

Virginia Polytechnic Institute and State University

*The Perceptions of the Army Combat Fitness Test (ACFT) by Female Army Officers on Event
Physical Effectiveness and Assessed Practicality of the Assessment's Scoring System*

*Captain Valerie N. Nostrant
US Army, Field Artillery*

*Master of Science
Agriculture & Life Science
Applied Nutrition and Physical Activity Concentration
Committee: Dr. Angela Anderson, Dr. Stella Volpe, Dr. Madlyn Frisard, Dr. John Tisson
Spring 2024*

Table of Contents

Abstract	pg 3
Introduction	pg 4
Literature Review	pg 6
Methods	pg 18
Results	pg 22
Discussion	pg 45
References	pg 56

Abstract

During nine years of active-duty service in the United States Army as an Officer, it has always been abundantly clear that there are discernable differences between male and female soldiers, which is amplified when you compare their physical abilities and performances. With the development of the Army Combat Fitness Test (ACFT), set to replace the Army Physical Fitness Test (APFT), it was a concern from Congress to the US public what would the impacts of this test change would be on the careers of service members. The female US Army Officer population was underrepresented in both the beta testing and in the development of the ACFT events. This survey investigated the perceptions of these women on individual event effectiveness and the ongoing revisions of the scoring systems. Overwhelmingly, this survey showed the efficacy of many of the events in which the desired targeted skills were not being evaluated and that a gender based scale was required so as not to have a biased or unfair test of physical abilities.

Introduction

The United States Army began implementing physical fitness assessments before the American Civil War.¹ These tests have had a variety of evaluations and standards, with the longest-standing test being employed in 1980, the Army Physical Fitness Test (APFT).² After over thirty years of use, this test was replaced in October 2022 by the newly revised and more comprehensive Army Combat Fitness Test (ACFT).² The road to the ACFT was not straightforward. It was in the making for over a decade with changes to scoring standards, events, and even a Congressional inquiry into the grading standards. Initially, the test was designed to focus on skills and fitness required to support combat operations in the Global War on Terrorism Era. The initial evaluation criteria intent was to create a gender and age-neutral test that would equally prepare Soldiers, regardless of job specialty, and increase the lethality across the force. After a Congressional intervention in 2020, calling for a pause in the implementation of the ACFT, the US Army commissioned an independent review of the scoring disparities between female and male servicemembers to determine if the test was, indeed, created equal.

During the development of the ACFT, female officers were a population primarily left out of beta testing and design. With this fact in mind, this research was developed to assess the perspective of female US Army Officers, examine their perceived viewpoint on the effectiveness of the tests, and determine if there are potential disparities due to gender present in the testing and scoring. The new US Army physical fitness test for official record, the ACFT, could have compounding effects on promotions, job competitiveness, and overall competency determinations on evaluations. Therefore, allowing a vast pool of voices to be heard in an ever-increasing competitive Officer environment is crucial. The purpose of this research is to provide insight into the scoring and effectiveness of the ACFT.

The overall perception of these six different events and their respective scoring scales were measured on the Likert scale with effectiveness values ranging from strongly disagree (1), disagree (2), neutral (3), agree (4), and strongly agree (5). After four revised scoring systems, initially developed to have an age and gender-neutral scale, it was overwhelmingly perceived that a gender-based scale would be required to maintain an equitable perception of male and female service members based on the current physical events. These events have been refocused on different “warrior tasks and skills” based on current US Army doctrine. Throughout the survey, there were twenty-three event effectiveness questions with quantitative perception values, with fourteen trending from strongly disagree to disagree and nine ranging from neutral to agreement.

Literature Review

Background

The United States (US) Army began assessing physical fitness in 1920 as part of a comprehensive review of the force. The initial studies and tests were geared to the war-fighting functionality of the soldier on the battlefield and skills related to combat operations of small echelon units, such as the Section, Squad, Platoon, or Company.¹ The US Army has shifted to a holistic health view in integrating a new fitness assessment as a biannual requirement for the active-duty force. This test focuses on functional movements used during combat in the wars in Iraq and Afghanistan. The integration of the new ACFT has been fraught with changes and refinements, with political and public opinion interdicted throughout the integration phases. The question remains, is this new test equitable for females and males once the scoring standards deviate from equality in the gender-neutral sphere? The test was initially designed to comply with the 2015 National Defense Authorization Act without differential standards or evaluation by gender.² However, in 2020, United States Senators Kristen Gillibrand and Richard Blumenthal submitted a letter to be included in the Fiscal Year (FY) 2021 National Defense Authorization Act (NDAA) to pause the implementation of the ACFT and examine the events and scoring standards based on the dramatic disparity between male and female soldier scores and pass rates.³

Early Tests and Influence

In 1858 the Army began testing the first-ever assessment of physical fitness at the United States Military Academy (USMA).¹ During this period, the United States Military Academy was the hub for creating and refining Army doctrine, including physical fitness. The first physical fitness test was proposed by First Lieutenant John Kelton, who traveled to Europe prior to the

outbreak of the Civil War to study "gymnastics" and create a comprehensive physical fitness test for USMA.⁴ His proposed test involved climbing a 15-foot wall, vaulting a horse 15 hands high (5 feet), leaping a 10-foot-wide ditch, running a mile in 8 minutes or two miles in 18 minutes, walking 4.5 miles in one hour, and walking 3 miles in one hour with a 20-pound knapsack, arms, and equipment. These recommendations and others were tested until they were put on pause in 1861 with the outbreak of the Civil War.²

The Army published *General Order Number 44* in 1906, which initiated a physical fitness and training regime for infantry, artillery, and horse-mounted soldiers, including ruck marches of 12 miles, 18 miles for mounted troops, riding tests, and a three-day 45-mile dismounted movement.⁵ President Theodore Roosevelt was a driving force behind these reforms following his service with the infamous Rough Riders during the 1898 Spanish-American War. Not only was Roosevelt concerned with the physical fitness of the military officers, but he directed that all officers considered for promotion be tested for skill and endurance, the results of these tests determining both promotion and continued service.

As the study of physical fitness and fitness testing implementation began to take off in American Universities,³ so did it in the US Army. The War Department published the *Mass Physical Training for Use in the Army and Reserve Officers' Training Corps*⁶ to ensure that Officers entering the Battlefields following World War I would be adequately prepared to lead troops in future combat situations. This test includes an obstacle course designed with a 3-foot hurdle vault, 10-foot wire entanglement, climbing a 5-foot ramp, negotiating a 1-foot wide, 20-foot-long plank, jumping from the top of a 5-foot ramp over a 10-foot trench, and climbing an 8-foot, smooth wall. The test was designed to incorporate aerobic endurance between obstacles and test the physical skills necessary to navigate trench warfare.⁶

In the 1930s, Dr. Charles McCloy and Dr. Arthur Esslinger, physical educators at the University of Iowa and Stanford University respectively, were the first to conduct scientific efforts to determine military physical readiness before the outbreak of World War II.⁷ Their research aimed to create testing to examine circulo-respiratory endurance, muscular contraction speed, muscular endurance, and mesomorphic build-up. As a result of their research, the US published its first hard copy physical fitness doctrine, *Field Manual (FM) 20-21*⁸, in 1941. *FM 20-21* outlined a program to evaluate "...the condition and aptitude of the men to be trained. The best method of determining this condition and aptitude of the group is by comparison with known standards."⁸

One year after *FM 20-21* was published, the Women's Auxiliary Corps, later the Women's Army Corps (WAC), was established and remained in effect, separating the female and male Army forces until 1978.⁹ *FM 35-29* was published in 1943 outlining the physical training guidance for the WAC and includes a self-test comprised of bent knee sit-ups, push-ups, wing lifts, squat thrusts, running, and stork stand.¹⁰ Push-ups were executed with knees on the ground. Wing lifts involved prone trunk extensions with the fingers interlocked behind the head. Running could be in place (stationary) or for an unspecified distance to progressively increase time or distance. The stork stand involved balancing on one leg with the unsupported leg on the supported knee, eyes closed, and arms crossed in front of the body.¹⁰ Unlike *FM 20-21* for male US Army soldiers, there were no set standards, but instructors were encouraged to use these guidelines to evaluate the fitness of WAC candidates.

Post-World War II and Korean War

As World War II ended, the US Army continued to take lessons learned and refine its physical training and testing standards. For example, in 1944, the *War Department Pamphlet 21-*

9¹¹ was published with an entire chapter devoted to this fitness testing. The requirements were meant to be minimal on both equipment and time, resulting in seven events, each worth 100 points with a scale of very poor to excellent. The scale provided such qualitative meanings as “excellent” was more than 600, “good” was 445-600, “average” was 247-444, “poor” was 114-246, and “very poor” was less than 114. This point scale was to provide unit commanders with a quantitative assessment of the physical readiness of an individual Soldier or unit.¹

Following the conclusion of World War II and throughout the Korean War, the US Army continued to refine and adjust physical fitness training and testing doctrine. These reforms mirrored the changes taking place with the American public concerning an increased emphasis on aerobic testing. In 1957 the revision of *FM 20-21* and the introduction of the complimentary *Technical Manual* ^{TM 21-200}¹² led to many scoring changes and revisions of the “ew "Physical Fitness Tests." This new aerobically based test stated that a 1-mile run should be administered with four days of the testing. This was the first time a run longer than 150 meters was directly included in testing standards. In 1961, the Army published a change to *TM 21-200* to create an increased focus on "combat readiness"¹³ based on physical lessons learned during the Korean War. As a result, the Physical Combat Proficiency Test was born, with eight events designed to test and assess the essential functional skills needed in combat. The events of the new assessment were a 40-yard low crawl, horizontal ladder movement, "run, dodge, jump obstacle," grenade throw, and the 1-mile run. Each event maintained a maximum 100-point score.¹³

The Cold War and Vietnam War

Change 2 to the *TM 21-200* ¹³ was published in 1961, which signified a return of a combat-ready force following the Korean War. The revision to Army fitness testing focused on eight events designed to test skills deemed necessary for combat. The “combat” required

essential skills: running, jumping, dodging, climbing, or traversing, vaulting, carrying, balancing, falling, and swimming.¹⁹ The 1961 Physical Combat Proficiency Test contained a scoring scale of 100 points per event. It required 60 point minimum in each event for all male service members under 40 years of age to maintain, including recruits in Basic Combat Training (BCT) for initial entry.¹³ In 1963, *DA Pam 21-1* was published to include “Minimum Physical Fitness Test-Male,” which lightened the physical requirements for those performing duty only in sedentary or administrative capacities. This would be the first-time different standards were published based on duty position. However, this test was still mandatory as an annual requirement.¹⁴ In conjunction, *DA Pam 21-2* was published in 1963 to include an “Army Minimum Physical Fitness Test-Female,” which outlined the minimum physical requirements for women under 40 years of age, similar to the test standards in *DA Pam 21-1*.¹⁵

For the first time, in 1965, an Army regulation was published, *AR 600-9*, that mandated twice yearly physical fitness testing would occur for both women and men. This standard has been included in physical regulations since and is in effect today.¹⁶ By 1969, the fifth revision to *FM 21-20* was published, outlining four tests commanders could utilize for various needs and formation types.¹⁷ The tests included the Physical Combat Proficiency Test, the Minimum Fitness Test-Male, Airborne Fitness Test, and the Inclement Weather Test. The 60-point minimum standard for any event was retained, as well as the utilization of these tests for initial recruits and the active force.¹⁷

A significant shift occurred in 1973 when a further test was added to *FM 21-20*.¹⁸ There were now two tests designed for the operational force; the Advanced Physical Fitness Test (for combat and combat support units) and the Staff and Specialist Fitness Test (for combat support units; including staff, faculty, and students assigned to Army schools). This duty position shifts

in testing requirements were also outlined with different scoring standards for male and female service members in the Staff and Specialist Fitness Test, which included no minimum passing standard and some tests, such as the push-up, to be completed with as many as possible. A woman's standard was not included in the Advanced Physical Fitness Test.¹⁸

The US Congress directed the disestablishment of the WAC in 1978, and women were integrated into the regular Army.⁹ This included revising and combining the physical training and testing doctrine *FM 21-20* (for men) and *FM 35-20* (for women).²⁰ In addition, the Army was directed to create a gender-integrated assessment, easy to administer in any environment or location and required minimal or no equipment.⁴ The disestablishment of WAC and the decision to create a singular fitness assessment led to the birth of the longest-standing Army physical fitness assessment and test, the Army Physical Fitness Test (APFT).

The Thirty-Year Physical Assessment: Army Physical Fitness Test (APFT)

In February 1980, General (GEN) Donn Starry, the Commander of US Army Training and Doctrine Command (TRADOC) requested the establishment of the Army Physical Readiness Testing (APRT) Study Group, led out of Fort Benning, Georgia to update *FM 21-20* to include a comprehensive test that was gender integrated and would establish the baseline physical fitness requirements for all US Army personnel.²¹ By the end of their review, the APRT Study Group had submitted the approved changes to the 1980 edition of *FM 21-20*.²⁰ The three events were 2 minutes of push-ups and sit-ups respectively followed by a 2-mile run. The test was allowed to be conducted in shorts, t-shirt, and running shoes. Prior physical fitness tests were taken in their uniforms and boots. This new physical assessment also contained age-and gender-based standards. When the initial physical fitness test was developed in 1980, scoring tables did not include service-member standards for those over 40 years of age. After 1986, an

adjustment was made to include an increased range beyond 40 years of age with approved medical screening.²²

In 1985 the test was officially named the Army Physical Fitness Test (APFT).²³ The test continued to evolve with subsequent changes including alternate events to accommodate service members with physical limitations, commonly referred to as temporary or permanent profiles. These alternate test events included an 800-yard swim, 6.2-mile cycle and a 3-mile walk. The evaluation criteria for the 100-point standard were consistently raised which is reflected in the 1986 edition of *FM 21-20*.²² For example the increase in mile running time standard in the ninth and final edition of *FM 21-20* was published in 1992 with scoring changes only to the 2-mile run point system for women age 27-31.²⁴ Of all the changes made between 1980-1992, only the sit-up standards were identical for men and women.

Global War on Terror (GWOT) Influence

As a result of the Global War on Terror (GWOT), an increased number of US Army personnel were required to support the ventures: Operations Enduring Freedom (OEF) and Operation Iraqi Freedom (OIF), and then subsequently Operation Inherent Resolve (OIR).²⁵ These global campaigns highlighted the increased need for a fitter and more lethal force as the Department of Defense shifted its focus from resetting and sustaining a military force to instead, a combat-focused and more lethal force.²⁵ In 2003, the United States Army Physical Fitness School (USAPFS) proposed the PRT Field Manual (FM 3-25.20) revision to include a potential replacement for the 3-event APFT to a 6-event physical readiness test, yet to be named.²⁶ The revision was met with such opposition throughout development and preliminary testing that it was ultimately halted and not included in the 2005 revision of FM 3-25.20. The public opposition, however, did not deter the Department of the Army from moving forward to begin

changing. By 2005, multiple research studies showed an increased number of “Marginally fit”²⁷ soldiers entering the US Army with increased rates of injury upon entering initial and advanced training. This increase in injury rates continued to drive the need for USAPFS to begin the development of a new physical readiness test and training guidance.²⁶

By 2010, the Training Circular (TC) 3-22.20 stated in its introduction, “Combat readiness is the Army’s primary focus as it transitions to a more agile, lethal, and survivable force.”²⁸ Shortly after TC 3-22.20 was published, USAPFS hosted an APFT revision working group to create the calendar and roll out for an eventual new physical fitness test. At the beginning of summer 2011, USAPFS rolled out a feasibility study with a 5-event test consisting of a standing long jump, rower, 60-yard shuttle run, push-ups, and a 1.5-mile run.²⁸ The review released in late 2011 showed concerns about the developmental process and the potential events.

Following the results of the 2011 study, USAPFS was ordered to conduct a secondary study in May 2012 that involved multiple organizations and took a measured three-pronged approach.⁴ This new approach included systematic literary reviews, examining the linkages between capacity and functional fitness assessments to Common Soldier Tasks (CST), and validating the final testing events against CST.²⁹ The US Military Academy, the Initial Military Training Center of Excellence, the US Army Research Institute of Environmental Medicine, the US Army Public Health Command, and civilian researchers participated in the study examination.³⁰

Army Combat Fitness Test (ACFT) Implementation & Changes

In 2013, the US Army began development of the Army Combat Fitness Test (ACFT), which was designed to increase "combat readiness" and address the "declining health and fitness standards of incoming recruits" by focusing on creating events to support the 113 essential

"warrior tasks and drills."³¹ In the fall of 2018, ACFT 1.0 was released for initial testing. This fielding included a test made up of six events: 3-repetition max deadlift, standing power throw with a 10lb ball, hand release push-ups, sprint drag carry, leg tuck, and a 2-mile run.³² The scoring scale used for testing was based on Military Occupational Specialties (MOS) broken into three tiers, Black, Gray, and Gold, in descending minimum passing scores required but was age and gender-neutral.³¹ These three tiers were focused on Combat Arms, Combat Support, and Specialty MOSs. The purpose was to emphasize specific MOS having a higher baseline for physical fitness based on the Specialties' physical demands.³² However, initial testing results from 63 units across the Regular Army, Army Reserves, and National Guard saw an increased disparity between male and female service members and those with highly intellectual or specialized job requirements (medical, legal, and cyber personnel). Overall, the pass rate for the ACFT in fiscal year 2019 was 72%, with only 21% of women passing and 81% of men passing. Passing is determined by reaching 60 out of 100 points in each event.³⁵

In the spring of 2020, the implementation of the ACFT, like the rest of the world, was affected by the COVID-19 Pandemic. ACFT 1.0 was set to become the physical fitness test for record in October 2020. With the COVID-19 mitigations put in place by the Department of Defense, the test could only be administered for diagnostic purposes and not for record use in military schools, Officer Evaluation Records (OER), or Noncommissioned Officer evaluation records (NCOER) and in the placement on Order of Merit Lists (OML) for promotion.³³ During FY20, the Army rolled out ACFT 2.0, which dismissed the tiered scoring system based on MOS but remained gender and age neutral. The Leg Tuck was the overwhelming event that both male and female service members routinely failed, so the option to do a plank for two minutes was

provided as an alternate event; however, regardless of the time held, an individual could not score above 60 points on the event.³⁵

Senators Kirsten Gillibrand (D-NY) and Richard Blumenthal (D-Conn) submitted a letter to the chairmen and ranking members of the House and Senate Armed Services Committees requesting a halt to the ACFT 3.0 roll-out and calling for further evaluation of grading standards.³⁷ In the fall of 2020, beginning in FY21, the US Army moved to the ACFT 3.0, which officially called for the plank as an optional event instead of the leg tuck, and up to 100 points could be scored.³⁴ A second scoring standard was also released for promotion values based on gender. However, the test for military records, remained age and gender neutral. The Senators requested a suspension of the test until an independent study could be conducted and included in the final version of the 2021 National Defense Authorization Act (NDAA).³⁸ Over 54% of women were still failing the ACFT, compared to only 7% of men failing. The Senators also called into question the Army predictability measures, stating that the initial testing only included 16 volunteered women with an average age of 23.³⁸

The ACFT implementation to the US Army has been fraught with constant criticism from exercise professionals, social media, and politicians. When the Senate appointed an independent evaluation by the Rand Corporation to conduct an independent review of the ACFT's tests and scoring, the results shaped the final version more than any other data to date. The RAND Corporation is a nonpartisan organization whose purpose is to use researches to develop solutions to public policy challenges on a global scale. This Independent review was conducted by the RAND Arroyo Center's Personnel, Training, and Health Program, federally funded research sponsored by the United States Army.³⁷ Using a multidimensional approach involving evaluation of previously gathered data, interviews, and review assessments of the evidence used

by the US Army, the RAND review findings were encompassed by the following: The evidence base to support the ACFT is incomplete, ACFT scores collected during the diagnostic period show groups failing at significantly higher percentages, research in multiple military settings has shown training can improve pass rates and that the US Army would benefit from a formal management structure to oversee the refinements to the ACFT over time.³⁷

The findings by the RAND Corporation drove the US Senate to push the US Army for another round of scoring changes to account for the disparity in pass rates amongst women and to consider the potential implications for personnel management that would result in higher failure rates among specific demographics. In a final attempt to create a balanced ACFT, the US Army rolled out the revised ACFT on April 1st, 2022, with a mandatory bi-annual requirement without administrative consequences until April 1st, 2023.³⁶ Administrative consequences could include separation from service, negative reflections in performance reports, and used as a negative factor for promotions. The most significant change resulting from the report was the implementation of the Plank as the fifth event tested with the complete elimination of the leg tuck. In terms of scoring, a gender and age scoring scale were introduced.³⁶ As it stands presently, all service members are required to take the ACFT twice a year for the record, which may now appear in personnel records.

Conclusion that leads into research

Historically speaking, the US Army has a long and varied history of physical readiness training and testing that has evolved to follow the requirements of an ever-evolving force. One key finding in the Rand Corporation research study was that the ACFT and similar physical assessments were disproportional in scoring and the testing of females versus males. After reading the Rand Study, the desire to conduct similar but smaller scale research focused on the

female Officer population was developed. The test subject population used to develop the ACFT was disproportional in the testing numbers of females to males. It also never directly addressed the physiological differences in the male and female musculoskeletal and circulo-skeletal systems.

To address those gaps in the development of the events, this study looks to investigate the perceptions of female US Army Officer on the desired physical outputs of the ACFT events and the division of the scoring scale to aggregate it by gender. The questions were designed to derive an answer to the following: Does the ACFT address the physical skills it was intended to? Was the idea and design of the ACFT effective in being a gender agnostic assessment? Is the most recent implementation of a gender-based scoring standard more effective for limiting the ACFT as a discriminator between men and women? It is important to note that this study is purely based on the perceptions of a specific population: female US Army Officers. This population was chosen because it was underrepresented in the initial testing of the events.

Methods

This survey and the questions contained in it were designed to determine if there were any perceived specific biases in the events that comprise the ACFT and the approved scoring system that now separates events by gender and age, contrary to the initial intent of the ACFT in relation to the perception from female officers. In 2022, the Rand Corporation published a study showing that female enlisted and Officer servicemembers scored significantly lower than their male counterparts on the AFCT.³⁹ The reflections of the Commissioned Officers and particularly with reference to the annual Officer Evaluation Reports (OERs), is that the ACFT scores potentially impact the ability of a servicemember for nominative positions and promotions.

Study Design

To initially test the reliability and validity of this survey, 11 female officers were chosen to take the survey, the group was diverse in terms of Military Occupational Specialty (MOS) and in geographic location. Approximately 120 days later the test group was re-surveyed to test the validity and reliability of the survey questions. At this time 9 of the original 11 completed the survey. One was excluded due to leaving the US Army and one was excluded for personal reasons.

Survey Population

The survey was released via social media to a closed Facebook Group comprising only female Army Officers. The main aim of this Facebook Group is to provide mentorship and career advice without gender discrimination. During the month the survey was open, 55 participants completed it in its entirety. Demographic collected data ensured that there were no duplications and that all participants met the baseline criteria of being female in gender and as a US Army

Officer. Demographically, gender, race/ethnicity, Officer rank, time in the Army and time as an Officer were measured to determine trends across any group.

Of the participants, 42 participants were Caucasian or White, 8 were Hispanic/Latino, 4 were Asian, and 1 identified as multiracial (Table1). 19% identified as a Second or First Lieutenants (O-1, O-2), 41% identified as Captains (O-3) and 40% as Majors, Lieutenant Colonels or Colonels (O-4, O-5, O-6) (Table 2). The majority of participants fell into the 5 to 8 years of service category. Table 2 indicates that 43.68% of participants were Captains with 5 to 8 years of service in the Army and as an Army Officer. Tables 1 and 2 demonstrate that the average participant of this survey was a Caucasian or White female with the rank of CW2-3 (Chief Warrant Officer 2 or 3) or O-3 (Officer 3; Captain) with 5 to 8 years of service in the US Army.

Table 1. Survey Participants Demographics for Rank and time as a US Army Officer and in the US Army

Demographic Characteristic	<i>n</i>	M	SD
What is your current rank?	55	2.27	0.76
WO1, O1-O2	9		
CW2-3, O3	23		
CW4, O4-O5	22		
CW5, O6-O7	1		
How many years have you been in the US Army?	55	2.84	1.20
0 to 4 years	4		
5 to 8 years	24		
9 to 13 years	12		
14 to 18 years	7		
18+ years	8		
How many years have you been an Officer in the US Army?	55	2.56	1.20
0 to 4 years	11		

5 to 8 years	19		
9 to 13 years	12		
14 to 18 years	9		
18+ years	4		

Table 2. Survey Participants Demographics for gender and race/ethnicity.

Demographic Characteristic	<i>n</i>	M	SD
What is your gender?	55	1.00	0.00
Male	0		
Female	55		
What is your race/ethnicity?	55	5.07	1.87
American Indian or Alaska Native	0		
Asian	4		
African American or Black	0		
Hispanic or Latino	8		
Caucasian or White	42		
Native Hawaiian or Other Pacific Islander	0		
Multiracial	1		
Another race/ethnicity not provided here	0		

Statistical Analysis

The survey (Appendix A) was divided into four sections. The sections comprised in the survey are Administrative & Demographic, ACFT Individual Event Assessments, ACFT Event Scoring Assessments, and Overall ACFT Assessments. Except for the Administrative and Demographic section, the scores are based on the Likert scale. Each section is broken down with a Macro analyses of data for all 55 participants and then broken down into three sub-groups. The subgroups are based on the administrative data representing their ranks in the US Army. The first group comprises Second and First Lieutenants, known as O-1 and O-2, respectively, for the duration of the results analyses. The second group is comprised of Captains, known as O-3. The

final group combines two groups, first the Major and Lieutenant Colonel ranks, O-4, and O-5. Then, the 1 participant who identified as a Colonel, or O-6, was added to this cohort. The final three ranks are all considered Field Grade Officers, a qualification based on their time in service and military education level. The data analyzed through Prism (GraphPad Software, San Diego, CA), and the frequency of reported results on the Likert Scale was compared for the perceptions of the remaining three sections. Mean differences between ranks were assessed using a One-way ANOVA with a Tukey's multiple comparisons test.

Results

Origin and Validation Survey

After validation survey results were collected, the survey responses of the nine participants who completed both surveys were plotted on an XY-graph with the means for each question used as the data points to determine if there were significant differences in answers between the surveys. Of the means for seventy-five questions on the surveys, our data showed that the R-squared value of Survey #1 was 79%, and the R-squared value of Survey #2 was 86%. The P value was 0.518. This data depicted in Figure 1.

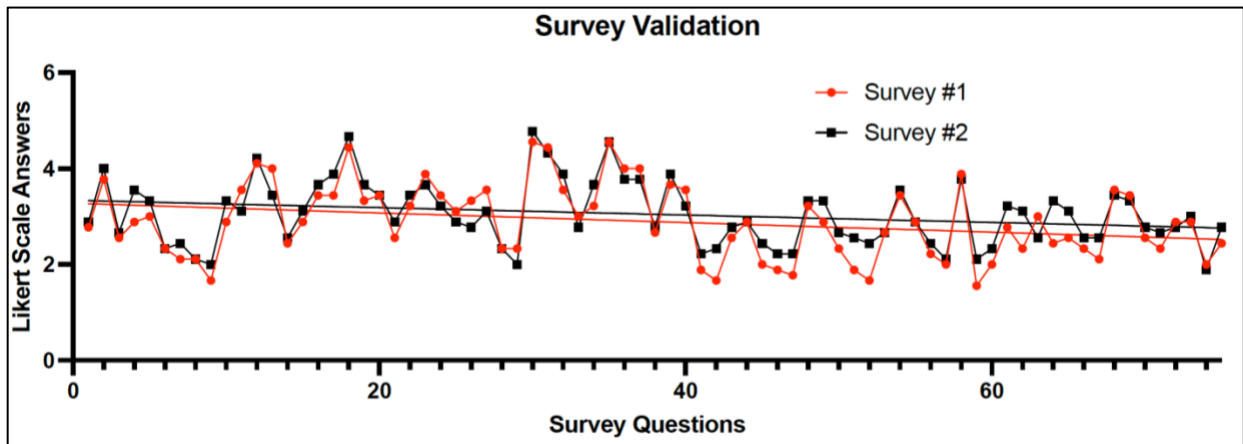


Fig. 1. Validation Survey Results with R-Square value. Survey #1 consisted of 11 participants who completed all 75 questions. 8 weeks later Survey #2 was completed by the same 11 participants and then the linear regression was plotted to determine reliability of the Survey.

ACFT Individual Event Assessment

There are six events in the ACFT, each with specific tasks that the event is either to train service members to perform or provide a particular physiological test. Of note is that the performance standards for Event One were designed to assist in preventing injuries. As injury prevention is one of the reasons for developing and implementing the ACFT, this was directly asked in the survey if injury prevention resulted. Figures 2a-d through 7a-d each represent one of the six events, and then a specific population is denoted in the title of each figure. There is the

overall assessment, and then the data further broken down by rank in the three subgroups previously discussed.

Figure 2.a: Overall Event Effectiveness Assessment **Figure 2.b: O-1/O-2 Event Effectiveness Assessment**
Event 1: 3 Repetition Max Deadlift **Event 1: 3 Repetition Max Deadlift**

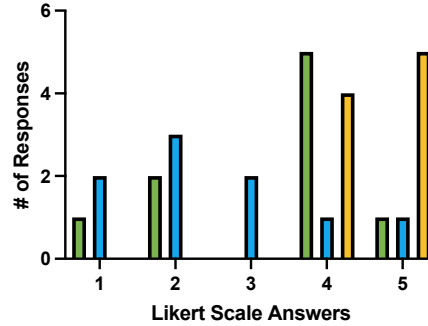
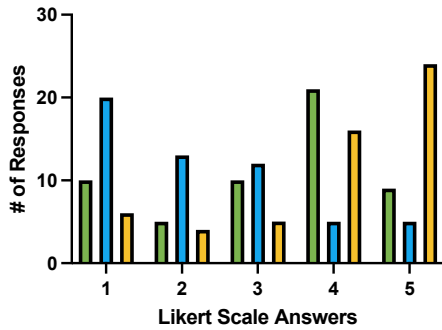


Figure 2.c: O-3 Event Effectiveness Assessment **Figure 2.d: O-4/O-5/O-6 Event Effectiveness Assessment**
Event 1: 3 Repetition Max Deadlift **Event 1: 3 Repetition Max Deadlift**

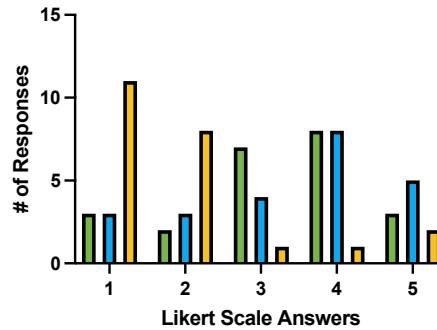
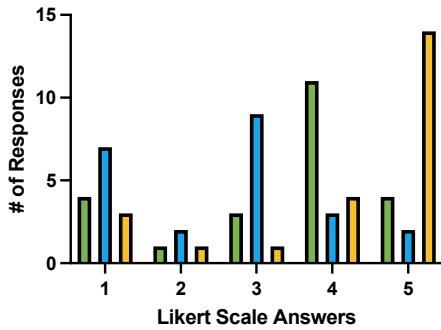
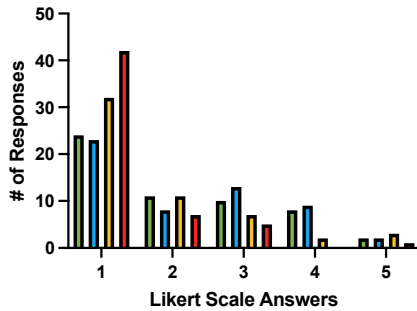
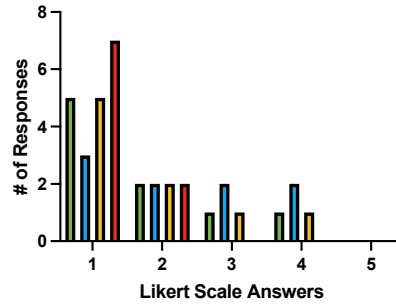


Fig. 2a-d. Display the responses to the following questions about the effectiveness of Event 1: 3 Repetition Max Deadlift. Green indicates the responses to the question: Is this event effective in measuring the ability to extract and carry a casualty on a litter? Blue indicates the responses to the question: Is this event effective in measuring the ability to lift & carry heavy loads from the ground? Yellow indicates the responses to the question: Is this event effective in preventing lower back injuries?

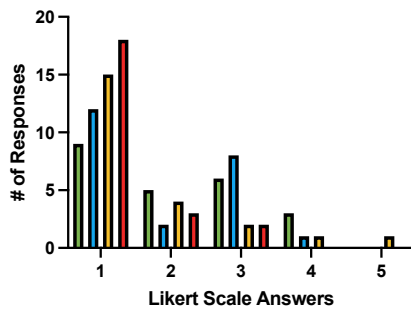
**Figure 3.a: Overall Event Effectiveness Assessment
Event 2: Standing Power Throw**



**Figure 3.b: O-1/O-2 Event Effectiveness Assessment
Event 2: Standing Power Throw**



**Figure 3.c: O-3 Event Effectiveness Assessment
Event 2: Standing Power Throw**



**Figure 3.d: O-4/O-5/O-6 Event Effectiveness Assessment
Event 2: Standing Power Throw**

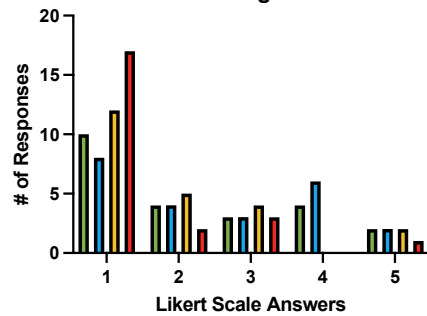
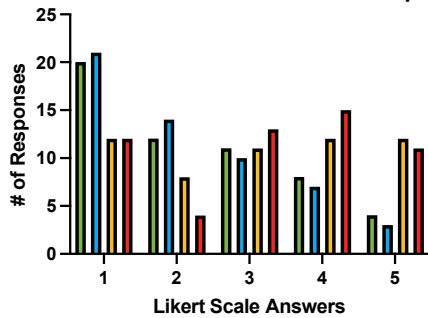
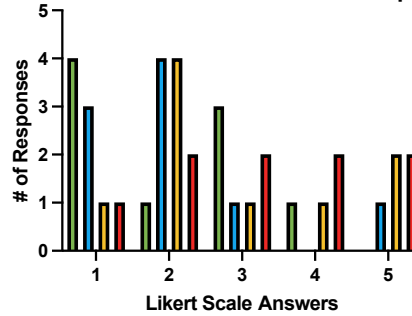


Fig. 3a-d. Display the responses to the following questions about the effectiveness of Event 2: Standing Power Throw. Green indicates the responses to the question: Is this event effective in measuring explosive power, balance, and flexibility? Blue indicates the responses to the question: Is this event effective in measuring the ability to throw equipment over or onto an obstacle? Yellow indicates the responses to the question: Is this event effective in measuring the ability to lift soldiers up or assist them over a wall? Red indicates the responses to the question: Is this event effective in measuring the ability to jump across & over obstacles?

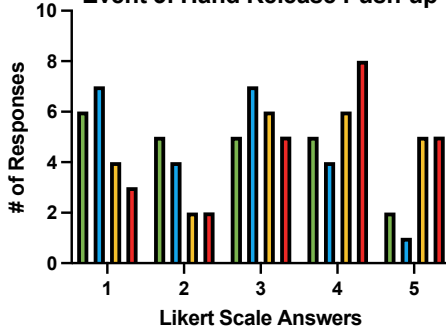
**Figure 4.a: Overall Event Effectiveness Assessment
Event 3: Hand Release Push-up**



**Figure 4.b: O-1/O-2 Event Effectiveness Assessment
Event 3: Hand Release Push-up**



**Figure 4.c: O-3 Event Effectiveness Assessment
Event 3: Hand Release Push-up**



**Figure 4.d: O-4/O-5/O-6 Event Effectiveness Assessment
Event 3: Hand Release Push-up**

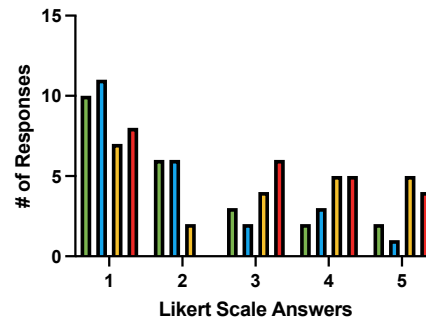
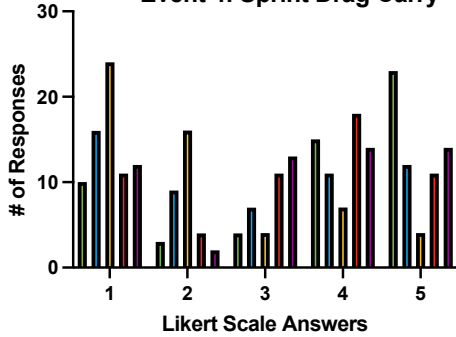
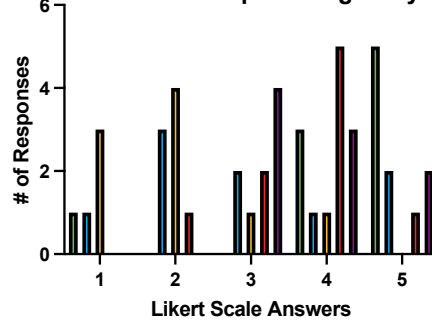


Fig. 4a-d. Display the responses to the following questions about the effectiveness of Event 3: Hand Release Push-up. Green indicates the responses to the question: Is this event effective in measuring your ability to move obstacles equal to or greater than your body weight? Blue indicates the responses to the question: Is this event effective in measuring your ability to push an opponent away during man-to-man contact? Yellow indicates the responses to the question: Is this event effective in measuring your ability to get to and from the ground during evasion & maneuvering? Red indicates the responses to the question: Is this event effective in preparing you to move from the prone position in order to shoot, take cover or low crawl?

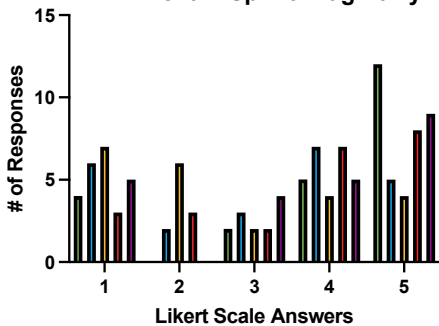
**Figure 5.a: Overall Event Effectiveness Assessment
Event 4: Sprint Drag Carry**



**Figure 5.b: O-1/O-2 Event Effectiveness Assessment
Event 4: Sprint Drag Carry**



**Figure 5.c: O-3 Event Effectiveness Assessment
Event 4: Sprint Drag Carry**



**Figure 5.d: O-4/O-5/O-6 Event Effectiveness Assessment
Event 4: Sprint Drag Carry**

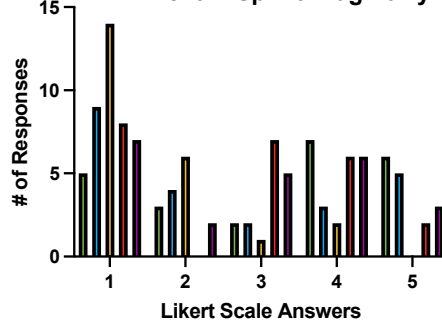
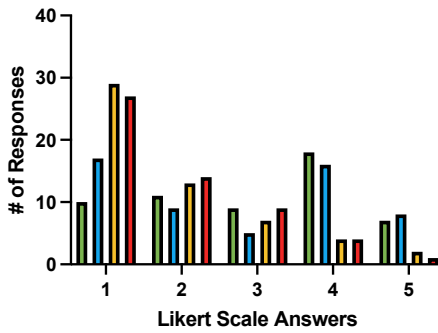
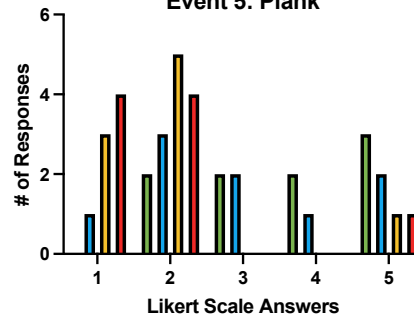


Fig. 5a-d. Display the responses to the following questions about the effectiveness of Event 4: Sprint Drag Carry. Green indicates the responses to the question: Is this event a good test of strength, agility, muscular endurance & anaerobic capacity? Blue indicates the responses to the question: Is this event effective in measuring your ability to react faster to direct or indirect fire? Yellow indicates the responses to the question: Is this event effective in measuring your ability to build hasty fighting positions? Red indicates the responses to the question: Is this event effective in measuring your ability to extract and carry a casualty? Purple indicates the responses to the question: Is this event effective in measuring your ability to carry ammunition to a fight position from a vehicle?

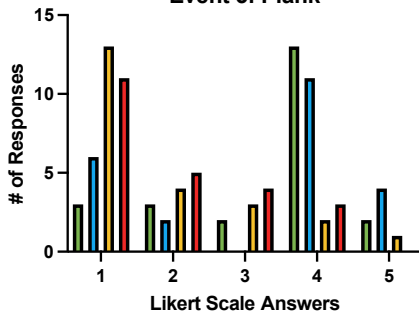
**Figure 6.a: Overall Event Effectiveness Assessment
Event 5: Plank**



**Figure 6.b: O-1/O-2 Event Effectiveness Assessment
Event 5: Plank**



**Figure 6.c: O-3 Event Effectiveness Assessment
Event 5: Plank**



**Figure 6.d: O-4/O-5/O-6 Event Effectiveness Assessment
Event 5: Plank**

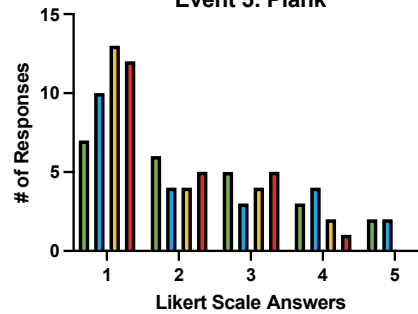


Fig. 6a-d. Display the responses to the following questions about the effectiveness of Event 5: Plank. Green indicates the responses to the question: Is this event is a good test of strength, agility, muscular endurance & anaerobic capacity? Blue indicates the responses to the question: Is this event assists in preparing you for load bearing exercises and avoiding back injuries? Yellow indicates the responses to the question: Is this event prepares you to move over walls & obstacles? Red indicates the responses to the question: Does this event prepares you for climbing, descending, and traversing ropes?

Figure 7.a: Overall Event Effectiveness Assessment
Event 6: 2-Mile Run

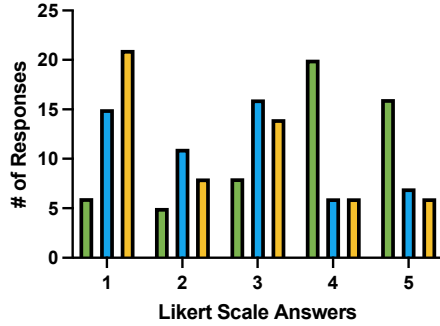


Figure 7.b: O-1/O-2 Event Effectiveness Assessment
Event 6: 2-Mile Run

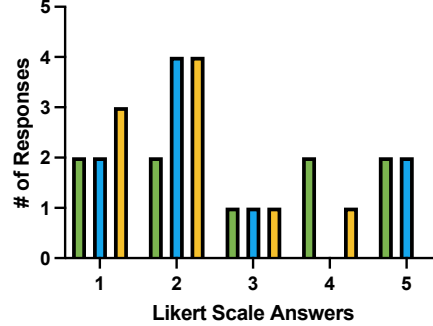


Figure 7.c: O-3 Event Effectiveness Assessment
Event 6: 2-Mile Run

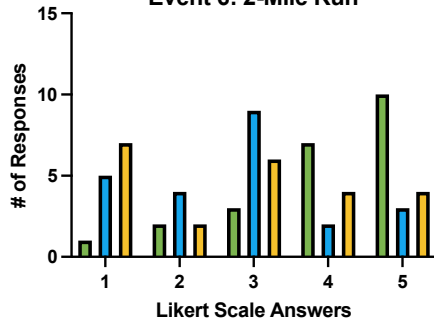


Figure 7.d: O-4/O-5/O-6 Event Effectiveness Assessment
Event 6: 2-Mile Run

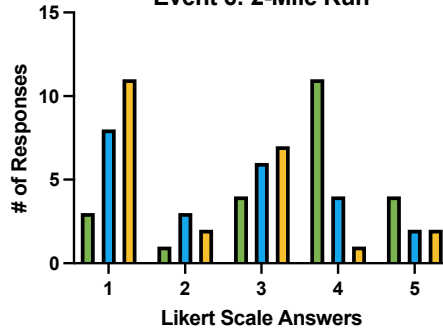
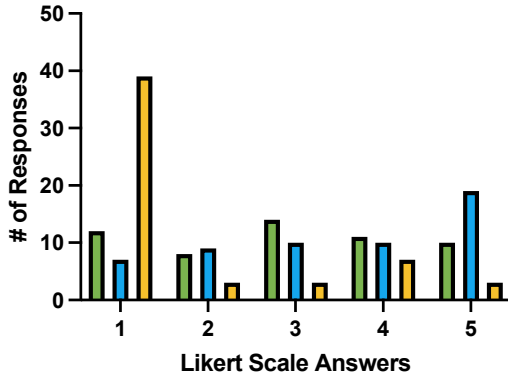


Fig. 7a-d. Display the responses to the following questions about the effectiveness of Event 6: 2-Mile Run. Blue indicates the responses to the question: Is this event is a good test of aerobic fitness? Green indicates the responses to the question: Is this event effective in measuring your ability for dismounted movements, ruck marching or infiltration with combat load? Yellow indicates the responses to the question: Is this event is an appropriate distance to complete after the other five events of the ACFT?

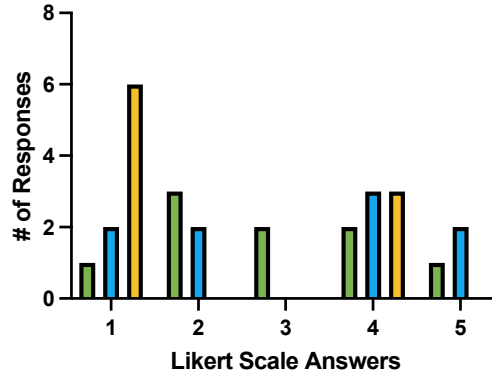
ACFT Event Scoring Assessment

All six events in the ACFT's version 4.0 have a scale that is separated by gender. The original test was designed to be gender agnostic but was later changed as supported in the Literature Review. Each figure in this section examines the perceptions of female US Army Officers on whether or not an event is favored towards one gender, that the current scoring scale is practical with the gender division, and that the scoring scale is better divided between male and female subjects. Figures 8a-d through 13a-d each represent one of the six events, and then a specific population is denoted in the title of each figure. There is the overall assessment, and then the data further broken down by rank in the three subgroups previously discussed.

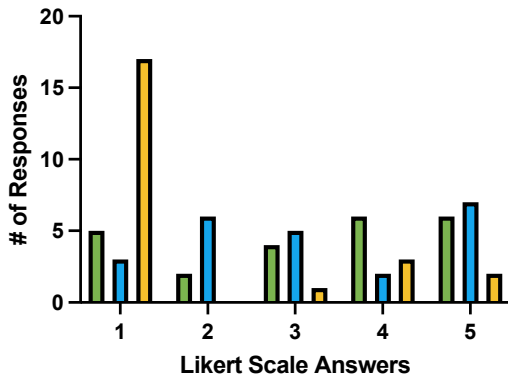
**Figure 8.a: Overall Scoring Assessment
Event 1: 3 Repetition Max Deadlift**



**Figure 8.b: O-1/O-2 Scoring Assessment
Event 1: 3 Repetition Max Deadlift**



**Figure 8.c: O-3 Scoring Assessment
Event 1: 3 Repetition Max Deadlift**



**Figure 8.d: O-4/O-5/O-6 Scoring Assessment
Event 1: 3 Repetition Max Deadlift**

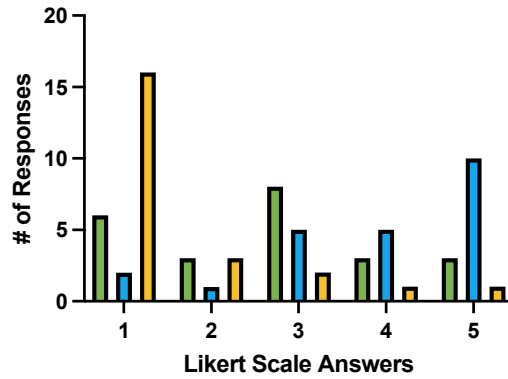
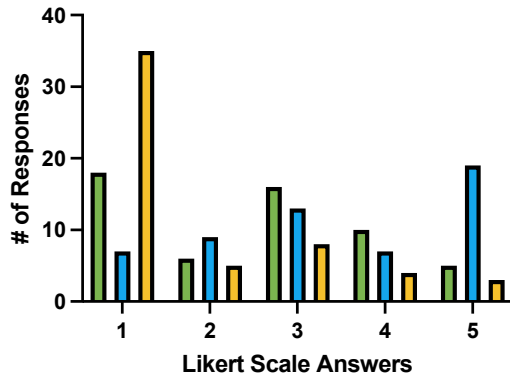
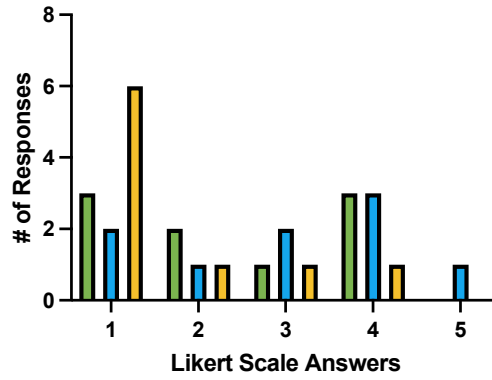


Fig. 8a-d. Display the responses based off the scoring assessment statements for Event 1: 3 Repetition Max Deadlift. Green indicates the responses to the statement: Scoring is fair regardless of gender or age. Blue indicates the responses to the statement: Scoring is biased towards one gender. Yellow indicates the responses to the statement: Scoring would be more effective w/o gender specific standards.

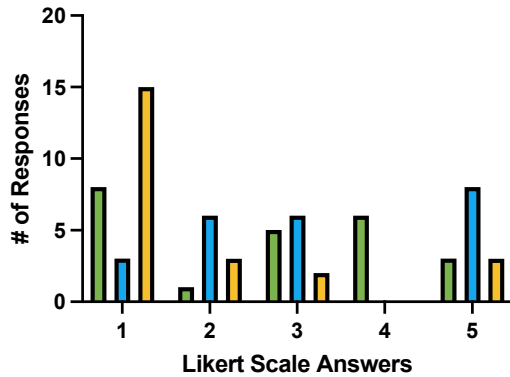
**Figure 9.a: Overall Scoring Assessment
Event 2: Standing Power Throw**



**Figure 9.b: O-1/O-2 Scoring Assessment
Event 2: Standing Power Throw**



**Figure 9.c: O-3 Scoring Assessment
Event 2: Standing Power Throw**



**Figure 9.d: O-4/O-5/O-6 Scoring Assessment
Event 2: Standing Power Throw**

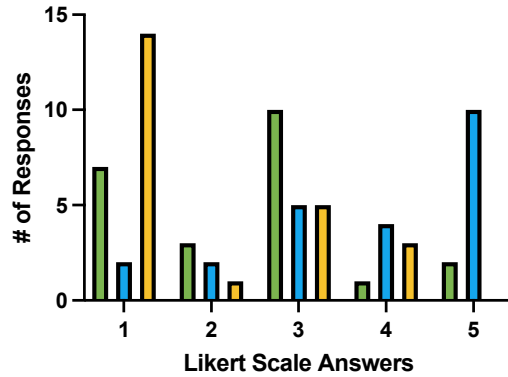
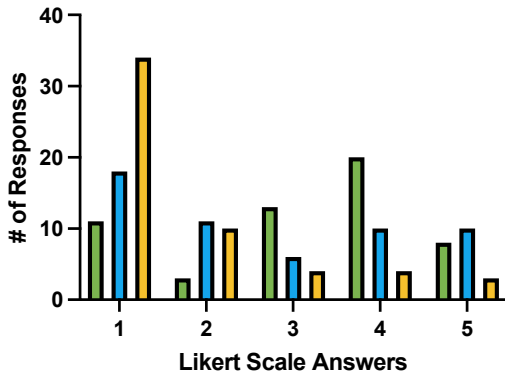
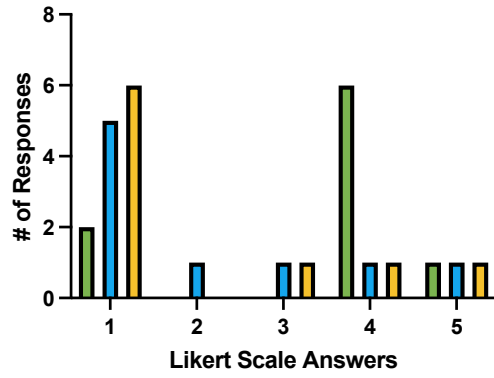


Fig. 9a-d. Display the responses based off the scoring assessment statements for Event 2: Standing Power Throw. Green indicates the responses to the statement: Scoring is fair regardless of gender or age. Blue indicates the responses to the statement: Scoring is biased towards one gender. Yellow indicates the responses to the statement: Scoring would be more effective w/o gender specific standards.

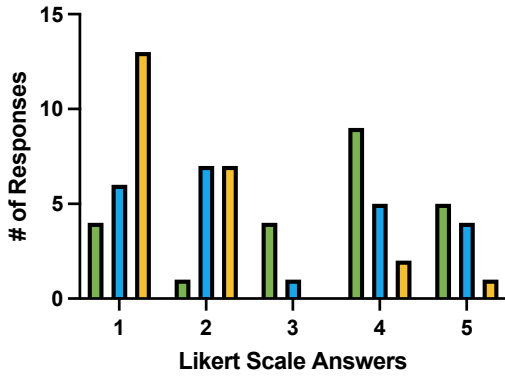
**Figure 10.a: Overall Scoring Assessment
Event 3: Hand Release Push-Ups**



**Figure 10.b: O-1/O-2 Scoring Assessment
Event 3: Hand Release Push-Ups**



**Figure 10.c: O-3 Scoring Assessment
Event 3: Hand Release Push-Ups**



**Figure 10.d: O-4/O-5/O-6 Scoring Assessment
Event 3: Hand Release Push-Ups**

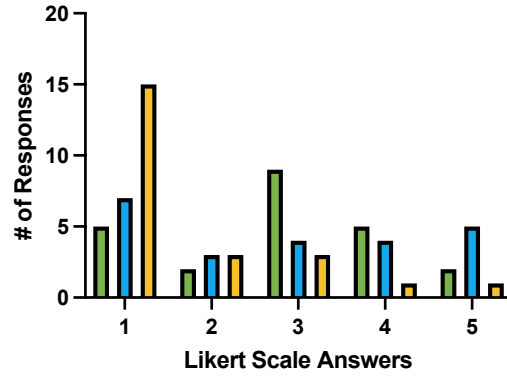
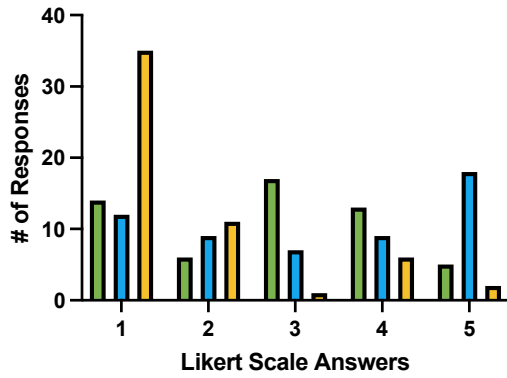
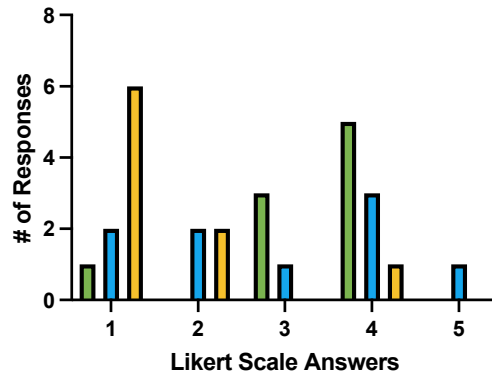


Fig.10a-d. Display the responses based off the scoring assessment statements for Event 3: Hand release Push-Ups. Green indicates the responses to the statement: Scoring is fair regardless of gender or age. Blue indicates the responses to the statement: Scoring is biased towards one gender. Yellow indicates the responses to the statement: Scoring would be more effective w/o gender specific standards.

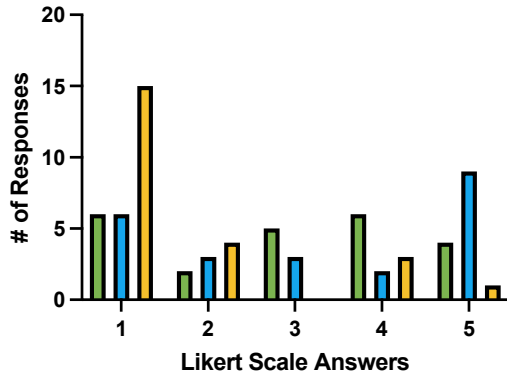
**Figure 11.a: Overall Scoring Assessment
Event 4: Sprint Drag Carry**



**Figure 11.b: O-1/O-2 Scoring Assessment
Event 4: Sprint Drag Carry**



**Figure 11.c: O-3 Scoring Assessment
Event 4: Sprint Drag Carry**



**Figure 11.d: O-4/O-5/O-6 Scoring Assessment
Event 4: Sprint Drag Carry**

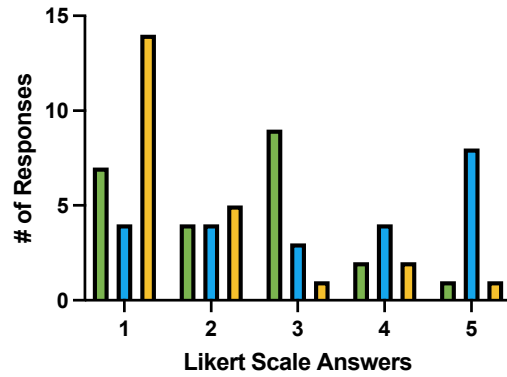
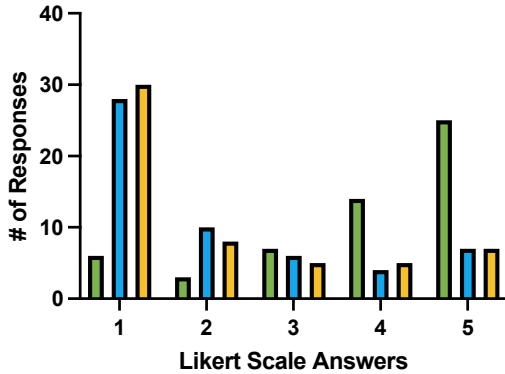
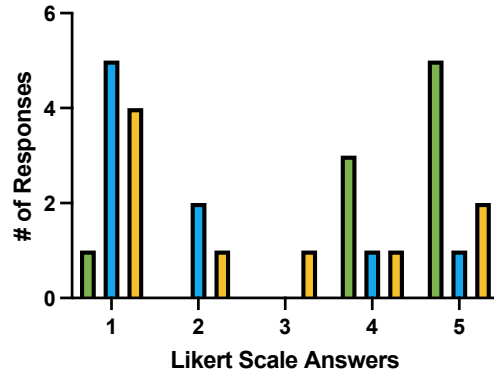


Fig. 11a-d. Display the responses based off the scoring assessment statements for Event 4: Sprint Drag Carry. Green indicates the responses to the statement: Scoring is fair regardless of gender or age. Blue indicates the responses to the statement: Scoring is biased towards one gender. Yellow indicates the responses to the statement: Scoring would be more effective w/o gender specific standards.

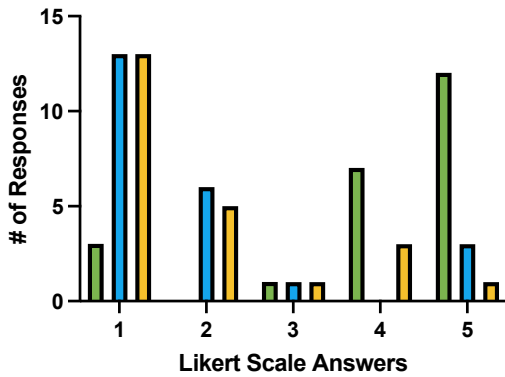
**Figure 12.a: Overall Scoring Assessment
Event 5: Plank**



**Figure 12.b: O-1/O-2 Scoring Assessment
Event 5: Plank**



**Figure 12.c: O-3 Scoring Assessment
Event 5: Plank**



**Figure 12.d: O-4/O-5/O-6 Scoring Assessment
Event 5: Plank**

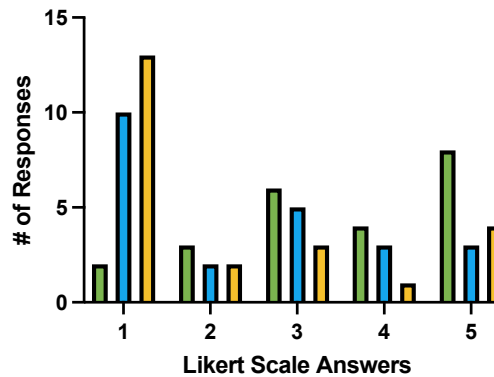
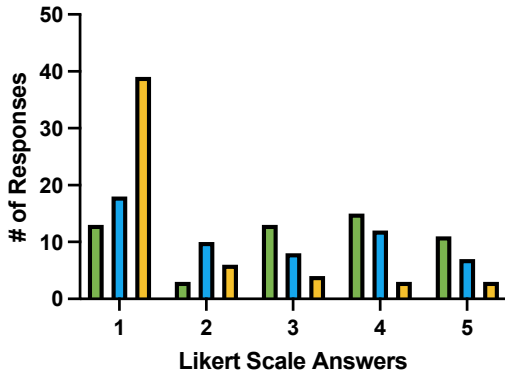
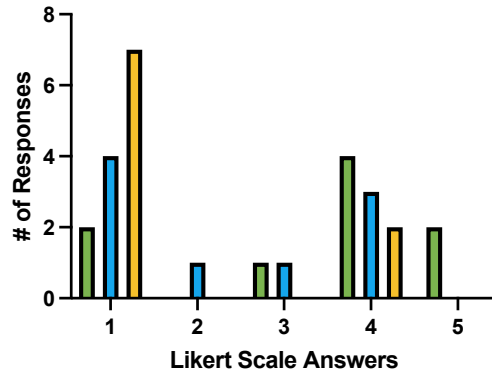


Fig. 12a-d. Display the responses based off the scoring assessment statements for Event 5: Plank. Green indicates the responses to the question: Scoring is fair regardless of gender or age. Blue indicates the responses to the question: Scoring is biased towards one gender. Yellow indicates the responses to the question: Scoring would be more effective w/o gender specific standards.

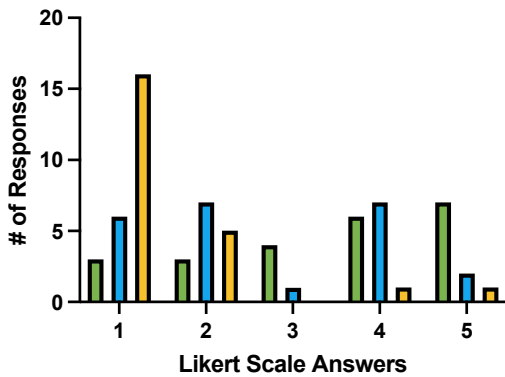
**Figure 13.a: Overall Scoring Assessment
Event 6: 2-Mile Run**



**Figure 13.b: O-1/O-2 Scoring Assessment
Event 6: 2-Mile Run**



**Figure 13.c: O-3 Scoring Assessment
Event 6: 2-Mile Run**



**Figure 13.d: O-4/O-5/O-6 Scoring Assessment
Event 6: 2-Mile Run**

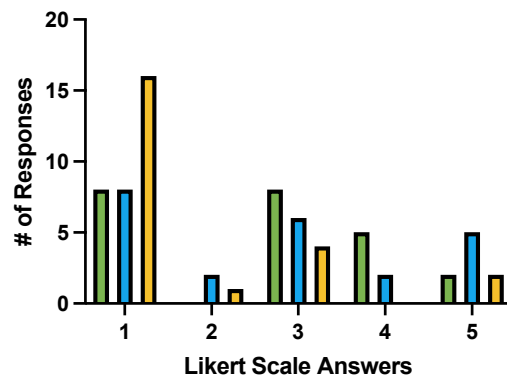


Fig. 13a-d. Display the responses based off the scoring assessment statements for Event 6: 2-Mile Run. Green indicates the responses to the statement: Scoring is fair regardless of gender or age. Blue indicates the responses to the statement: Scoring is biased towards one gender. Yellow indicates the responses to the statement: Scoring would be more effective w/o gender specific standards.

Overall ACFT Assessment

The overall ACFT perspective compares the whole survey population and the three subgroups which are divided by rank. Figures 14a-d compare these four groups and then show the values by group for the four overall ACFT questions asked to survey participants.

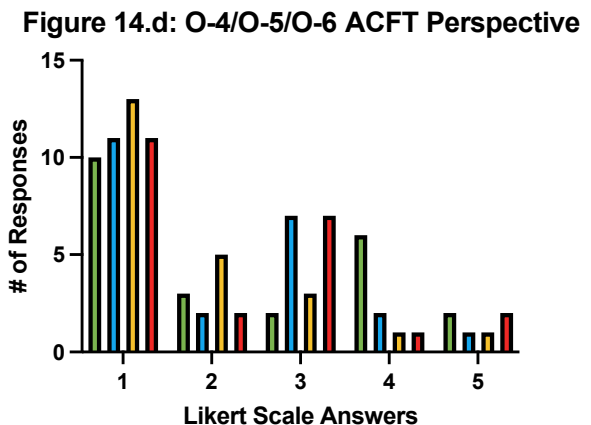
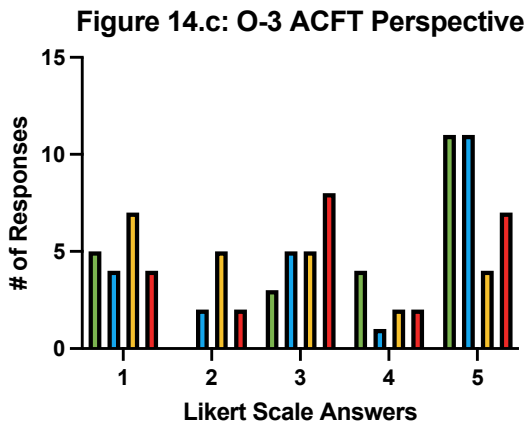
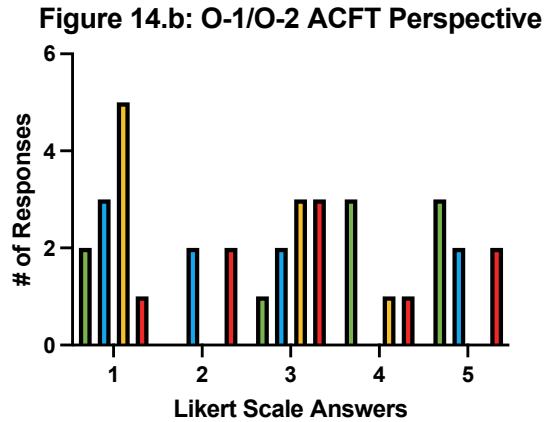
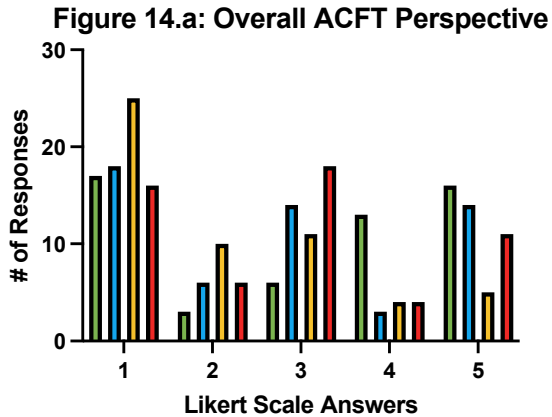


Fig14a-d. Displays the responses to the statements regarding the holistic overview of the ACFT as a physical fitness test. Green indicates the responses to the statement: The ACFT more adequately assesses the preparation of soldiers for the large-scale combat operations. Blue indicates the responses to the statement: The ACFT has positively changed the way Company sized units conduct physical training. Yellow indication the responses to the statement: The ACFT training has reduced the rate of injuries within your organization. Red indicates the responses to the statement: The ACFT has increased the overall physical fitness of your organization.

ACFT Mean and Standard Deviation Data

Tables 3-8 show the Means and Standard Deviations for the whole study population and the three rank-based subgroups. On the left side of each table are all the questions that measured the perceived effectiveness of the ACFT events. On the right side of each table are the three questions related to the perceived scoring assessment in relation to each ACFT event. The mean shows the average score on the Likert scale based on the survey populations. It allows a comparison between the three groups to determine where there are potential trends and deviations. The standard deviation shows how dispersed the data across the survey population in relation to the mean.

Table 3. ACFT Event 1:3-Repetition Maximum Deadlift Mean and Standard Deviation for event effectiveness perceptions and scoring perceptions by study population.

ACFT EVENT 1 Descriptive Statistics	<i>n</i>	M	SD		<i>n</i>	M	SD
Effectiveness Q1 Study Population	55	3.25	1.35	Scoring Q1 Study Population	55	2.98	1.41
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	3.33	1.32	WO1, O1-O2	9	2.89	1.27
CW2-3, O3	23	3.44	1.34	CW2-3, O3	23	3.26	1.51
CW4-CW5, O4-O6	23	3.26	1.21	CW4-CW5, O4-O6	23	2.74	1.36
Effectiveness Q2 Study Population	55	3.87	1.35	Scoring Q2 Study Population	55	3.45	1.44
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	4.56	0.53	WO1, O1-O2	9	3.11	1.62
CW2-3, O3	23	4.09	1.44	CW2-3, O3	23	3.17	1.47
CW4-CW5, O4-O6	23	3.39	1.34	CW4-CW5, O4-O6	23	3.87	1.29
Effectiveness Q3 Study Population	55	2.31	1.30	Scoring Q3 Study Population	55	1.76	1.32
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	2.56	1.33	WO1, O1-O2	9	2.00	1.500
CW2-3, O3	23	2.61	1.31	CW2-3, O3	23	1.83	1.47
CW4-CW5, O4-O6	23	1.91	1.24	CW4-CW5, O4-O6	23	1.61	1.12

Table 3. Effectiveness Q1 indicates the responses to the question: Is this event effective in measuring the ability to extract and carry a casualty on a litter? Effectiveness Q2 indicates the responses to the question: Is this event effective in measuring the ability to lift & carry heavy loads from the ground? Effectiveness Q3 indicates the responses to the question: Is this event effective in preventing lower back injuries? Scoring Q1 indicates the mean & standard deviation to the statement: Scoring is fair regardless of gender or age. Scoring Q2 indicates the mean & standard deviation to the statement: Scoring is biased towards one gender. Scoring Q3 indicates the mean & standard deviation to the statement: Scoring would be more effective w/o gender specific standards.

Table 4. Event 2: Standing Power Throw Mean and Standard Deviation for event effectiveness perceptions and scoring perceptions by study population.

ACFT EVENT 2 Descriptive Statistics	<i>n</i>	M	SD		<i>n</i>	M	SD
Effectiveness Q1 Study Population	55	2.15	1.24	Scoring Q1 Study Population	55	2.60	1.36
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	1.78	1.09	WO1, O1-O2	9	2.44	1.33
CW2-3, O3	23	2.13	1.10	CW2-3, O3	23	2.78	1.51
CW4-CW5, O4-O6	23	2.30	1.43	CW4-CW5, O4-O6	23	2.48	1.24
Effectiveness Q2 Study Population	55	2.25	1.27	Scoring Q2 Study Population	55	3.40	1.44
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	2.33	1.23	WO1, O1-O2	9	3.00	1.41
CW2-3, O3	23	1.91	1.04	CW2-3, O3	23	3.17	1.50
CW4-CW5, O4-O6	23	2.57	1.44	CW4-CW5, O4-O6	23	3.78	1.35
Effectiveness Q3 Study Population	55	1.78	1.15	Scoring Q3 Study Population	55	1.82	1.25
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	1.78	1.09	WO1, O1-O2	9	1.67	1.12
CW2-3, O3	23	1.65	1.11	CW2-3, O3	23	1.83	1.40
CW4-CW5, O4-O6	23	1.91	1.24	CW4-CW5, O4-O6	23	1.87	1.18
Effectiveness Q4 Study Population	55	1.38	0.80				
Sub-Populations							
WO1, O1-O2	9	1.22	0.44				
CW2-3, O3	23	1.30	0.64				
CW4-CW5, O4-O6	23	1.52	1.04				

Table 4. Effectiveness Q1 indicates the responses to the question: Is this event effective in measuring explosive power, balance, and flexibility? Effectiveness Q2 indicates the responses to the question: Is this event effective in measuring the ability to throw equipment over or onto an obstacle? Effectiveness Q3 indicates the responses to the question: Is this event effective in measuring the ability to lift soldiers up or assist them over a wall? Effectiveness Q4 indicates the responses to the question: Is this event effective in measuring the ability to jump across & over obstacles? Scoring Q1 indicates the mean & standard deviation to the statement: Scoring is fair regardless of gender or age. Scoring Q2 indicates the mean & standard deviation to the statement: Scoring is biased towards one gender. Scoring Q3 indicates the mean & standard deviation to the statement: Scoring would be more effective w/o gender specific standards.

Table 5. ACFT Event 3: Hand Release Push-Ups Mean and Standard Deviation for event effectiveness perceptions and scoring perceptions by study population.

ACFT EVENT 3 Descriptive Statistics	<i>n</i>	M	SD		<i>n</i>	M	SD
Effectiveness Q1 Study Population	55	2.35	1.31	Scoring Q1 Study Population	55	3.20	1.34
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	2.11	1.17	WO1, O1-O2	9	3.44	1.42
CW2-3, O3	23	2.65	1.34	CW2-3, O3	23	3.44	1.38
CW4-CW5, O4-O6	23	2.13	1.33	CW4-CW5, O4-O6	23	2.87	1.25
Effectiveness Q2 Study Population	55	2.22	1.24	Scoring Q2 Study Population	55	2.69	1.54
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	2.11	1.27	WO1, O1-O2	9	2.11	1.54
CW2-3, O3	23	2.48	1.24	CW2-3, O3	23	2.74	1.51
CW4-CW5, O4-O6	23	2.00	1.24	CW4-CW5, O4-O6	23	2.87	1.58
Effectiveness Q3 Study Population	55	3.07	1.46	Scoring Q3 Study Population	55	1.76	1.20
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	2.89	1.45	WO1, O1-O2	9	2.00	1.58
CW2-3, O3	23	3.26	1.39	CW2-3, O3	23	1.74	1.14
CW4-CW5, O4-O6	23	2.96	1.58	CW4-CW5, O4-O6	23	1.70	1.15
Effectiveness Q4 Study Population	55	3.16	1.42				
Sub-Populations							
WO1, O1-O2	9	3.22	1.39				
CW2-3, O3	23	3.44	1.31				
CW4-CW5, O4-O6	23	2.87	1.55				

Table 5. ACFT Effectiveness Q1 indicates the responses to the question: Is this event effective in measuring your ability to move obstacles equal to or greater than your body weight? Effectiveness Q2 indicates the responses to the question: Is this event effective in measuring your ability to push an opponent away during man-to-man contact? Effectiveness Q3 indicates the responses to the question: Is this event effective in measuring your ability to get to and from the ground during evasion & maneuvering? Effectiveness Q4 indicates the responses to the question: Is this event effective in preparing you to move from the prone position in order to shoot, take cover or low crawl? Scoring Q1 indicates the mean & standard deviation to the statement: Scoring is fair regardless of gender or age. Scoring Q2 indicates the mean & standard deviation to the statement: Scoring is biased towards one gender. Scoring Q3 indicates the mean & standard deviation to the statement: Scoring would be more effective w/o gender specific standards.

Table 6. ACFT Event 4: Sprint Drag Carry Mean and Standard Deviation for event effectiveness perceptions and scoring perceptions by study population.

ACFT EVENT 4 Descriptive Statistics	<i>n</i>	M	SD		<i>n</i>	M	SD
Effectiveness Q1 Study Population	55	3.69	1.51	Scoring Q1 Study Population	55	2.80	1.31
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	4.22	1.30	WO1, O1-O2	9	3.33	1.00
CW2-3, O3	23	3.91	1.51	CW2-3, O3	23	3.00	1.48
CW4-CW5, O4-O6	23	3.23	1.54	CW4-CW5, O4-O6	23	2.39	1.16
Effectiveness Q2 Study Population	55	2.89	1.56	Scoring Q2 Study Population	55	3.22	1.58
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	3.00	1.41	WO1, O1-O2	9	2.89	1.45
CW2-3, O3	23	3.13	1.54	CW2-3, O3	23	3.22	1.70
CW4-CW5, O4-O6	23	2.61	1.64	CW4-CW5, O4-O6	23	3.35	1.56
Effectiveness Q3 Study Population	55	2.11	1.30	Scoring Q3 Study Population	55	1.71	1.17
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	2.00	1.00	WO1, O1-O2	9	1.56	1.01
CW2-3, O3	23	2.65	1.53	CW2-3, O3	23	1.74	1.25
CW4-CW5, O4-O6	23	1.61	0.94	CW4-CW5, O4-O6	23	1.74	1.18
Effectiveness Q4 Study Population	55	3.25	1.40				
Sub-Populations							
WO1, O1-O2	9	3.67	0.87				
CW2-3, O3	23	3.61	1.44				
CW4-CW5, O4-O6	23	2.74	1.42				
Effectiveness Q5 Study Population	55	3.29	1.46				
Sub-Populations							
WO1, O1-O2	9	3.78	0.83				
CW2-3, O3	23	3.57	1.56				
CW4-CW5, O4-O6	23	2.83	1.47				

Table 6. Effectiveness Q1 indicates the responses to the question: Is this event a good test of strength, agility, muscular endurance & anaerobic capacity? Effectiveness Q2 indicates the responses to the question: Is this event effective in measuring your ability to react faster to direct or indirect fire? Effectiveness Q3 indicates the responses to the question: Is this event effective in measuring your ability to build hasty fighting positions? Effectiveness Q4 indicates the responses to the question: Is this event effective in measuring your ability to extract and carry a casualty? Effectiveness Q5 indicates the responses to the question: Is this event effective in measuring your ability to carry ammunition to a fight position from a vehicle? Scoring Q1 indicates the mean & standard deviation to the statement: Scoring is fair regardless of gender or age. Scoring Q2 indicates the mean & standard deviation to the statement: Scoring is biased towards one gender.

Scoring Q3 indicates the mean & standard deviation to the statement: Scoring would be more effective w/o gender specific standards.

Table 7. Event 5: Plank Mean and Standard Deviation for event effectiveness perceptions and scoring perceptions by study population.

ACFT EVENT 5	n	M	SD		n	M	SD
Descriptive Statistics							
Effectiveness Q1 Study Population	55	3.02	1.34	Scoring Q1 Study Population	55	3.89	1.34
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	3.67	1.23	WO1, O1-O2	9	4.22	1.30
CW2-3, O3	23	3.35	1.23	CW2-3, O3	23	4.09	1.35
CW4-CW5, O4-O6	23	2.44	1.31	CW4-CW5, O4-O6	23	3.57	1.34
Effectiveness Q2 Study Population	55	2.80	1.51	Scoring Q2 Study Population	55	2.13	1.44
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	3.00	1.41	WO1, O1-O2	9	2.00	1.50
CW2-3, O3	23	3.22	1.54	CW2-3, O3	23	1.87	1.36
CW4-CW5, O4-O6	23	2.30	1.43	CW4-CW5, O4-O6	23	2.44	1.50
Effectiveness Q3 Study Population	55	1.85	1.13	Scoring Q3 Study Population	55	2.11	1.47
Sub-Populations				Sub-Populations			
WO1, O1-O2	9	2.00	1.23	WO1, O1-O2	9	2.56	1.74
CW2-3, O3	23	1.87	1.22	CW2-3, O3	23	1.87	1.25
CW4-CW5, O4-O6	23	1.78	1.04	CW4-CW5, O4-O6	23	2.17	1.59
Effectiveness Q4 Study Population	55	1.87	1.06				
Sub-Populations							
WO1, O1-O2	9	1.89	1.27				
CW2-3, O3	23	1.96	1.11				
CW4-CW5, O4-O6	23	1.78	0.95				

Table 7. Effectiveness Q1 indicates the responses to the question: Is this event is a good test of strength, agility, muscular endurance & anaerobic capacity? Effectiveness Q2 indicates the responses to the question: Is this event assists in preparing you for load bearing exercises and avoiding back injuries? Effectiveness Q3 indicates the responses to the question: Is this event prepares you to move over walls & obstacles? Effectiveness Q4 indicates the responses to the question: Does this event prepares you for climbing, descending, and traversing ropes? Scoring Q1 indicates the mean & standard deviation to the statement: Scoring is fair regardless of gender or age. Scoring Q2 indicates the mean & standard deviation to the statement: Scoring is biased towards one gender. Scoring Q3 indicates the mean & standard deviation to the statement: Scoring would be more effective w/o gender specific standards.

Table 8. Event 6:2-Mile Run Mean and Standard Deviation for event effectiveness perceptions and scoring perceptions by study population.

ACFT EVENT 6 Descriptive Statistics	<i>n</i>	M	SD		<i>n</i>	M	SD
Effectiveness Q1 Study Population Sub-Populations	55	3.64	1.30	Scoring Q1 Study Population Sub-Populations	55	3.15	1.45
WO1, O1-O2	9	3.00	1.58	WO1, O1-O2	9	3.44	1.51
CW2-3, O3	23	4.00	1.17	CW2-3, O3	23	3.48	1.41
CW4-CW5, O4-O6	23	3.52	1.24	CW4-CW5, O4-O6	23	2.70	1.40
Effectiveness Q2 Study Population Sub-Populations	55	2.62	1.34	Scoring Q2 Study Population Sub-Populations	55	2.64	1.46
WO1, O1-O2	9	2.56	1.51	WO1, O1-O2	9	2.33	1.41
CW2-3, O3	23	2.74	1.29	CW2-3, O3	23	2.65	1.40
CW4-CW5, O4-O6	23	2.52	1.38	CW4-CW5, O4-O6	23	2.74	1.57
Effectiveness Q3 Study Population Sub-Populations	55	2.42	1.38	Scoring Q3 Study Population Sub-Populations	55	1.64	1.18
WO1, O1-O2	9	2.00	1.00	WO1, O1-O2	9	1.67	1.32
CW2-3, O3	23	2.83	1.50	CW2-3, O3	23	1.52	1.04
CW4-CW5, O4-O6	23	2.17	1.34	CW4-CW5, O4-O6	23	1.74	1.9

Table 8. Effectiveness Q1 indicates the responses to the question: Is this event is a good test of aerobic fitness? Effectiveness Q2 indicates the responses to the question: Is this event effective in measuring your ability for dismounted movements, ruck marching or infiltration with combat load? Effectiveness Q3 indicates the responses to the question: Is this event is an appropriate distance to complete after the other five events of the ACFT? Scoring Q1 indicates the mean & standard deviation to the statement: Scoring is fair regardless of gender or age. Scoring Q2 indicates the mean & standard deviation to the statement: Scoring is biased towards one gender. Scoring Q3 indicates the mean & standard deviation to the statement: Scoring would be more effective w/o gender specific standards.

Mean differences between ranks (O-1 to O-2, O-3, O-4 thru O-6) for effectiveness for each of the events was analyzed. Significant differences in the Event 4, question 3, (“Is this event effective in measuring your ability to build hasty fighting positions?”) and in the Event 5, question 1, (“Is this event is a good test of strength, agility, muscular endurance & anaerobic capacity?”) (Figure 15a-b).

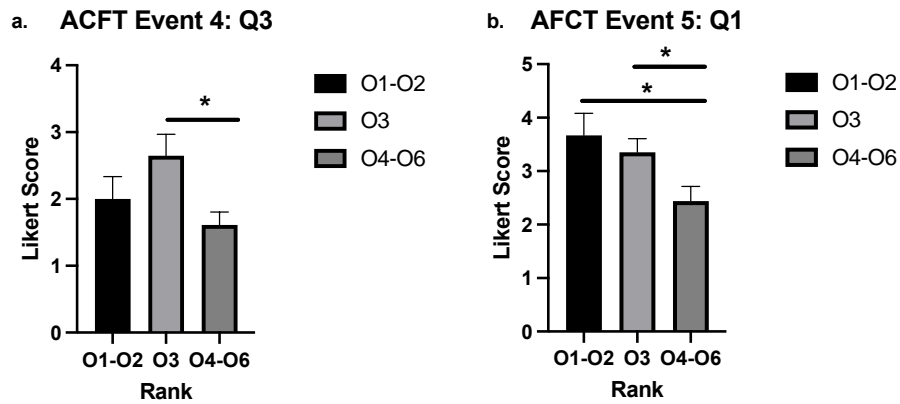


Fig 15a-b. Significant differences registered by the O-4 thru O-6 population for Event 4, the Sprint Drag, Carry Effectiveness Question 3: Is this event effective in measuring your ability to build hasty fighting positions? and Event 5. The Plank, Effectiveness Question 1: Is this event is a good test of strength, agility, muscular endurance & anaerobic capacity?

Discussion

In 2005, during the height of the Global War on Terror (GWOT),²⁵ the US Army determined that a change had to be made to the physical fitness of incoming soldiers. By 2011, the US Army began designing and testing a new physical fitness test to replace the current test that had been in effect for almost thirty years. The intent was to create a more holistic test that retained gender and age-neutral scoring. This would not be the case as the test design began to show significant passing rate differences between male and female service members. The US Senate used the RAND Corporation independent study³⁷ to redesign the scoring and events to ensure a fairer test. But is the test practical, and is the scoring system fair between the genders now that it is gender-separate? This study was designed to gauge the perspectives of female US Army Officers, an underrepresented population in the US Army, and from the initial testing groups when the ACFT was intended. The assertion is that any form of physical assessment is unfair when the events are not based on physiological differences between men and women, particularly regarding weight-bearing exercises, and that a gender-specific scoring scale must be used. The assumption is also present that many of the events, in particular the plank and the standing power throw, do not effectively test the designed physical skills required.

Conclusions: Overall ACFT Assessment

As discussed in the literature, the primary reasons for creating the ACFT were to refocus the force on conducting large-scale combat operations (LSCOs) and to decrease the number of injuries in the force. It was also designed to provide a more holistic approach to fitness compared to the test from the thirty years prior, the ACFT. The cohort of participants who completed the study are rather inexperienced in terms of time in the Army and time as an Officer. This group had an average time in the Army and as an Officer of 4 to 10 years. When the Officer's

perceptions of the AFCT are broken down by rank, there are some interesting trends (Figure 14a-d). As shown in Figure 14d, those at a O-4 thru O-6 ranks with time in the Army between 12 to 18+ years, showed that only 39% rated the question, “The ACFT more adequately assesses the preparation of soldiers for the large-scale combat operations,” positively. It is relevant to note that the more senior group of participants does not see the ACFT overall preparing the force for LSCOs compared to the middle group, seeing that it does. This is most likely due to the experiences of the senior group, which had all served during the Global War on Terror (GWOT) era, where the Army was focused on smaller-scale tactics.³¹ Only the most senior members in the lower rank of O-3 would have served during this time. This means that these two groups of differing ranks have had different focuses on how military operations should be conducted and experience levels with the ACFT and APFT.

Another interesting trend when analyzing the Officer’s overall perception of the AFCT was in response to the question, “The ACFT training has reduced the rate of injuries within your organization.” As shown in Figure 14a, 64% of total survey participants feel that the ACFT does not reduce injury rates in their organizations. Figures 14b-d show that participants strongly disagreed that the ACFT reduced injury rates across each of the three groups. Event 1 had performance measures deliberately put in place to reduce injury rates, such as using a hexagonal bar versus a straight bar. As seen in Figure 3a, 33 participants disagree that Event 1 prevents lower back injuries and as seen in Figure 14a, 35 participants disagree that the ACFT prevents injury. These two responses correlate. It appears in Figure 14d that most O-4 thru O-6 participants disagreed with all four questions, whereas the O-3 rank Officers and the O-1 and O-2 ranked Officers (Figures 14b and 14c) see a more evenly distributed response. Since the

population for the O-4 thru O-6 Officers (Figure 14d) was 23 of the 55 participants, it also stands that their responses mirror the overall responses depicted in Figure 14a.

Conclusions: ACFT Event Effectiveness

When examining the event effectiveness for the AFCT, Event 1, the 3-Repitition Maximum Deadlift, had the highest effectiveness means across all six events (Table 3) with a low of 3.25 and a high of 3.87 with an average of 3.56 as an effectiveness score. The data trends show the perception was mostly neutral but trended positive. This equated to 54% of the participants believing that the event effectively measured the ability to extract and carry out a casualty, with 72% believing that the event effectively measured the ability to lift heavy loads from the ground. However, 60% of survey participants firmly believed that this event led to increased lower back injuries. The event was designed to utilize a hexagonal bar instead of a straight bar, with the justification that this would decrease the injury rate.

The event with the lowest mean effectiveness rate was Event 2, the Standing Power Throw. As seen in Table 4, a mean response of 1.38 was registered for event's effectiveness in measuring the ability to jump or move across obstacles for all ranks. With a mean score of 1 or 2 this means that the perception of effectiveness trended negatively or was viewed as an effective test. In the O-1 to O-2 population, 88% of participants answered with a mean score of 1.22, which means that this population strongly disagree with the effectiveness of the event, the lowest of the three populations. The other two more senior populations with time in the Army and as Officers gave it a slightly higher score, with a mean of 1.30 for the O-3 population and 1.52 for the O-4 thru O-6 population (see Table 4). The standard deviation was 0.80, which overwhelmingly indicates that the event was commonly rated as ineffective.

Only two of the six events in the ACFT are similar to those in the previous physical fitness assessment, the APFT. These events are the 2 minutes of hand release push-ups (Event 3), which replaced 2 minutes of regular push-ups, and the 2-mile run (Event 6), which remains unchanged from the APFT. In Event 3, the Hand Release Push-up, there were four associated effectiveness questions (see Table 5), with an average score of 2.7, this trends towards a neutral perception with the most significant deviation between the scores of the O-3 and O-4 thru O-6 populations. The 4 effectiveness questions for the hand release push-up each had an average mean deviation of 0.4675 between these two populations. This can be attributed to age and the potential for a physical decline in upper body strength, where the O-4 thru O-6 population consistently disagreed that this event was effective compared to the O-3 population. The O-4 thru O-6 population has an increased time in the Army as Officer by an average of 6 years compared to the O-3 population, as seen in Table 1.

Event 4 is the Spring Drag Carry, which had a mean of 3.05 across all five effectiveness questions (see Table 6) and the highest single question score of 3.69 for effectiveness question 1, measuring if the event is effective for measuring agility, muscular endurance, and anaerobic capacity. This means that in general all populations felt neutrally about the events and that question 1 trended to agree that the event was perceived as effective. In this question, the two groups on opposing ends were the O-1 & O-2 and O-4 thru O-6 populations, with the highest and lowest ratings of 4.22 and 3.23, which are also graphically depicted as opposing scores (Figures 7b and 7d). The most junior population with a rating 4.22 is demonstrating agree to strongly agree with the score of 3.23 for the more senior population rating the effectiveness neutral with only a quarter of the population agreeing. As with Event 3 (hand release push-ups), time in the Army is the most significant difference between these two populations, and most likely, the

perceived difficulty of this event would be expected of the more senior officers compared to the youngest population in terms of age and experience. It is a more extended duration event than all the other events anaerobically with the multiple sprinting sections, a 90lb sled drag, and a 80 lb kettlebell carry.

The Plank, Event 5 in the ACFT, was the last event added to the test, which replaced the Leg Tuck was replaced as the event from the earliest iterations of the test. The Leg Tuck was the event that was causing the highest rate of female failures, which drove the data to support the Congressional memorandum to halt the implementation of the ACFT in 2020. With that historical context noted, this event needs a high pass rate but should not diminish the effectiveness of measuring core strength or detract from the desired intents. Out of all six events, the plank had the lowest overall effectiveness score of 2.385, which shows a trend of either strongly disagreeing or disagreeing with all effectiveness questions (Figure 6a). This means that 130 of 220 indicted event ineffectiveness. Questions 3 and 4 ask the participants about the effectiveness of the plank to gauge their ability to conduct dynamic movements. In these two questions, 76% and 74% of participants respectively gave a rating of strongly disapprove or disapprove (Table 7), with 74% of the O-4 through O-6 population giving a strongly disapprove and disapprove rating (Figure 6d) and an average of 1.78 for both questions (Table 7).

The final event of the ACFT, the 2-mile run, remains the only event not changed or modified from the previous test, the APFT. The 2-mile run is conducted at the end of the six-event ACFT, similar to how it was the final event in the three-event APFT. The data responses for this test showed an average assessment of 3.64 for effectiveness as an aerobic test, which is neutral and trending towards agree, but with the O-3 population average registering a score of 4.00, which agrees with the effectiveness (Table 8). Question 3 (Is this event an appropriate

distance to complete after the other five events of the ACFT?) registered the lowest average answer of 2.42, which disagreed a neutral perception with the O-1 to O-2 population responding with a 2.00 of a firm disagree, the lowest across the three populations (Figure 7c).

Conclusions: ACFT Scoring

Unlike in the Event Effectiveness portion of the survey, all six events had the same three statements asking about concerns on scoring the AFCT:

- Scoring is fair regardless of gender or age.
- Scoring is biased towards one gender.
- Scoring would be more effective without gender-specific standards.

In Event 1, the 3-Repitition Maximum Deadlift, 71% of the participants strongly disagreed that the event scoring would be more effective without gender-specific standards (Figure 8a). This result correlates with the physiological differences and typical body weight between males and females, meaning that the expectation is that a female will typically weigh less than a male and then, in turn, would normally be able to lift less weight from the ground. Based on the perception values (Table 3), Event 1 registered the second highest score, 3.45, and was neutral in agreeing with Statement 2, “Scoring is biased towards one gender.”

Event 2, the Standing Power Throw, like many events, showed a trend of strongly disagreeing to disagree with the assessment that the scoring would be fair regardless of gender (Figures 9a-d). The O-4 through O-6 population registered a 3.78 (Table 4) neutral to agree that the event scoring is based on gender. The youngest of the three populations, O-1 to O-2, registered a value of 3.00 (Table 4), a neutral perception. This has been a continued trend that these two populations separated by the most significant gap in both age and time in the US Army is noted.

Event 3, the Hand Release Push-Ups, and Event 6, the 2-mile Run, are the two events that are only slightly adjusted from the APFT to the ACFT. For Event 3, the only scoring assessment of note was that when asked if the scoring was fair regardless of age or gender, only the O-4 through O-6 population registered a mean score of disagree, 2.87 (Table 5), with the other two populations' means scoring of 3.44, neutral to agree. Event 6 remains utterly unmodified from the APFT. This event had the lowest mean, 1.64 (Table 8), of any of the events, disagreeing with the statement, "Scoring would be more effective without gender-specific standards." Like event 1, the perception here makes sense based on the standard ability of males versus females regarding the speed and pace of an aerobic event.

In Sprint, Drag Carry, and Event 4, all three populations registered mean values that aligned with strongly disagreed with the statement "Scoring would be more effective without gender-specific standards": 1.56, 1.74, and 1.74 (Table 6), respectively. When asked about statements 1 and 2 ("Scoring is fair regardless of gender or age" and "Scoring is biased towards one gender"), the O-1 to O-2 population was the only one to register 0 strongly agree responses (Figure 11b); however, for question 1, a majority of participants in this population registered an agree to statement 1, so the value for this statement, 3.33 appears neutral even though there were more agrees compared to the other two populations, leaving the mean value to be, perceived as neutral (Table 6)

Event 5, the newest of the events, the Plank, was explicitly designed so that more female service members could pass the event. The current scoring scale for this event is the only one that does not differentiate between genders. The results from the survey showed that the O-3 population was the only one with mean values indicating strongly disagree scores, 1.87 to statements 2 and 3 ("Scoring is biased towards one gender" and "Scoring would be more

effective without gender-specific standards”). These values are contradictory because, in statement 1 (“Scoring is fair regardless of gender or age”), the O-3 population agreed, 4.09 (Table 7), that scoring was fair regardless of gender but strongly disagreed, 1.87 (Table 7) with statement 3, that the scoring would be more effective without gender-specific standards. This deviation could be caused by confusion with the wording of the statements but causes a slight deviation in the overall mean values (Figure 12c) of the scoring assessment of this event. However, the results for this event mirror the other events in which a gender scale is necessary. Otherwise, the event is biased towards one gender.

Overall Take-aways

Of the twenty-three questions related to event effectiveness, only nine of the effectiveness questions showed overall means that were neutral to positive, with scores over 3.00. The two events that demonstrated the three highest effectiveness questions answers that were neutral to agree were Event 1, question 2 (“Is this event effective in measuring the ability to lift & carry heavy loads from the ground?”), 3.87 (Table 3), Event 4, question 1 (“Is this event a good test of strength, agility, muscular endurance & anaerobic capacity?”), 3.69 (Table 6), and Event 6, question 1 (“Is this event is a good test of aerobic fitness?”), 3.64 (Table 8). The four lowest scores, perceived as strongly disagree, were exclusive to two events. Event 2, question 4 (“Is this event effective in measuring the ability to jump across & over obstacles?”), 1.38, and question 3 (“Is this event effective in measuring the ability to lift soldiers up or assist them over a wall?”), 1.78. (Table 4) and Event 5, question 3 (“Is this event prepares you to move over walls & obstacles?”), 1.85, and question 4 (“Does this event prepares you for climbing, descending, and traversing ropes?”), 1.87 (Table 7). Throughout the survey, there were deviations in specific populations.

Still, most of the overall values showed no significant differences based on the results with numerical data from the Likert scale. Only two measured deviations (Figure 15a-b), which indicate the two questions that when a one-way ANOVA test was conducted, showed significant differences in Event 4, question 3, “Is this event effective in measuring your ability to build hasty fighting positions?” and in the Event 5, question 1, “Is this event is a good test of strength, agility, muscular endurance & anaerobic capacity? (Figure 15a-b). The p-value for Figure 15a is 0.0215, and the p-value for Figure 15b is 0.0173. These p values less than 0.05 are indicators of strong significance against the questions of effectiveness being asked with the relationship between the respective ranks. In addition, Overwhelmingly, the participants in the survey disagree that the scoring for any of the events is fair regardless or practical without gender-specific standards and agree that the scoring is biased towards one gender or the other.

Limitations

This Project and Report faced a few limitations. The main limitations were the narrow scope of the population being surveyed and the survey length. With only 15.7% of the US Army active-duty component consisting of females,⁴⁰ reaching a large group of voluntary participants virtually was limiting. However, females make up 20% of the Officer Corps,⁴⁰ so the scope was narrowed to focus on the target population of female US Army Officers. This population was reached through various social media platforms. It was accessible and reliable due to the pre-vetting of the participants as the moderators of the Facebook group, in which the participants were recruited, have confirmed their military service, officer rank, and gender as access to these groups. The demographic data collected identified another major limitation of this study: the race and ethnicity breakdown of the survey needed to mirror the race and ethnicity breakdown of the active component female Officer population and, in the end, did not. Figures 18a and 18b

graphically depict the demographics by race and ethnicity. In the present survey, the only registered responses to the ethnicities and races were White, not Hispanic, Hispanic, Asian, and Multiracial/Other. Our survey had a more significant proportion of White, not Hispanic participants, 76%, compared to the 58% serving in the US Army. 18% of female Officers in the Active component of the US Army register as Black, not Hispanic, the present survey registered 0% in this category. Another limitation was survey completion. 97 participants initially started the survey, but 42 participants failed to complete the survey and did not register scores. After re-examining the questions, the survey could have focused the event effectiveness section questions purely on the physical aspects and omitted the questions around the feasibility of training and testing the events, as that ended up not being the focus of this research.

Figure 18.a: US Army 2022 Demographics for Female Officers by Race/Ethnicity

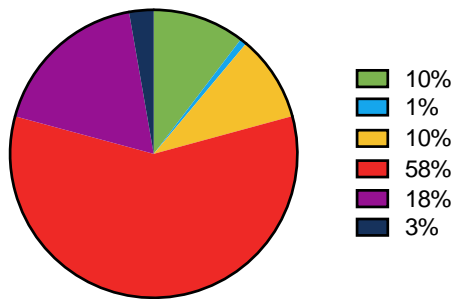


Figure 18.b: ACFT Survey Demographics of US Army Female Officers by Race/Ethnicity

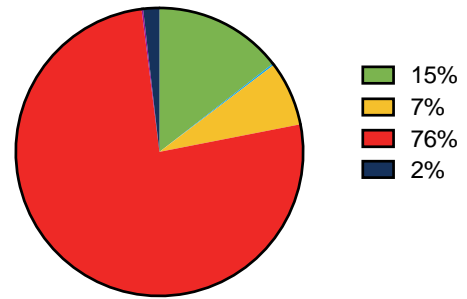


Fig 18a-b. Displays the Races/Ethnicities of US Army Female Officers and the Virginia Tech Survey US Army Female Officers. Green indicates Hispanic. Light Blue indicates American Indian or Alaskan Native. Yellow indicates Asian or Pacific Islander. Red indicates White, not Hispanic. Purple indicates Black, not Hispanic. Dark Blue indicates Other.

Dissemination

The uniqueness of this survey and the surveyed population lend itself to multiple dissemination opportunities. The results of this survey can be presented through US military information and education outreach in conjunction with the Army Women’s Foundation, which could potentially provide a platform for a virtual seminar or presentation during a leadership

development opportunity. This Project and Report can be used as a resource for articles related to female-based issues; many of these articles are published through military leadership-based websites. The problems and results identified in this project and report also provide insight into lessons learned when developing future fitness tests. Approaching dissemination from this angle, there could be the potential for publication or presentation through the Journal of Military Learning (JML) or other research-based publications; Military Medicine, Military Medical Research, and BMJ Military Health.

Future Directions for this Project and Report

This survey was initially developed with a specific population to increase representation for an underrepresented group in the beta testing. However, after reviewing the results of this survey, the scope could be expanded to track comparable population data, such as male US Army Officers, or to develop it further to include enlisted ranks. According to the FY24 National Defense Appropriation Act (NDAA) that will be passed into law within the next twelve months, there is the potential for continued changes to the scoring scale, including reverting to a MOS-based scale for close combat operations specialties or the potential to relook having a gender-neutral scoring system. The US Army was tasked to report back at the end of the fiscal year to Congress with recommended ways forward for any potential refinements based on the data now provided that the test is an official test for record and required biannually for active-duty soldiers.

This survey also continues to spark discussion about other potential research avenues for which this data could be a base. One of the possible research areas to examine is the potential for a weight-based scoring system instead of gender-based. One example is shown in the ACFT 1.0 with the gender-neutral and age-neutral scoring. On the original scale, 340 lb for the 3-repetition

max deadlift was worth up to 100 points. A 120 lb soldier would be lifting 283% of their body weight, while a 220 lb soldier would only be lifting 142% of their body. The Standing Power favors taller Soldiers, and the Spring, Drag, Carry has similar limitations based on the 90 lb sled drag and 80 lb kettlebell carry. If the US Army is going to relook at the scoring scale again, perhaps a study on a weight-based scale is more feasible and realistic instead of based on age and gender. Lastly, examining the injury rates in different populations would be possible now that the test has been required since October 2022. Studies to determine if there have been decreased rates in training and testing are just two examples of potential branches for which this survey and future ones could act as the base.

References

1. Knapik, J. J., & East, W. B. (2014). *History of United States Army physical fitness and physical readiness training*. U.S. Army Medical Department journal, 5–19.
3. Grier, T., Canham-Chervak, M., Anderson, M. K., Bushman, T. T., & Jones, B. H. (2015). *The effects of cross-training on fitness and injury in women*. U.S. Army Medical Department journal, 33–41.
2. U.S. Government Publishing Office. (2015). Carl Levin and Howard P. ‘‘Buck’’ McKeon National Defense Authorization Act for Fiscal Year 2015. GovInfo. Retrieved from <https://www.govinfo.gov/app/details/COMPS-11977>
4. East WB. *A Historical Review and Analysis of Army Physical Readiness Training and Assessment*. (2013) Fort Leavenworth, KS: Combat Studies Institute Press.
5. *General Order No. 44*. Washington, DC: US War Department. March 1, 1906.
6. Raycroft JE. *Mass Physical Training for Use in the Army and the Reserve Officers’ Training Corps*. Washington, DC: United States Infantry Association;1920.
7. McCloy CH. *The measurement of general motor capacity and general motor ability*. Res Q. 1934; 5:46-61.
8. *War Department Field Manual 21-20: Basic Field Manual Physical Training*. Washington, DC: US War Department; March 6, 1941.
9. Morden BJ. *The Women’s Army Corps, 1945-1978*. Washington, DC: US Government Printing Office; 1990.
10. *War Department Field Manual 35-20: Physical Training*. Washington, DC: US War Department; July 15, 1943.
11. *War Department Pamphlet 21-9: Physical Conditioning*. Washington, DC: US War Department; May 1, 1944.
12. *Technical Manual 21-200: Physical Conditioning*. Washington, DC: US Dept of the Army; December 31, 1957 [obsolete].
13. *Technical Manual 21-200: Physical Conditioning [change 2]*. Washington DC: US Dept of the Army; July 25, 1961 [obsolete]
14. *Department of the Army Pamphlet 21-1: Physical Fitness Program for Specialists and Staff Personnel*. Washington, DC: US Dept of the Army; January 7, 1963 [obsolete].

15. *Department of the Army Pamphlet 21-2: Physical Fitness Program for Female Special Staff Personnel and TD Organizations*. Washington, DC: US Dept of the Army; January 7, 1963 [obsolete].
16. *Army Regulation 600-9: Army Physical Fitness Program*. Washington, DC: US Dept of the Army; January 5, 1965 [obsolete].
17. *Field Manual 21-20: Physical Readiness Training*. Washington, DC: US Dept of the Army; January 31, 1969 [obsolete].
18. *Field Manual 21-20: Physical Readiness Training*. Washington, DC: US Dept of the Army; March 30, 1973 [obsolete].
19. Cousins GF. *A factor analysis of selected wartime fitness tests*. Res Q. 1955; 26:277-288.
20. *Field Manual 21-20: Physical Readiness Training*. Washington, DC: US Dept of the Army; October 31, 1980 [obsolete].
21. Department of the Army, AR 350-15 *The Army Physical Fitness Program* (Washington, DC: US Government Publishing Office, 1982), 1; “Soldier Physical Fitness Center: An Historical Review”
22. *Field Manual 21-20: Physical Fitness Training [change 1]*. Washington, DC: US Dept of the Army; June 23, 1986 [obsolete].
23. *Field Manual 21-20: Physical Fitness Training*. Washington, DC: US Dept of the Army; August 28, 1985 [obsolete].
24. *Field Manual 21-20: Physical Fitness Training*. Washington, DC: US Dept of the Army; September 30, 1992 [obsolete].
25. Patzkowski, J. C., Rivera, J. C., Ficke, J. R., & Wenke, J. C. (2012). The changing face of disability in the US Army: the Operation Enduring Freedom and Operation Iraqi Freedom effect. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, 20, S23-S30.
26. Richard P. Mustion, “Sustaining Our Army Then and Now,” *Professional Bulletin of the United States Army Sustainment PB700-09-06 41*: (November-December 2009)
27. Department of the Army, Regulation 350-1, *Training: Active Duty Training for FORSCOM Units-FORSCOM* (Fort McPherson, Georgia: United States Army Forces Command, 2002), 19.
28. Training Circular [TC] 3-22-20. *Army Physical Readiness Training*. Washington, DC: Headquarters, Department of the Army, 2010
29. Buckner, K. (2019). What the critics miss: the army combat fitness test is going to make us a more combat-read force. *Modern War Institute at West Point*, 9.

30. Physical Fitness Training Information Paper (2013), prepared for General Raymond T. Odierno.
31. Gabremariam, S. (2019). One Army, One Standard: The New Fitness Test Should Have a Single Scoring System. *Army Law.*, 80.
32. Eustis, T. F. (2019). TRADOC–ARMY COMBAT FITNESS TESTING (ACFT) DESIGN.
33. Brading, T. (2020). ACFT 2.0: Changes Sparked by COVID-19. *US Army, June, 15.*
34. Brading, T. (2021). ACFT 3.0: exploring a more inclusive scoring assessment, planks stay. *Army News Service.*
35. Lynch, A. (2021). The History of the US Army’s Physical Fitness Testing.
36. Novak, K. A. (2021). A Critical Review of the Baseline Soldier Physical Readiness Requirements Study. *arXiv preprint arXiv:2110.03062.*
37. Hardison, C. M., Mayberry, P. W., Krull, H., Setodji, C. M., Panis, C., Madison, R., & Wong, J. (2022). *Independent Review of the Army Combat Fitness Test: Summary of Key Findings and Recommendations* (p. 92). Santa Monica, CA, USA: RAND Corporation.
38. Gillibrand, K., & Blumenthal, R. (2020). Limitation on the Implementation of the Army Combat Fitness Test. Washington, DC; United States Senate.
39. Hardison, Chaitra M., et al. (2022) "Independent Review of the Army Combat Fitness Test."
40. United States Department of the Army (2022) Active Component Demographics. Washington, DC Department of Defense.