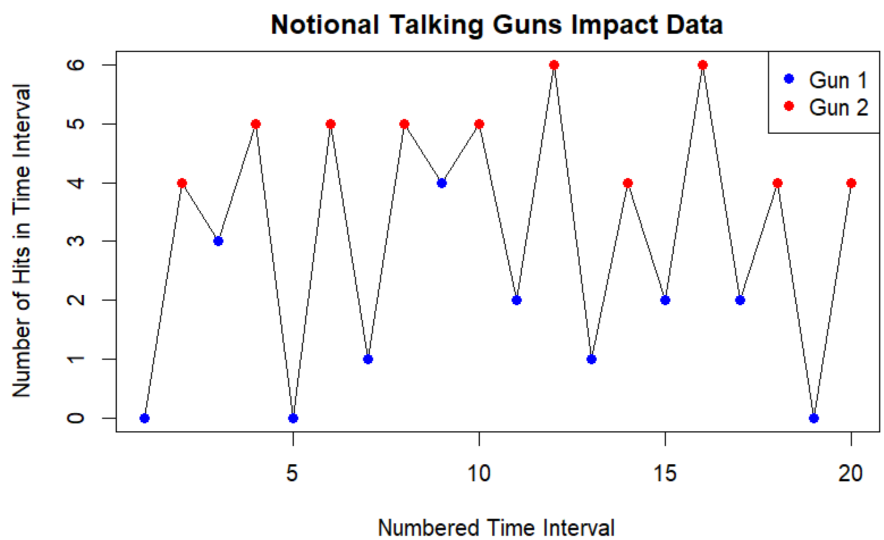


Figure 15. Notional Medium Machine Gun By-Gun Effectiveness

Observe that Gun 2 had significantly more impacts over the one-minute period of firing than did Gun 1. Leveraging a plot like Figure 15 for use as a debrief aid provides objective and quantitative feedback far beyond that of current methods. Other technical skills including "Gun Up" time, disassembly and assembly of all machine gun platforms, and manipulation of mounted machine guns can all be easily measured and recorded in various settings. Uploading this data and comparing it to other Marines across the force is a low effort means to truly understand unit technical proficiency.

D. 0341 MORTARMAN

Mortars are invaluable assets within an infantry unit as they have historically provided the only organic indirect fire capability. Emerging technologies are changing this precedent, offering additional capabilities for beyond line-of-sight targeting. However, mortars will continue to remain a key weapon within infantry units. At the company level, there are three 60mm mortars within the mortar section. A single 60mm mortar is unique in that it can be employed in handheld mode, meaning the weapon can be rapidly moved and employed alongside maneuver forces, as shown in Figure 16.

Figure 16. Marines Employ the 60mm Mortar in Handheld Mode. Source: MCOTEA (2015).

Weighing in at just 18 pounds, and with each round weighing four pounds, a team of two Marines can efficiently provide indirect fires while maneuvering with the 60mm mortar (DON 1992). The GCEITF employed 60mm mortars in handheld mode, with two teams of two Marines. Each team moved to an initial Mortar Firing Position (MFP) and fired three rounds, then displaced to a second MFP, and fired another three rounds, as shown in Figure 17.

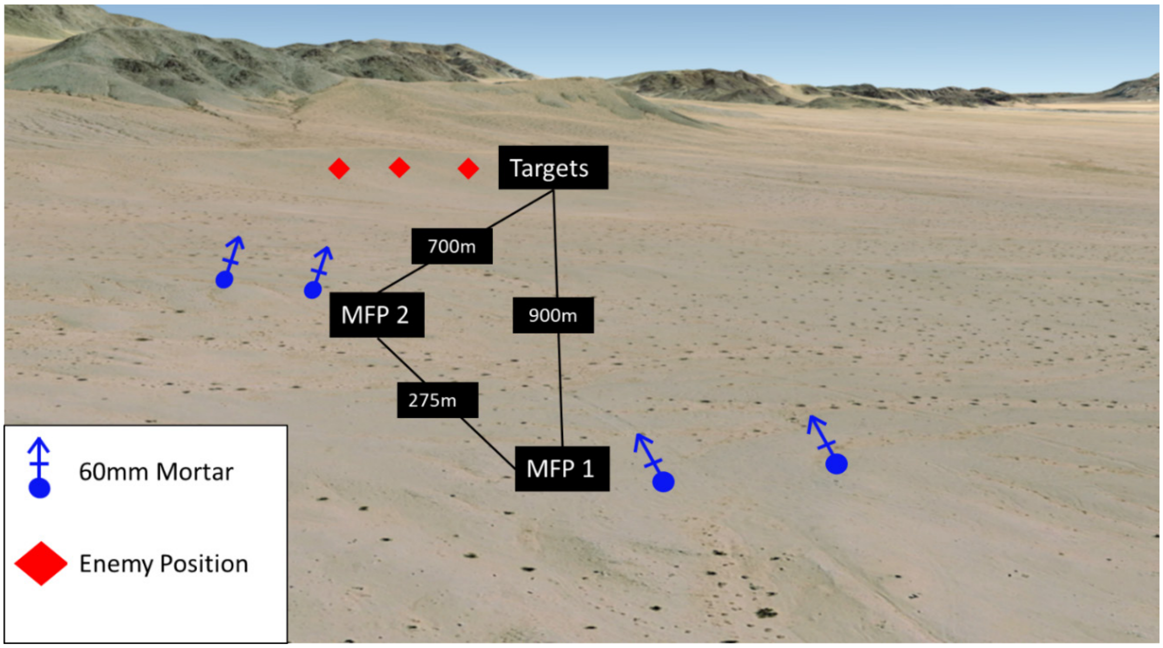


Figure 17. Mortar GCEITF Scheme of Maneuver

During the GCEITF attacks, the impact of each round was observed by VECTOR 21 binoculars connected to a Defense Advanced Global-Positioning-Receiver Receiver (DAGR) to record the exact location of impact. Despite recording the data, the GCEITF study did not draw any conclusions from the recorded impacts. Indirect fire is highly variable to conditions outside the controls of an experiment like the GCEITF study. It is known that employing the 60mm mortar in handheld mode is less accurate than employing the weapon in conventional mode. However, the true effect of handheld employment has not been quantified. Each trial fired 12 rounds, with three trials per day for 21 days, totaling 756 rounds fired. After data cleaning, there are 440 valid impacts from the GCEITF trials, shown in black in Figure 18.

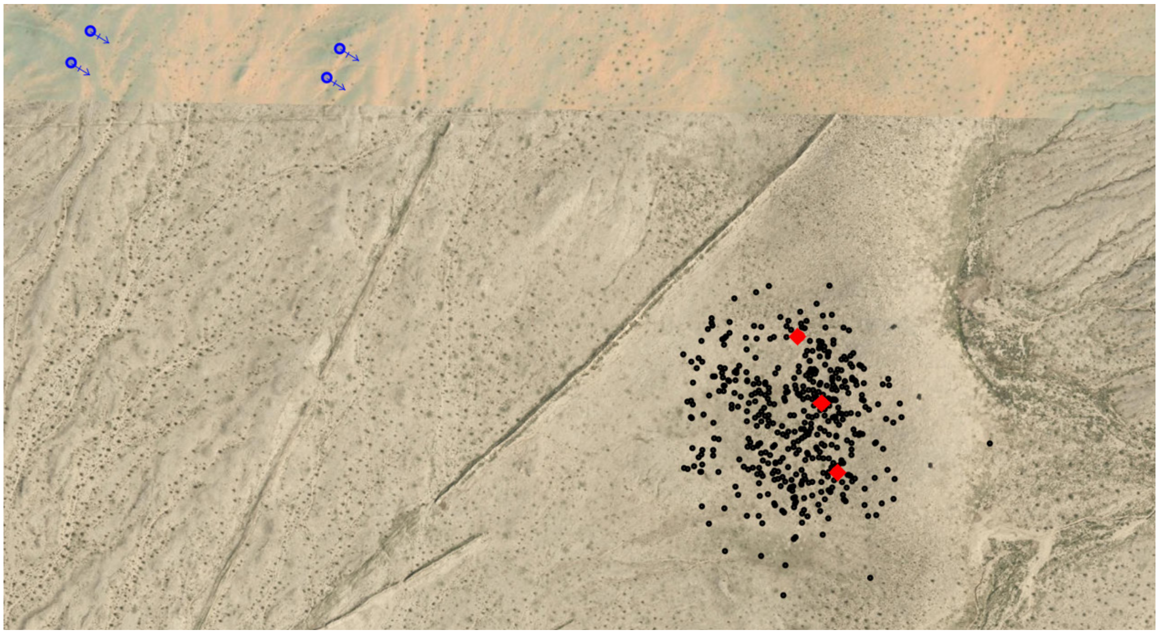


Figure 18. Mortar Impacts

The 60mm mortar, as compared to the 81mm mortar or 155mm artillery, is a light form of indirect fire. While employed in handheld mode, it is likely the commander seeks to suppress the enemy force. MCWP 3-15.2, *Tactical Employment of Mortars*, defines mortar suppression as "an effect produced in the mind of the enemy that prevents him from returning fire or carrying on his duties" (DON 1992, p. 237). A more thorough definition of mortar suppression states "if a 60-mm mortar round lands within 35 meters of a target, there is a 50 percent chance it will be suppressed. Beyond 50 meters, little suppression takes place" (DON 1992, p. 238). Since there is no way to determine which of the three targets each mortar was fired at, the distance measure for each impact was optimistically computed as the distance from impact to the nearest target and computed as the average distance from impact to each of the three targets. Under the optimistic minimum distance case, 103 of the 440 rounds landed within 35m of the target, which means approximately 25% of rounds fired achieve a doctrinal 50% chance of suppression. A mortar squad consisting of three Marines and one 60mm mortar, in the attack and employing the mortar in handheld mode, can likely carry 8 rounds per person. Based on the GCEITF data and the MCWP 3-15.2 doctrinal standard, and assuming a squad fires all rounds, this implies that a mortar squad will only fire 6 rounds effectively. However, the MCWP 3-15.2

estimate of suppression distance seems too conservative. In the same section as previously quoted, MCWP 3-15.2 states “if an 81-mm mortar round lands within 75 meters of a target, there is a 50 percent chance that the target will be suppressed” (DON 1992, p.238). The initial suppressive effects of mortars landing in proximity to an enemy position is not likely to change whether it is a 60mm mortar or an 81mm mortar. Therefore, assume that any mortar landing within 75m of an enemy position will have an initial suppressive effect. The GCEITF data shows that 260 out of the 440 rounds fired landed within 75m of the nearest target. By adopting a more realistic standard of suppression, 75m instead of 35m, the proportion of effectively fired rounds increases from 23% to 60%. Returning to the example of a mortar squad with 24 rounds, the expected number of effectively fired rounds increases from 6 to 14. For a firing distance reference, the maximum effective range for the 60mm mortar employed in the handheld mode is 1,340m (DA 2000). All rounds fired were approximately 750m or 900m from the targets, which equates to 55% or 70% of the maximum effective range. While there was no significant difference in accuracy between the rounds fired at each distance, it is likely that accuracy will increase as the firing distance decreases.

When using indirect fire, it is crucial to observe the initial impact and make a correction to get the next round closer to the target. Sometimes this is done with the use of an observer relaying information to the firing agency, but in the case of the GCEITF study, the Marines firing mortars could see the target and their impacts. Using the optimistic distance measurement of impact to nearest target, the rounds fired from each MFP were sorted and plotted, as shown in Figure 19.

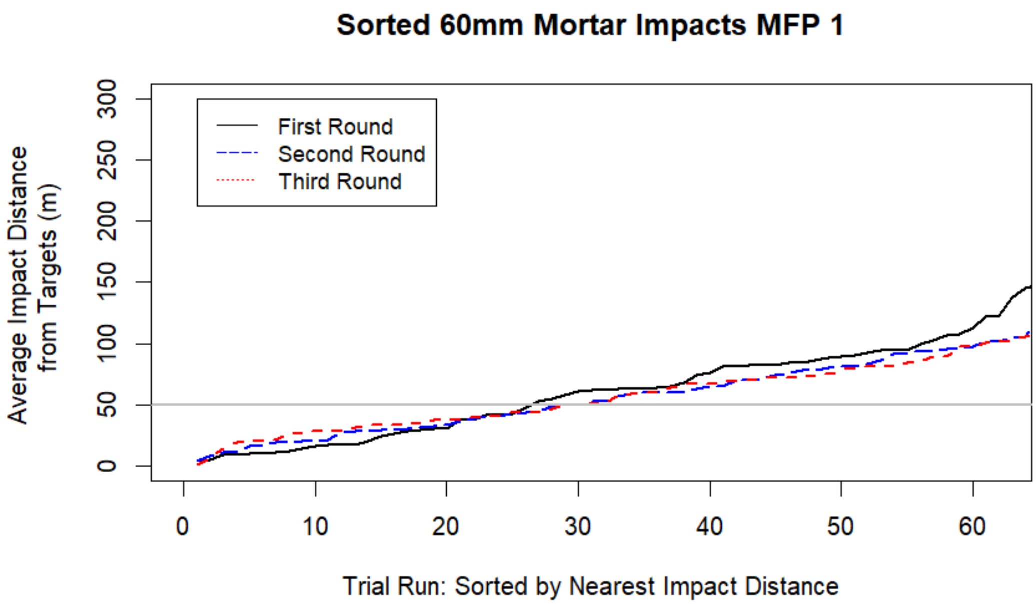


Figure 19. Mortars Fired from MFP 1

Note that the rounds plotted in Figure 19 are sorted from minimum impact distance to maximum impact distance for the first, second, and third rounds fired from MFP 1. The horizontal line at 50m is given to visually depict the distance past which corrections were likely applied. If the first impact was within 50m of the target, then the Marine firing will likely try to repeat that outcome. Figure 19 suggests that past 50m, the second round and third round impact closer to the target, but this must be verified using the three-round impact data from each trial run. To test if this difference statistically significant, the first three rounds from each mortar team were collected. After removing missing or invalid data entries there are 66 observations. Of those 66, there are 40 cases where the first round impacted 50m or farther from the target. Since the data are not normally distributed, a Wilcoxon Signed Rank Test was conducted with the null hypothesis being that the difference between the first round and second round impacts equals zero against the alternative hypothesis that the difference between the first round and second round impacts is greater than zero (Wilcoxon 1945). The p-value for this test is 0.002, thus the null hypothesis is rejected at any commonly used significance level. The practical conclusion is that effective corrections were made from the first to the second round. Using the same methodology for the second and third rounds from MFP 1 and repeating the procedure for the three rounds fired from MFP 2, all p-values were below 0.05, suggesting that by round corrections are effectively applied.

The T&R task relating to this section is 0341-WPNS-1005 "Fire a Mortar in Handheld Mode," with the standard "to achieve effects on target" (DON 2020, p. 14-11). There are six mortar rounds allotted per Marine to achieve this task. Achieving effects on target is too vague to be a useful metric when considering the range of Marines who will be evaluating this task. However, as previously noted, indirect fire is difficult to control and there are extraneous conditions that effect where the mortar round will land. The standard could be clarified with additional information, proposed as "evaluator notes." From the impact analysis, one note could read: "Effects on target as defined in MCWP 3-15.2 are rounds impacting within 35m of the target. It is expected that three of the six rounds will impact within 75m of the target, thus achieving suppressive effects." Another note pertaining to corrections may read "[If] the first round impacts more than 50m from the target, it is expected that the subsequent rounds will be closer. If subsequent rounds are not closer to the target, coach the Marine on proper corrections." By adding quantitative components, as in this example with the evaluator notes, based on actual data, the infantry force will train to the same objective standard and begin to see areas the unit excels in or is deficient in.

E. 0351 ASSAULTMAN

Infantry units require weaponry that can be used to engage covered positions, such as enemy bunkers and vehicles that may have armor protection. Rockets are the weapon of choice for such targets, and at the time of the GCEITF study the Shoulder-Launched Multipurpose Assault Weapon (SMAW) was the rocket assigned to 0351 Assaultman. The Marine Corps no longer has 0351 Assaultman and has transitioned from the SMAW to the Carl Gustaf Multi-Role Anti-Armor Anti-Personnel Weapon System (MAAWS) employed by 0311 riflemen. Despite the change with respect to the weapon and MOS employment, there is still value in examining the rocket data from the GCEITF to serve as a baseline metric for rocket employment. The MAAWS is a much more capable weapon system than the SMAW, so if any units employing the MAAWS have lower performance than the GCEITF employing the SMAW, that will be indicative of a gap in proficiency (U.S. Marine Corps 2021).

The GCEITF live-fire employed two teams of two Marines, each firing two rockets for a total of four rockets per trial. Each team engaged a target at 150m and then engaged a second target at 250m. Due to different techniques in firing the SMAW and the MAAWS, the time of engagement is not considered in this analysis. Accuracy, however, is a valid metric even when comparing the two different systems and is the focus of this analysis. When considering the squad (both teams) as a whole, there are 180 rocket shots recorded. For the first and second teams, respectively, there are 88 and 86 rocket shots recorded. This implies that six shots were not attributed to teams, but it cannot be determined which of those six shots should be attributed to each team, so the difference is not considered when comparing teams to squad performance. Figure 20 shows the team accuracy for all trials.

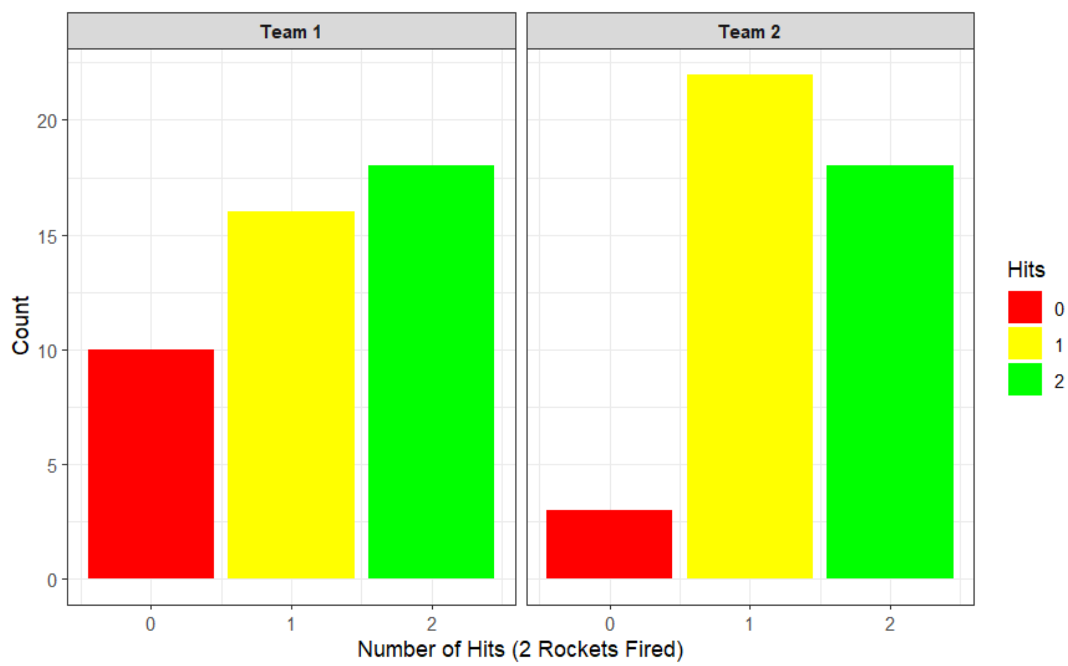


Figure 20. SMAW Accuracy by Fire Team

A Chi-Square test conducted with the null hypothesis being that there is no difference between Team 1 and Team 2 results against the alternative hypothesis that there is a statistically significant difference between the results yields a p-value of 0.004, thus concluding that there is a statistically significant difference. It is difficult to explain the difference in the distribution between the two teams. The selection process was done randomly to negate the difference in individual skill. The most important difference is that the first team had a significantly higher number of trials with no hits than did the second team. The most likely explanation is that the squad leader was announcing adjustments for the second team based on the impacts of the first team. This rationale is supported by the overall hit percentages of 59% for the first team and 67% for the second team. However, testing the equality of these proportions, under the null hypothesis that the probability of hit is the same for the first team and second team against the alternative hypothesis that the probability of hit is different for the two teams yields a p-value of 0.34. Thus, the null is not rejected and there is a conclusion that the difference in proportions of hits between the two teams is not statistically significant.

One limitation of the data available is that hits and misses are recorded by team and squad as a percentage rather than a sequence of individual shots. For example, if the first team hit one of two shots, and the second team hit both shots, the first teams' recorded accuracy is 50%, the second teams' recorded accuracy is 100%, and the squads' recorded accuracy is 75% for the trial. This method of data collection precludes determining exact probabilities of hit, since each shot is not an independent event. If the data from the example presented was available in the form of "0-1-1-1," conditional probabilities could be calculated to give more accurate expectations. However, it is still possible to provide approximate probabilities. When looking at the 180 shots for the squad trials, there are 110 hits, which yields an expected hit percentage of 61%. This can be used as the probability of success while treating the volley of four rocket shots as a binomial distribution. Thus, a commander can expect the following probabilities of having one or more, two or more, three or more, or four hits, respectively, as: 0.98, 0.83, 0.49, 0.14. From a slightly different perspective, by considering a geometric distribution with a 0.61 probability of success, the question of how many rockets it will take to achieve a hit with a certain probability can be answered. From the squad data, the probabilities of achieving at least one hit while firing one, two, three, or four rockets respectively are: 0.61, 0.85, 0.94, 0.98.

An applicable T&R task is 0351-SMAW-1005 "Engage a target with a MK153 SMAW using a Day Optical Device." As mentioned at the beginning of this section, the SMAW has been phased out for the MAAWS, but the T&R manual has not yet been updated to reflect this change. The standard for 0351-SMAW-1005 is "to achieve a hit on the target" and the condition is that each Marine is given one round to fire. This standard is as clear as possible and does not need refinement. However, the condition that each Marine is given one round makes remediation of the task unrealistic. Rockets are crucial to the infantry force, as high explosives are the primary killing assets, and thus every effort should be made to secure additional rocket rounds to fire in training. If the condition were instead three rounds per Marine, the standard could read "to achieve two hits on targets between 150m and 250m." This is reflective of the expected value from all rocket rounds fired during the GCEITF and provides a realistic baseline proficiency to strive for.

F. DATA ANALYSIS CONCLUSION

The analysis presented here was intentionally limited in scope to those metrics which are not quantitatively measured by current training standards and have not been analyzed in detail in the GCEITF EAR. This is the first step in bringing objectivity to the evaluation of unit training. However, there is much work to be done that can enhance the perspective of unit proficiency by thorough data collection and simple analysis. By developing guidelines for data collection that is neither resource nor time intensive, and creatively displaying results in a manner which is intuitive and informative, unit leaders can objectively benchmark their units' proficiency against the rest of the force. Further details and examples of this type of evaluation are presented in Chapter IV.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. PERFORMANCE COMPARISON SYSTEM

As the Marine Corps continues to develop people, equipment, and concepts to prepare for combat with a peer adversary, more responsibility will be placed on leaders at all levels. Marines will be more capable, equipment will be more advanced, and emerging concepts will need to be tested and critiqued. Changes are being made with a sense of urgency as it will be the adversary, not the United States, who determines when the next war will begin. Commanders at all levels will lead formations that are trained, equipped, and organized differently than they were just a few years ago. To be successful, commanders must have a clear understanding of their unit's capabilities and limitations.

Data collection and analysis can assist with evaluating unit training performance. In *Training and Education 2030,* the Commandant of the Marine Corps tasked Training and Education Command (TECOM) to "revise infantry battalion T&R standards based on the past 24 months of Maneuver Warfighting Exercise (MWX) lessons learned in order to facilitate force design implementation and readiness" (U.S. Marine Corps 2023, p. 13). Furthermore, he posed the question "We have nearly three years of data, findings, and lessons learned from Force-on-Force experimentation in support of Force Design 2030. Has this influenced T&R standards?" (U.S. Marine Corps 2023, p. 13). General Berger recognizes that our training standards must be updated to align to the missions that the force will be expected to accomplish. This section proposes a tool to aid commanders in understanding training standards and urges the necessity for quantitative components of such training standards for the purpose of evaluation.

One such implementation of a performance comparison tool is a dashboard-style interface that allows a unit to compare their performance to the rest of the force, which will be referred to as Expectations for Infantry Marines (EIM). Unlike MCTIMS, which has requirements and mandates for its use, EIM can be hosted as a standalone tool for unit leaders to understand how their unit is performing. The benefits of having a tool whose use is optional will be discussed at the conclusion of this chapter. The sequence of use for EIM is given next.

First, users visit the site and explore general information for specific MOS fields, as shown in Figure 21. For the example presented here, all data is from the GCEITF study as presented in Chapter III though the goal for the tool is to build a data repository based on unit uploads.

	Metric/Statistic	~1km Hike	100m Movement to SBF	Mount and Load M240B	Engage Targets (400 rds)	Displace 300m to LOA	Mount M240B and Engage Targets	Total
0	Mean	09:47	01:59	00:57	06:14	04:21	01:51	25:08
1	Standard Deviation	00:59	00:25	00:20	01:32	00:44	00:57	04:57
2	Minimum	08:00	01:19	00:05	03:09	02:41	00:05	15:19
3	Maximum	11:57	03:26	02:06	09:00	06:22	05:00	37:51

Figure 21. Screenshot of Example MOS-Specific Summary Information

The page shown in Figure 21 extends to encompass all figures presented in the 0331 sections of Chapter III. After reviewing general information, users go to the home page and select a T&R task, as shown in Figure 22. For this example, INF-MGUN-3002 “mount a heavy machinegun on a tactical vehicle” will be used.

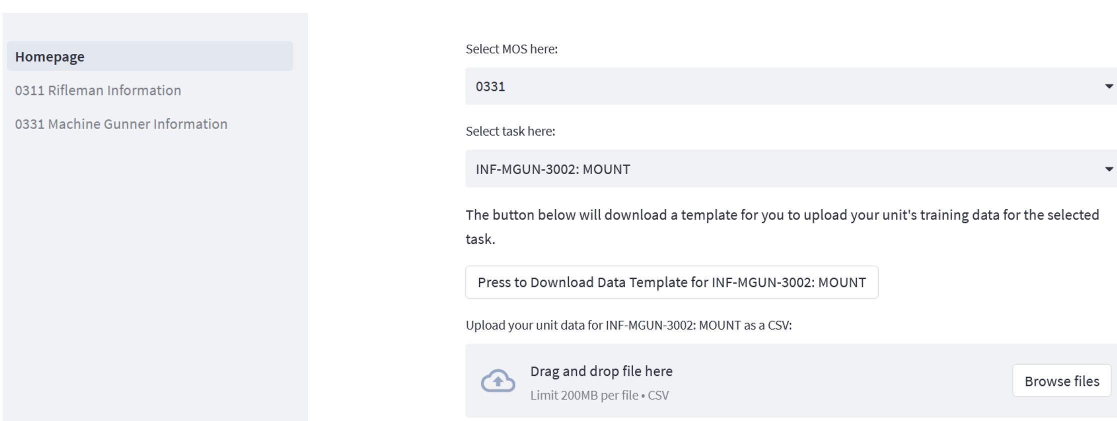


Figure 22. Screenshot of Task Selection

After selecting a task, users can download a template in the form of an Excel file with instructions to collect data. After completing the data collection, users can upload their results, as shown in Figure 23.

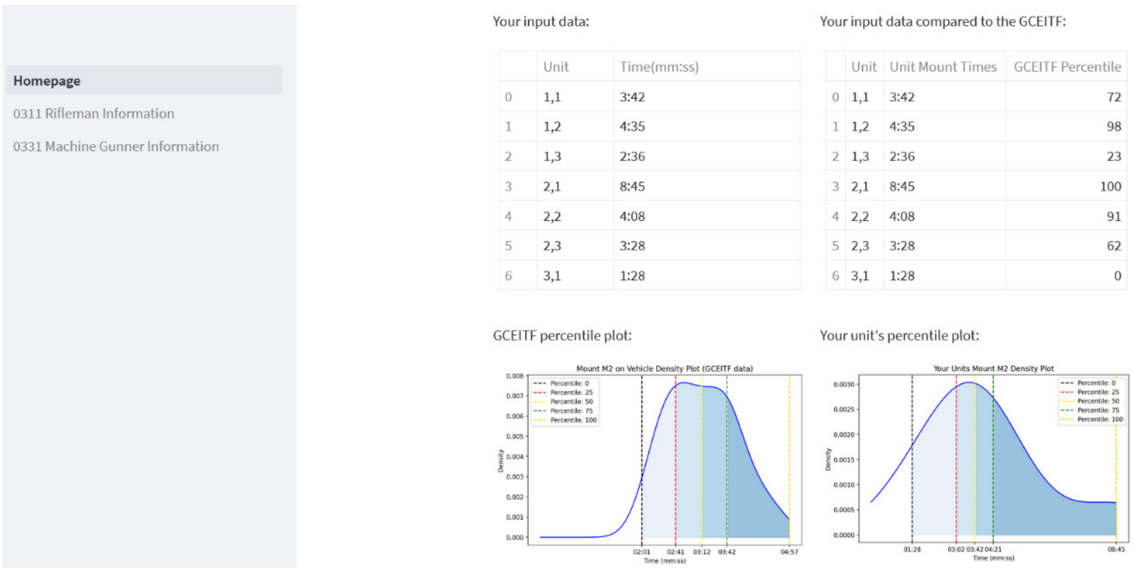


Figure 23. Screenshot of Results for Unit Data

The standard for INF-MGUN-3002 is "to prepare for mounted operations." As mentioned in Chapter III, this task is not generally done with speed as the primary goal. However, it is an easy metric by which to gauge vehicle crew proficiency. By looking at times of subordinate units, as shown in Figure 23, leaders can get a detailed breakdown with percentiles of subordinate unit proficiency. More importantly, the subordinate unit times are benchmarked by the GCEITF times. If, instead of the GCEITF times, the benchmark was Marine Corps averages based on other uploads to EIM, the comparison would be even more valuable and could evolve over time. This information provides much more context about how well a unit performs a given T&R task than does just the T&R standard.

EIM is envisioned as a tool to help leaders adapt to the increasing pace of change occurring within the infantry community. New technologies arriving to the infantry force, such as organic precision fires, are complicated systems that do not have combat-proven doctrine to guide their employment. With force restructure, units are quite literally writing doctrine based on trial and error. To succeed, all leaders in the force need the capability to assess how their units are performing, and this cannot be done without comparing with other units. It is important that this lateral information sharing is done with complete transparency and honesty. MCTIMS is a great snapshot of unit training, but it lacks the depth of understanding that a tool like EIM can provide. However, MCTIMS is also time intensive to upload the required data for every individual completing training events. Because the requirements are time intensive, data collection can be oversimplified or overlooked, which implies that the uploaded data is not as accurate as needed for a tool like EIM to succeed. For this reason, it is paramount that EIM remain an optional tool. It is not intended to function as a system for evaluating unit leaders on fitness reports or otherwise measure commander's performance. By keeping EIM an optional tool to be used at the commander's discretion, thorough data collection can be prioritized, and results will have the pedigree required for an accurate comparison. Last, EIM is not limited to identifying and correcting deficiencies in unit performance. If uploaded data shows that the unit is outperforming the broader force, the commander can recognize the unit for their achievement. More importantly, the commander can then evaluate the tactics, techniques, and procedures (TTPs) that led to the superior performance and share those TTPs with the broader force. As the Marine Corps prepares for peer-conflict, lateral information sharing of TTPs that are proving effective will be critical to success in combat.

V. CONCLUSION

The Marine Corps is rapidly changing, as it should, to prepare for peer conflict. As these changes happen, Marines should be scrutinous of all existing processes and systems to create the most lethal force possible. Advancements in technology are fundamentally changing the battlespace, as unmanned equipment with reconnaissance and strike capabilities leveraging machine learning to optimize employment have extended the range of unit operations farther than ever before. The Marine Corps infantry force will fight much differently than it has traditionally, and there is no historical precedent for employing the emerging weapons and concepts. Cooperation amongst units to find, through trial and error, the best methods of employment must be shared throughout the force. Additionally, to understand how one's unit is performing, objective and quantitative standards must be published and compared to a set benchmark to ensure the progress of the entire force.

Leveraging data from the GCEITF to set an initial benchmark has one distinguished benefit: some data collected will likely not be measured again on as large of a scale and with the appropriate level of detail. Placing sensors on all weapons and targets is unique to the GCEITF, and those findings should be understood given such experiments will likely not be conducted again. Those findings are summarized here.

- For 0311 Rifleman, conducting fire and movement as part of a squad attack, 22% of rounds expended are hits and 43% are effectively fired (meaning a hit or within one meter of the target). Of the 43% of effectively fired rounds, 65%-85% are hits when the shooter is within 100 meters of the target. Farther than 100 meters, less than 50% of the effectively fired rounds are hits.

- When employing the 40mm grenade launcher, squads achieved a mean of 6.5 effective rounds per trial with a standard deviation of 3.5, with 18 rounds issued per trial. There is no significant difference in accuracy between a collection of 0311 Rifleman acting as a squad and a similar collection of non-infantry Marines, though recent updates to entry-level 0311 training are producing better marksman.

- Though the SMAW rocket is no longer employed, the finding of 61% hits firing at targets between 150 meters and 250 meters is a useful benchmark to compare performance with the MAAWS.
- For 0331 Machine Gunners, physical fitness is paramount to providing timely offensive medium machine gun support. Over the course of the trials, there is consistent correlation with movement times throughout the sequence of support. The Marines who began the initial movements quickly also finished the final movements quickly, as opposed to fatiguing throughout the event.
- For heavy machine gun employment, accuracy decreased from the 450 meter target to the 800 meter target, though these ranges are less than 50% of the weapon's published capability.
- For 0341 Mortarman, employing 60mm in handheld mode, 23% of rounds fired landed within the doctrinal measure of suppression of 35 meters. Extending the concept of effective fire to mortars, 60% landed within 75 meters of the target. Additionally, corrections were made effectively. If the initial round fell more than 50m from the target, the following rounds landed closer to the target.

There are three primary paths for recommended future work. First, there are ranges in the Marine Corps that are supported by contractors providing reactive targets. Some of these targets collect data that can be distributed to units upon the completion of training. Probability of hit data from training events involving moving shooters and targets could be valuable sources of data for simulations of small unit combat (Thompson et al. 2020). Such data is notoriously limited. A standard report can be developed that provides relevant information to unit leaders on the accuracy of their units instead of just the raw data. Second, recommended data collection procedures and evaluator notes can be created for specific T&R tasks that provide additional context beyond the standard for each task. In addition to providing the procedures and evaluator notes, a more robust version of EIM can be created to allow for direct comparison of performance. By actively seeking how other units are operating, and how well their TTPs are working, the force will quickly adapt to new technologies and concepts. Last, analysis presented throughout this work and continued analysis as recommended throughout Chapter III can be included in current combat models. Probabilities of hit with various weapon systems, and movement speeds while hiking or in the attack, can be found within the GCEITF dataset. Instead of using averages or a best-guess, realistic distributional information can instead be used as inputs to combat models and produce a more accurate output.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Department of the Army (2000) Mortars. FM 23-90. Washington, DC, https://www.marines.mil/Portals/1/Publications/FM%2023-90%20W%20CH%201.pdf.

Department of the Navy (1992) Tactical employment of mortars. MCWP 3-15.2. Washington, DC, https://www.trngcmd.marines.mil/Portals/207/Docs/TBS/MCWP%203-15.2%20Tactical%20Employment%20of%20Motars.pdf.

Department of the Navy (1996) Machine guns and machine gun gunnery. MCWP 3-15.1. Washington, DC, https://www.marines.mil/Portals/1/MCWP%203-15.1.pdf.

Department of the Navy (2010) Systems approach to training user's guide. NAVMC 1553.1. Washington, DC, https://www.marines.mil/Portals/1/Publications/NAVMC%201553.1_1.pdf.

Department of the Navy (2015) How to conduct training. MCRP 3–08B. Washington, DC, https://www.marines.mil/Portals/1/Publications/MCRP%203-0B.pdf.

Department of the Navy (2016) Unit training management guide. MCTP 8-10A. Washington, DC, https://www.marines.mil/portals/1/Publications/MCTP%208-10A.pdf?ver=2017-03-16-121330-570.

Department of the Navy (2019) Offensive and defensive tactics, MCWP 3-01. Washington, DC, https://www.marines.mil/Portals/1/Publications/MCWP%203-01.pdf?ver=lyn9EYWTlhzLMJdUijR8Yw%3d%3d.

Department of the Navy (2020) Infantry training and readiness manual. NAVMC3500.44D. Washington, DC, https://www.marines.mil/Portals/1/Publications/NAVMC%203500.44D%20Infantry%20T-R%20Manual.pdf?ver=2020-05-21-104009-273.

Emmel DC, George B (2022) The 21C Infantry Marine. *Marine Corps Gazette* 106(6).

Johnson PC, Pinelis J (2019) Day 47: Nerds to the Rescue. Nelowet S, Pritz S, eds. *The Experiment of a Lifetime: Doing Science in the Wild for the United States Marine Corps* (Kindle Direct Publishers) 40–49.

Marine Corps Operational Test and Evaluation Activity (2014) Ground Combat Element Integrated Task Force: Experimental Assessment Plan (Marine Corps Operational Test and Evaluation Activity, Quantico, VA).

Marine Corps Operational Test and Evaluation Activity (2015a) Ground Combat Element Integrated Task Force: Experimental Assessment Report (Marine Corps Operational Test and Evaluation Activity, Quantico, VA).

Marine Corps Operational Test and Evaluation Activity (2015b) Ground Combat Element Integrated Task Force: Experimental Test Plan (Marine Corps Operational Test and Evaluation Activity, Quantico, VA).

Marshall SLA (1947) *Men Against Fire: The Problem of Battle Command* (University of Oklahoma Press, Oklahoma).

Mattis J, West B (2019) *Call Sign Chaos,* (Random House, New York).

McCoy BP (2011) The *Passion of Command: The Moral Imperative of Leadership* (Marine Corps Association, Quantico, VA).

Meehan M (2020) Moving the Corps into the information age: Data-driven training standards and analytics to support evaluation. Master's thesis, Department of Operations Research, Naval Postgraduate School, Monterey, CA, https://calhoun.nps.edu/handle/10945/65406.

Office of the Secretary of Defense (1994) Direct ground combat definition and assignment rule. Department of Defense, Washington, DC, https://www.govexec.com/pdfs/031910d1.pdf.

Office of the Secretary of Defense (2013) Elimination of the 1994 direct ground combat definition and assignment rule. Washington, DC, https://dod.defense.gov/Portals/1/Documents/WISRJointMemo.pdf.

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Report, Biometrika, Great Britain, https://doi.org/10.2307/2333709.

Smeeton NC, Sprent P (2007) *Applied Nonparametric Statistical Methods* (CRC Press, Boca Raton, FL).

Thompson, C., T.W. Lucas and M.I. McDonald, "Estimating the Effects of Combat Load Weight on Mission Outcomes," *PHALANX,* 53(1), 2020, 52–57.

U.S. Marine Corps (2016) Toward the chaos: 2/8 prepares for combat scenarios. Accessed 3 May 2023, https://www.marines.mil/Photos/igphoto/2001515176/.

U.S. Marine Corps (2020) Force design 2030. Washington, DC, https://www.hqmc.marines.mil/Portals/142/Docs/CMC38%20Force%20Design%202030%20Report%20Phase%20I%20and%20II.pdf?ver=2020-03-26-121328-460.

U.S. Marine Corps (2021) Changing the fight: Marine Corps fields new rocket system to infantry Marines. Accessed 14 May 2023, https://www.marines.mil/News/News-Display/Article/2678825/changing-the-fight-marine-corps-fields-new-rocket-system-to-infantry-marines/.

U.S. Marine Corps (2023) Training and education. Washington, DC, https://www.marines.mil/Portals/1/Docs/Training%20and%20Education%202030.pdf.

Wilcoxon F (1945) Individual comparisons by ranking methods. Report, Biometrics Bulletin, https://doi.org/10.2307/3001968.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California