Review Report

Baseline Soldier Physical Readiness Requirements Study

By

THE UNIVERSITY OF IOWA

1 April 2020

Contents

1. Executive Summary 3
2. Background 4
3. Evaluation 6
   3.1 Technical soundness of methodology 6
   3.2 Appropriate criterion variable 7
   3.3 Method of identification of physical test events 7
   3.4 ACFT raw scores vs APFT 8
   3.5/3.6 Definition of performance and measurements per event/task and identification of decision criteria for analytical approaches 10
   3.7 Assessment of technical/scientific analysis 11
4. Findings 12
   4.1 Primary Findings 12
5. Summary of Review 13
6. Reviewers' Biographies 14
   6.1 Karim A. Malek, PhD 14
   6.2 Kevin C. Kregel, PhD 15
   6.3 Laura A Frey-Law, PhD, PT 15
   6.4 Landon Evans, MS, RD, CSCS 16
   6.5 Rajan, Bhatt, PhD 16

1. Executive Summary

A team at the University of Iowa Virtual Soldier Research program has conducted a peer review of the Army Baseline Soldier Physical Readiness Requirements Study (BSPRRS). (See reviewers' biographies in Section 6 for expertise of the five-member review team.) The BSPRRS was conducted by the U.S. Army Center for Initial Military Training at Ft. Eustis, VA, to determine the physical requirements for Soldiers in a combat environment. The BSPRRS is a central portion of the U.S. Army's research effort to reform the Army physical fitness test and change the culture of fitness with the goal of better preparing soldiers and reducing injuries.

The University of Iowa review team submits that the stated objectives of the Study were successfully achieved, namely: (1) determining the baseline physical requirements of the Warrior Tasks and Battle Drills (WTBD) and Common Soldier Tasks (CST) in men and women; (2) determining that the Army Physical Fitness Test (APFT) explained less than half of the variability in WTBD/CST performance; and (3) determining that a set of six common physical fitness test events (ACFT) were more predictive (> 70% variability explained) of combat task performance than the APFT.

First, the BSPRRS applied a combination of qualitative and quantitative approaches to develop an obstacle course as a surrogate representation of WTBD/CSTs. Five core tasks were distilled to the final WTBD/CST assessment through practical considerations, pilot testing, field observation, and input from focus groups. Overall, this methodology, used to develop the criterion metric of soldier physical requirements (i.e. WTBD/CST obstacle course completion time), was scientifically valid. Both men and women were evaluated in a ratio reflecting the current proportions across all components of the Army (84% men, 16% women).

Second, both the classic three-task APFT and the new Army Combat Fitness Test (ACFT), a set of six fitness tasks aimed at better representing recognized fitness domains, were assessed as predictors of WTBD/CST performance. The BSPRRS showed that the APFT explains less than half of the variability observed in soldier WTBD/CST completion times. Accordingly, this finding validates the need for a better physical assessment to predict combat fitness. The six-task ACFT was developed through testing of soldiers, using both stepwise linear regression and stakeholder feedback, and further validated with additional testing of separate Army cohorts. The process of evaluating multiple potential fitness assessments representing various domains of fitness relative to the simulated WTBD/CST utilized scientifically appropriate and rigorous methodologies. The ACFT explained 70-85% of the variability in simulated WTBD/CST performance, nearly twice the original APFT. Further face validity of the ACFT is supported by its similarity to other fitness tests currently in use by the Army and Air Force (e.g., Ranger Athlete Warrior, and Battlefield Airman Fitness Assessment).

In summary, the University of Iowa review team submits that the BSPRRS Study successfully achieved the stated aims. The resulting six-event ACFT is based on a scientifically valid examination of experimental results and is predictive of combat soldier task performance in men and women. Key findings of this peer review include both strengths and inherent limitations that are addressed in a feasible manner. The University of Iowa assessment of the BSPRRS Study is that it was conducted by a well-qualified team of scientists and military personnel, the methodologies utilized were appropriate and rigorous, and the results provide strong baseline empirical validation for the Army Combat Fitness Test.

2. Background

The Baseline Soldier Physical Readiness Requirements Study (BSPRRS) was conducted by the U.S. Army Center for Initial Military Training at Ft. Eustis, VA. The main objective of this study was to determine the physical requirements for Soldiers in a combat environment. The study used physically demanding, commonly occurring and critical Warrior Tasks and Battle Drills (WTBD) and Common Soldier Tasks (CST) as a proxy for combat tasks. There were three stated objectives: (1) determine the baseline physical requirements of WTBD/CST; (2) determine combat task variability explained by the Army Physical Fitness Test (APFT); and (3) determine if other common physical fitness test events were more predictive of combat task performance.

The BSPRRS was conducted in three phases. In Phase I, researchers conducted a systematic review and Soldier interviews and surveys. They also deconstructed Army Warrior Tasks and Battle Drills and Common Soldier Tasks to identify those tasks that were physically demanding, commonly occurring and critical. In the final part of Phase I, male (243) and female (47) Soldiers (Ft. Carson) participated in the development of the WTBD simulation test (WTBD-ST). In Phase II, male (275) and female (46) Soldiers (Ft. Riley) performed the WTBD-ST, the Army Physical Fitness Test (APFT), and 23 common physical fitness test events to determine if other physical fitness events better predicted WTBD/CST performance. In Phase III, male (136) and female (16) Soldiers (Ft. Benning) performed the eight (8) physical fitness test events, identified in Phase II, sequentially with no programmed rest and the WTBD/CST test. Physical fitness test events were regressed against the WTBD/CST using multiple linear regression.

In the Phase II data analysis (Ft. Riley), the three-event Army Physical Fitness Test (APFT), while able to predict a portion of WTBD/CST performance ($R^2 = 0.432$, $p< 0.001$), relative to most standards for excellent prediction of human performance this fell well short (i.e., assuming standard target of $R^2 = 0.70$). Thus, this finding supported the need to identify a more comprehensive assessment of soldier preparedness.

In the initial step-wise regression model, where 23 fitness tasks were evaluated relative to WTBD/CST performance, researchers found a much higher predictive ability using seven (7) variables ($R^2 = 0.737$; $p< 0.001$): sled drag, two-mile run, ~1RM deadlift, sled push, push-ups, goblet squat, and power throw. While predictive validity is crucial to the final regression model, it was equally important to the Army to produce a model that assessed multiple domains of fitness, that could motivate Soldier training, and ultimately reduce injuries. After considering these qualitative factors, an eight (8) variable model was developed and analyzed, including the following tasks: sled drag, two-mile run, ~1RM deadlift, sled push, push-ups, 300yd shuttle run, leg tuck, and power throw. The results of this analysis showed that these eight (8) tasks predicted 73.5% of the variability in WTBD/CST completion time performance (R^2 = 0.735). Thus, these tasks were chosen as the final set.

In the Phase III data analysis (Ft. Benning), the eight (8) physical fitness test events were administered sequentially with no programmed rest. On different days, the WTBD-ST was also administered. Four primary predictors were identified (R^2 = 0.832; p= 0.001): sled drag, power throw, two-mile run, and 1RM deadlift, while three secondary measures accounted for additional variability: leg tuck, 300m shuttle run and push-ups (R^2 = 0.835, p< 0.001).

In summary, to ensure Soldiers are prepared to execute physically demanding combat tasks, physical fitness should be assessed across multiple primary physical fitness domains. Test events measuring muscular strength, aerobic and anaerobic endurance and anaerobic power were most predictive of combat task performance. Secondary measures of speed, core strength, and upper body muscular endurance accounted for additional variability and provided physiological balance to the test battery and focused physical training intending to reduce musculoskeletal injuries.

The US Army report lists the following conclusions of this study:

Three questions were addressed and answered: (1) What are the baseline physical readiness requirements of the physically demanding, commonly occurring and critical Warrior Tasks and Battle Drills and Common Soldier Tasks (WTBD/CSTs)?; (2) Does the current three-event Army Physical Fitness Test (APFT) adequately assess the baseline physical readiness requirements required to execute WTBD/CSTs?; and (3) If the three-event APFT does not assess the baseline physical requirements, what physical fitness test events better predict a Soldier's success on high physical demand WTBD/CSTs? Based on the findings of this study, the answers to these three questions are:

1. The baseline physical components required for Soldier success on high-demand Warrior Tasks and Battle Drills and Common Soldier Tasks are: muscular strength, anaerobic power and aerobic, anaerobic, and muscular endurance. Training in and assessment of these primary components of physical fitness are necessary to prepare Soldiers to over-match in multi-domain battle.

2. The Army Physical Fitness Test (APFT) explains less than half of the variance in WTBD/CST performance (R^2 = 0.42), showing the APFT does not ensure Soldiers are capable of performing physically demanding Warrior Tasks and Battle Drills and Common Soldier Tasks.

3. The eight (8) test event battery identified in the Baseline Soldier Physical Readiness Requirements Study is a relatively high predictor of WTBD/CSTs performance (R^2 = 0.74 - 0.84).

To reduce the burden of eight tasks, analyses were performed to estimate the predictive value of combining three (3) tasks into a single test: the sprint-drag-carry. Modeling the use of the resulting six (6) fitness assessment tasks also showed very high predictive ability. Thus, the final report uses a six-task physical fitness assessment as the Army Combat Fitness Test (ACFT).

3. Evaluation

In conducting this review, we have evaluated the Baseline Soldier Physical Readiness Requirements Study, received and reviewed associated videos, and assessed other ancillary material as deemed necessary. The review and evaluation are focused on the technical scientific aspects of the report and do not address any policy making issues. This review is not an audit, does not seek to replicate any of the experimental setups or exercises, and does not include any additional data from human subjects studies.

3.1 Technical soundness of methodology

Is the methodology used to develop the Warrior Tasks and Battle Drill Simulation Test (WTBD-ST) sound?

Our collective opinion is that the U.S. Army team that has developed the WTBD/CST test is well-qualified and has the appropriate mix of personnel, experience, and background. It is also our collective opinion that the study designs and analytic methods used in Phase I - the development of the WTB-ST - are well accepted and consistent with professional and scientific norms. This process was well-described and serves as the benchmark for the remaining analyses (Phases II and III). The combination of a systematic review along with Soldier interviews and surveys provided the necessary background to identify key demanding tasks to incorporate into the simulated battle drills (WTBD/CST). The identification of five core tasks was appropriate and clearly explained:

1. Move over long distances under heavy loads;
2. Build a hasty fighting position;
3. Move over-under-around-through obstacles on uneven terrain;
4. Employ progressive levels of force (combatives);
5. Extract and transport a casualty.

The team identified tasks to represent each core task, and distilled them to the final WTBD-ST through practical considerations, pilot testing, field observation, and input from focus groups. Overall, the methodology used was sound.

3.2 Appropriate criterion variable

Was the WTBD/CST an appropriate criterion variable for a concurrent validation model?

Overall, the explanatory report justifies the use of the WTBD/CST as a surrogate measure of soldier performance. The sound methodology used to develop the WTBD/CST provides a degree of construct validity. However, a potential weakness of the WTBD/CST as a criterion variable is that the only measure of performance is completion time. It is not clear the degree to which completion time indicates Soldier success and survivability. However, other quality metrics would require substantially more resources (e.g. motion capture technology) without clear improvements in outcomes. Thus, the use of the simple metric of completion time is reasonable.

Future investigations may benefit from exploring what WTBD/CST completion times mean in terms of acceptable performance. Additional analyses comparing poor performers (sub-threshold for passing performance) relative to acceptable performers would provide an additional measure of construct validity for its use as a concurrent validation metric. That is, the goal of determining the predictive value of the ACFT tests regarding poor vs acceptable WTBD/CST performance as opposed to predicting the linear change in WTBD/CST completion times may be an alternative strategy for the use of the WTBD/CST. However, there is currently no clear metric of performance quality, making this a potential area for future development. Thus, while completion time is only one aspect of success in performance of the WTBD/CST, it provides a reasonable, but limited, metric for concurrent validation studies.

3.3 Method of identification of physical test events

How were the physical fitness test events identified and what criteria were used for inclusion?

While we are unaware of how the 23 physical tasks were initially chosen, the tasks themselves represented a wide range of fitness domains: muscular strength, muscular endurance, cardiovascular endurance, flexibility, agility, coordination, power, speed, and balance. Thus, the tasks assessed were sufficient to provide improved insight to the physical domains of performance.

Two key considerations stressed in the report were that the potential physical test events needed to be feasible and not require excessive equipment. These considerations were likely factors that played a significant role in the selection of specific exercises.

While physical domains are not assessed in isolation, certain tasks are more prominent in certain domains. Based on the physical abilities of the Soldier, some tasks may be more of a “maximal” test vs a “capacity” test. For instance, an individual with substantial upper-body strength may display muscular endurance during a bench press test, while a weaker individual may display their maximal strength in performing the same test.

Below is a listing of the 23 physical fitness tests that were evaluated, and the physical domains engaged during the performance of these tests:

- Standing Long Jump: Power, coordination, balance, full-body
- Vertical Jump: Power, coordination, full-body
- Medicine Ball Throw: Power, coordination, balance, full-body
- Sled Push: Muscular strength (or muscular endurance), full-body
- Sled Drag: Muscular strength (or muscular endurance), full-body
- Sumo Squat: Muscular strength (or muscular endurance), flexibility, lower-body
- Bench Press: Muscular strength (or muscular endurance), upper-body
- Hex-bar deadlift: Muscular strength (or muscular endurance), full-body (lower-body)
- Dips: Muscular strength (or muscular endurance), upper-body
- Pull-Ups: Muscular strength (or muscular endurance), upper-body
- Bench Press Endurance: Muscular endurance, upper-body
- Modified Sit-Up (crunch): Muscular endurance, torso
- Leg Tuck: Muscular strength (or muscular endurance), flexibility, torso
- Weighted Trunk Rotations: Muscular strength (or muscular endurance), torso
- Abdominal Rower: Muscular endurance, flexibility, torso
- Kettlebell Squat Endurance: Muscular endurance, flexibility lower-body
- 300m shuttle run: Cardiovascular endurance, speed, agility, coordination (anaerobic/aerobic)
- Loaded 300m shuttle run: Cardiovascular endurance, speed, agility, coordination (anaerobic/aerobic)
- Illinois Shuttle Test: Speed, coordination, agility (anaerobic/aerobic)
- 400m sprint: Cardiovascular endurance, speed (anaerobic / aerobic)
- Push-Up: Muscular endurance, upper-body
- Sit-Up: Muscular endurance, torso
- Two-Mile Run: Cardiovascular endurance, speed (aerobic / anaerobic)

The only domain that was not specifically addressed was body composition at the level of percentage of lean tissue to total body mass. However, several of the tests involved the assessment of multiple fitness domains, thereby providing a more complete representation of overall fitness.

In summary, the final six-task ACFT provides a more complete representation of a Soldier’s true abilities, encompassing more fitness domains than the three-task APFT.

3.4 **ACFT raw scores vs APFT**

How do ACFT raw scores (number of repetitions or time; not a scale score) compare to the raw score performance on the Army Physical Fitness Test (APFT)?

While we cannot directly compare the raw ACFT scores to the raw APFT scores without access to the raw data, we can analyze several aspects related to these two performance-based physical fitness tests.

First, both tests were analyzed relative to the WTBD/CST outcomes to determine how well the fitness scores predicted WTBD/CST performance. The APFT predicted 43% of the variance in WTBD/CST completion times (i.e., $R^2 = 0.43$), which was described as relatively moderate to poor. However, this value translates to a correlation coefficient of R = 0.656, which is by many definitions moderate to strong (e.g., Cohen's definition of effect sizes for correlations). In comparison, the ACFT predicted from 63 to 83% of the variance in assessed WTBD/CST performance, based on the Phase II findings using eight tests or in Phase III using similar tasks combined into six tasks in a new cohort (i.e., $R^2 =$ 0.63 to 0.83). Thus, the ACFT scores explain nearly twice as much of the variance in WTBD/CST completion times than the original APFT. Overall, these comparisons imply that the raw ACFT and APFT scores are likely correlated to some degree; however, the ACFT scores show greater predictive capacity.

Second, the mean, standard deviation (SD), and ranges of the APFT and ACFT tests were reviewed. By calculating the coefficient of variation (CV = SD/mean) for each test, the range of variance observed for each test provides insight into how consistent or inconsistent the performances are across the cohorts examined (i.e., currently trained soldiers). The original three APFT tests resulted in relatively low CVs (9 – 19% for men; 9 – 31% for women). However, for the eight tests originally chosen from the 23 tasks, the CVs showed dramatic variation (6 – 58% for men; 8 – 139% for women). In particular, the leg tuck showed high CV (peak for both men and women). If that test were excluded, then the ranges would reduce to 6 – 23% for men, and 8 – 35% for women. This additional analysis suggests that even in a cohort of active military personnel, performance in the leg-tuck task is highly variable, suggesting core strength is inconsistent in current Army personnel. Future research may be needed to determine whether this variance is reduced with a greater emphasis on core strength, or if it may be an inherently variable fitness domain. Conversely, the tests with minimal variability may have limited ability to predict variance observed in WTBD/CST performance. In fact, two of the three APFT tests showed very low CVs, which may in part contribute to the lower predictive capability of the APFT relative to the simulated WTBD/CST.

In summary, the APFT and ACFT raw scores cannot be directly compared based on the report summary and data tables provided. However, using indirect assessments, it appears most of the ACFT tests involve reasonable levels of variability to allow for more accurate prediction of simulated WTBD/CST than observed using only the three APFT tests. Further the use of raw scores (repetitions, time, etc.) over standardized scores appear to be a valid methodology given the ability of the ACFT to so highly predict performance on the WTBD/CST performance.

3.5/3.6 Definition of performance and measurements per event/task and identification of decision criteria for analytical approaches

In this section we identify how performance was defined and measured per event/task and address the identification of the criteria used to select and evaluate the validity and performance standards of each event/task. This section also addresses the decision criteria for the analytical approaches (e.g., logically specified for analyses and consideration for how decisions may impact test development).

There were several analytical approaches employed to evaluate the predictive value of 23 potential physical fitness assessments. The decision criteria were stated clearly for a number of these analyses (with some limitations noted).

The use of regression was clearly defined, with a pre-defined criterion of R^2>= 0.70 stated as a well-accepted prediction level. While the 0.70 criterion may be arguably a reasonable criterion, one weakness of this approach is the lack of consideration of WTBD/CST test performance stability/precision. When the test was repeated, albeit under slightly different conditions, the correlation was reported to be R = 0.83, which translates to a coefficient of determination less than the criterion (R^2 =0.69). This indicates that any attempts to explain more variance than what the same task can explain under slightly different conditions is at risk for model over-fitting. This suggests that while high predictive ability is ideal, any series of fitness tasks that can explain 70% of the variance in WTBD/CST performance are practically as good as repeating the WTBD/CST assessments directly.

The proportion of women studied reflects the current proportion of women in the U.S. Army, resulting in an inherently unbalanced study design (women: 16.2% Phase I; 14.3% Phase II; 10.5% Phase III). The argument that men and women were considered together is valid, because, "… baseline Warrior Tasks and Battle Drills and Common Soldier Tasks are criterion tasks that apply equally to men and women." The potential risk associated with the low number of women studied is that the determination of which tasks best predict or represent WTBD/CST performance could be influenced towards strategies used predominantly by men. If women were to use different strategies to successfully complete the WTBD/CST, the analyses might not identify those alternate strategies. For example, subjective comments from soldiers noted "women's concerns related to effects of height and body mass. Taller/higher body mass Soldiers did not identify the same problems." To model whether a more equal distribution of men and women would suggest different predictors as being important in predicting women's performance (and potentially any individual who is shorter, with lower body mass, and/or less muscle mass), secondary analyses were performed and shared with the review team. These findings indicated that largely the same final predictors of the ACFT were identified in women alone, and in an oversampled imputation of women relative to men. While these analyses may need to be replicated experimentally in future endeavors, particularly if the proportion of women to men increases in the Army over time, these additional analyses support the validity of using a single set of tasks in the ACFT as predictors of soldier capability.

3.7 **Assessment of technical/scientific analysis**

Overall, there are several study designs and analytic choices that are well accepted and consistent with professional and scientific norms. Future research efforts could augment these initial studies using additional alternate approaches that may provide additional insights that could further advance the science of Soldier fitness and readiness assessment.

The test protocols described utilize reasonable methodology to address the aims, including: (1) a reasonably large sample size to promote stable results, (2) use of regression techniques to help determine predictive tasks (as opposed to only empirically choosing tasks based on expert opinion, and (3) repetition of testing at multiple Army sites. In terms of the tests themselves, the use of "reps to fatigue" to mathematically estimate 1RM (one repetition maximum) is a strength of the study, as it is less susceptible to motivational bias and reduces risk of injury.

The fact that five fitness tests explained as much variance in the WTBD/CST as was observed between repeating the WTBD/CST under different conditions suggests these five tests may be sufficient. However, the argument to include more tests is reasonable, as the WTBD/CST is a surrogate representation of Soldier tasks. The addition of more fitness tasks may prove to promote more well-rounded fitness training, but future studies will be needed to ascertain if that is indeed the case.

The need to complete the full battery of 23 fitness tasks, in addition, the WTBD/CST obstacle course practically limited the testing. Thus tests were completed on multiple days (i.e., four days apart in the FT Benning sample). While it is understandable that testing of so many tasks on one day is infeasible (e.g., fatigue and order effects, ability to test large numbers of individuals, etc.), there are also limitations to multi-day testing. In practice, the final ACFT will occur over a single day. However, given the constraints of feasibility, the multi-day testing design is reasonable.

While several of the statistical methods employed are commonly reported in the literature (correlations, regressions, etc.), they typically rely on assumptions of independence and linearity of the physical tests to WTBD/CST performance. Alternate approaches, including Bayesian theory or Information theory-based approaches (i.e., machine learning), may be appropriate, and considered in future investigations to further evaluate the questions posed in these studies. These alternate approaches are particularly powerful when assessing predictive power.

4. Findings

The primary findings of the Baseline Soldier Physical Readiness Requirements Study evaluating soldier physical readiness requirements in three phases are summarized again as follows:

Phase I: Identification of physically demanding tasks and the development of the WTBD/CST.

Phase II: Assessed WTBD/CST performance against APFT and 23 fitness assessments. Identified eight (8) tests that better predicted WTBD/CST (R^2 = 0.74) than the APFT (R^2 = 0.43).

Phase III: Repeat testing of the new fitness tests (combined to 6 tasks to result in the ACFT) vs WTBD/CST performance in a new sample (R^2 = 0.82)

4.1 Primary Findings

1. Phase I involves the development of the WTBD/CST. This is well-described and serves as the concurrent validation criterion for the remaining analyses. While completion time is a simple metric of success and performance, it is feasible for widespread use.

2. Phase II identifies which tasks to include in the ACFT, based on linear regression relative to the WTBD/CST completion times. While we identified both strengths and weaknesses in the approach, most weaknesses resulted from practical limitations. A threat to validity is the unbalanced sampling of men and women; however, this represents current U.S. Army proportions. This issue was secondarily examined using statistical modeling, indicating the final ACFT includes fitness tasks that best predict WTBD/CST performance in men and women.

3. Phase III repeats the testing to confirm the predictive value of ACFT relative to the WTBD/CST. Similar strengths and weaknesses were noted for the

Phase III efforts, due to the practical considerations of a large physical study. However, the design to replicate and repeat the evaluation is a strength that further supports the face validity of the ACFT as a reasonable set of tasks to predict Soldier fitness and combat readiness.

5. Summary of Review

The Baseline Soldier Physical Readiness Requirements Study (BSPRRS) is a central portion of the U.S. Army's research effort to reform the Army physical fitness test and change the culture of fitness with the goal of better preparing Soldiers and reducing injuries. Several relevant background facts include that approximately 84% of all U.S. Army Soldiers are men, thus only 16% are women. All Soldiers, regardless of Military Occupational Specialty (MOS), must be capable of executing Warrior Tasks and Battle Drills and Common Soldier Tasks (WTBD/CST). Some Soldiers have a higher physical requirement due to additional demands of their MOS/METL. Approximately 26% of Soldiers are in a "Heavy" MOS; ~23% in "Significant" MOS, and ~51% in "Moderate" MOS.

There were three stated objectives of the BSPRRS, that were appropriately met upon review of the study report and documentation:

(1) determining the baseline physical requirements of WTBD/CST;

(2) determining combat task variability explained by the Army Physical Fitness Test (APFT); and

(3) determining if other common physical fitness test events are more predictive of combat task performance.

First, a combination of qualitative and quantitative approaches was used to develop an obstacle course as a surrogate representation of WTBD/CST tasks. Systematic review, Soldier interviews and surveys provided multi-modal methodologies to identify key demanding tasks for inclusion into the simulated battle drills (WTBD/CST). Five core tasks were distilled to the final WTBD/CST through practical considerations, pilot testing, field observation, and input from focus groups. Overall, the methodology used to develop the criterion metric of soldier physical requirements was scientifically valid.

Second, the BSPRRS showed that the current three-task APFT explains less than half of the variability observed in soldier completion times of the simulated WTBD/CST obstacle course. Accordingly, this finding validates the need for a better physical assessment to predict combat fitness.

Finally, evaluating multiple potential fitness assessments representing various domains of fitness relative to the simulated WTBD/CST utilized scientifically appropriate and rigorous methodologies. A two-pronged approach, using both stepwise linear regression and stakeholder feedback, identified eight tasks that explained 70-83% of the variability in simulated WTBD/CST performance, across several army populations. Further, secondary analyses showed that the inherently unbalanced proportion of men and women tested did not negatively impact the identification of optimal fitness tasks of those evaluated in a substantive way. Accordingly, the choice of optimal fitness tasks (combined to six total, making up the Army Combat Fitness Test, ACFT) is based on a scientifically valid examination that is appropriate for men and women.

In summary, the University of Iowa review team submits that the stated objectives of the BSPRRS Study were successfully achieved. Key findings of this peer review include both strengths and inherent limitations that are addressed in a feasible manner. The University of Iowa assessment of the BSPRRS Study is that a well-qualified team of scientists and military personnel conducted it, has used appropriate and rigorous methodologies, and is technically sound, resulting in valid findings.

6. Reviewers' Biographies

6.1 Karim A. Malek, PhD

Dr. Karim A. Malek is nationally and internationally recognized in the areas of robotics and human simulation. He is a Professor of Biomedical Engineering at the University of Iowa and serves as the Director of the Iowa Technology Institute, a world-renowned research center with 8 departments, including a national lab. Under his leadership, the Center has grown from 30 researchers to about 180 (scientists, engineers, and support staff) and a substantial portfolio of research spanning the Modeling and Simulation area.

Dr. Malek is the founder and director of the SANTOS virtual soldier research program, where he leads projects with all services of the US Military (US Army and US Marines), and several industry partners including Ford, GM, Chrysler, Rockwell Collins, Caterpillar, and others. Dr. Malek has executive management experience and is an entrepreneur that has launched basic research into commercial products. His research on human modeling and simulation has attained international recognition, has been published in many prestigious journals such as the Journal of Biomechanics, and has been featured in several respected science media outlets such as the Discovery Channel.

Dr. Malek received his MS and PhD degrees in robotics and simulation from the University of Pennsylvania and a BS in Mechanical Engineering from the University of Jordan. He has authored or co-authored over 220 technical papers and 2 US Patents. He is a Fellow of AIMBE, served as the elected President of the International Society for Human Simulation (ISHS), and is the recipient of many international awards including the Board of Regents Faculty Excellence award, the College of Engineering Award for Exceptional Research, the Fulbright Scholarship, the John T. Parsons Outstanding Manufacturing Engineer award from the Society of Manufacturing Engineers (SME), The Arch T. Colwell Best paper award from the Society of Automotive Engineers (SAE). Dr. Malek founded three technology companies and serves on their board.

6.2 Kevin C. Kregel, PhD

Dr. Kevin Kregel is currently the Executive Vice Provost for Faculty at the University of Iowa (UI). He is also a Professor of Human Physiology and former Chair of the Department of Health & Human Physiology. He earned a bachelor's degree (Biology) and doctorate (Physiology & Biophysics) from the University of Iowa and subsequently performed an NIH Postdoctoral Fellowship at the University of Arizona. In 1993, Dr. Kregel joined the faculty at the University of Iowa, and currently holds the rank of full professor.

Dr. Kregel is internationally recognized for his work on physiological adaptation to stressors, especially related to challenges such as exercise and heat stress, and his research program has consistently been funded by the NIH and Department of Defense. He has been an active member of the Virtual Soldier Research Program in the Iowa Technology Institute for the past decade, overseeing the merging of human physiological testing with physics-based digital human modeling and simulation tools for the development of predictive models for enhanced performance and injury prevention.

Dr. Kregel is a Collegiate Fellow in the UI's College of Liberal Arts & Sciences and a past recipient of an Alexander von Humboldt Foundation Research Fellowship. He has been a member of several national and international review panels, advisory councils, and oversight committees. He has also served for many years as an associate editor of several leading journals in the fields of exercise and human physiology.

6.3 Laura A Frey-Law, PhD, PT

Dr. Laura Frey-Law is an Associate Professor in the Department of Physical Therapy & Rehabilitation Science in the Carver College of Medicine. Her background is in biomedical engineering (BSE, Iowa and MS, Michigan) and physical therapy rehabilitation (MPT and PhD, Iowa). She is the Director of the Neuromuscular Biomechanics Laboratory and has been on faculty at the University of Iowa since 2005.

Dr. Frey-Law is nationally and internationally recognized for her work on muscle pain, fatigue and strength; incorporating human psychophysical testing and mathematical modeling studies to advance our understanding in these areas. Her work has been supported by the NIH, Department of Defense, and private foundations. Dr Frey-Law has been an active member of the Virtual Soldier Research (VSR) Program in the Iowa Technology Institute since 2005, overseeing the development of joint-based strength and fatigue models for use in the physics-based digital human modeling and simulation tools in Santos.

Dr. Frey-Law has over 50 peer-reviewed publications, has served in leadership positions in national and international professional organizations, and is an Associate Editor for the Journal of Applied Biomechanics. She serves as a reviewer for over 20 scientific journals, has served as a reviewer for grant and fellowship applications for the APTA, the APS, and the Center for sensorimotor Systems at the University of Aalborg. She has garnered several awards including the APTA Eugene Michels Young Investigator Award; the VSR Excellence in Research Leadership and Collaboration Award; the Most Outstanding Small Group Presentation Award at the SAE Digital Human Modeling Conference, Lyon, France; and was a Hancher-Finkbine

Medallion recipient, Rhodes Scholar Finalist, and NASA Space Life Science Training Program participant as a student.

6.4 **Landon Evans, MS, RD, CSCS**

Landon Evans is Director of Sports Science at the University of Iowa (UI). He joined the UI Athletics department in August 2012. His primary duties include directing the sports science department and strength & conditioning duties with the men's and women's sprints and jumps. Prior to July 2016, he oversaw all strength & conditioning duties with track & field, and was the nutrition coordinator for Olympic sports. Before joining the Iowa staff, Evans has been the Director of Sports Nutrition and Assistant Strength & Conditioning Coach at Illinois State from 2010-2012. He served as the Director of Athletic Development at the Performance Training Center in Beaverton, OR from June 2008- August 2009 and as Associate Director of Strength & Conditioning at Illinois State from January 2005 - May 2008.

While a student, he worked in strength & conditioning at Iowa State and Central College. Evans received a bachelor's of Science in Health & Human Performance from Iowa State in 2005, a master's of Science in Kinesiology and Recreation from Illinois State in 2007 and DPD Concentration in Family & Consumer Sciences, with emphasis on food, nutrition and dietetics from Illinois State 2011.

He is a Registered Dietitian (RD), Certified Strength and Conditioning Specialist (CSCS) from the National Strength and Conditioning Association, Strength & Conditioning Coach Certified (SCCC) though the Collegiate Strength & Conditioning Coaches Association. He is CPR/AED certified.

6.5 **Rajan, Bhatt, PhD**

Dr. Rajan Bhatt is an assistant director at The University of Iowa's Virtual Soldier Research (VSR) program. He received his M.S. (Feb 04) and Ph.D. (Feb 07) degrees in mechanical and aerospace engineering from the State University of New York at Buffalo and a B.E. in mechanical engineering (Feb 01) from the MS University of Baroda, India. His research interests include modeling, simulation, and control of constrained multibody dynamic systems. After receiving his bachelor's degree, Rajan worked for a year for a private firm in his hometown, Baroda, designing and analyzing (via motion simulation and finite element analysis) various mechanical components of screw air compressors, surface-finish vibrators, and injection molding machines for geographically distributed clients in India, Europe, and the USA.
He then received his M.S. and Ph.D. degrees from the University at Buffalo, specializing in the field of robotics. During this time, he worked on various projects like optimal path generation for a group of nonholonomic mobile robots, musculoskeletal analysis of jaws of sabertooth cats, and dynamic modeling and decoupled control of redundant nonholonomic mobile robots. He also instructed a graduate-level course in robotics and assisted in teaching many graduate- and undergraduate-level mechanical engineering courses.

After receiving an M.S. and before continuing for a Ph.D., he did a seven-month internship at Children's Hospital Los Angeles (starting Jan 04) and worked under researchers from Jet Propulson Labs in Pasadena on a project to build a wireless heterogeneous distributed sensor network platform with autonomous event detection and sensor processing capabilities, and real-time multisensor control and resource management.

Dr Bhatt, since joining the University of Iowa, has led the research effort in physics-based human modeling and simulation at the Virtual Soldier Research(VSR) Program. He has managed many DOD and private company projects for VSR and served as a mentor to graduate and undergraduate students working on various research projects. He has also served on the boards of many international conferences and was awarded College of engineering's Excellence in Research award and best paper award at the Human Modeling and Simulation conference in Montreal, Canada in 2016.