TECHNICAL REPORT
NATICK/TR-17/002

AD ________________

RELIABILITY ANALYSIS OF TIME TO COMPLETE THE OBSTACLE COURSE PORTION OF THE LOAD EFFECTS ASSESSMENT PROGRAM (LEAP)

by
K. Blake Mitchell
Jessica M. Batty
Megan E. Coyne
Linda L. DeSimone
and
Carolyn K. Bensel*

***Oak Ridge Institute for Science and Education (ORISE), Oak Ridge Associated Universities (ORAU)
Belcamp, MD 21017

October 2016

Final Report
October 2014 – June 2015

Approved for public release; distribution is unlimited

U.S. Army Natick Soldier Research, Development and Engineering Center
Natick, Massachusetts 01760-5020

DISCLAIMERS

The findings contained in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of trade names in this report does not constitute an official endorsement or approval of the use of such items.

DESTRUCTION NOTICE

For Classified Documents:

Follow the procedures in DoD 5200.22-M, Industrial Security Manual, Section II-19 or DoD 5200.1-R, Information Security Program Regulation, Chapter IX.

For Unclassified/Limited Distribution Documents:

Destroy by any method that prevents disclosure of contents or reconstruction of the document.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
25-10-2016

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
October 2014 – June 2015

4. TITLE AND SUBTITLE
RELIABILITY ANALYSIS OF TIME TO COMPLETE THE OBSTACLE COURSE PORTION OF THE LOAD EFFECTS ASSESSMENT PROGRAM (LEAP)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
63001

6. AUTHOR(S)
K. Blake Mitchell, Jessica M. Batty, Megan E. Coyne, Linda L. DeSimone, and Carolyn K. Bensel*

5d. PROJECT NUMBER
J50

5e. TASK NUMBER
52

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Natick Soldier Research, Development and Engineering Center
ATTN: RDNS-SEW-THB
10 General Greene Avenue, Natick, MA 01760-5020

8. PERFORMING ORGANIZATION REPORT NUMBER
NATICK/TR-17/002

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
*Oak Ridge Institute for Science and Education (ORISE), Oak Ridge Associated Universities (ORAU) Maryland, 4692 Millennium Drive, Suite 101, Belcamp, MD 21017

14. ABSTRACT
This report documents a study conducted on an obstacle course that is part of a new battery of physical tests for assessing effects of military clothing and equipment on the performance of service members. Activities performed on the course include running, crawling, and balancing. The battery, referred to as the Load Effects Assessment Program (LEAP), emanated from work by the U.S. Marine Corps to identify an objective method for quantifying effects of service members' loads on execution of combat-relevant movements and tasks. The U.S. Marine Corps and Army and Canadian and Australian military establishments are using or planning to use the LEAP. This study was conducted to establish the reliability of time to complete the LEAP obstacle course. Participants, 19 U.S. Army enlisted men, were scheduled for one session per day for five days. The first session was an orientation that included familiarization with the obstacle course and having the men traverse the course at 50% and at 75% of maximal effort. Testing was conducted at the four remaining sessions and consisted of a participant expending 100% of maximal effort to complete the obstacle course twice at each session, for a total of eight trials. The men wore a basic uniform, a helmet, and combat boots. Time to complete the entire course was recorded, as were times to complete segments of the course. Statistical analysis indicated that course completion time stabilized after four trials. An intraclass correlation coefficient computed on the data for Trials 5 and 6 demonstrated a high degree of relative reliability of the course completion time measure, *ICC*(2,1) = .93. The standard error of measurement and the limits of agreement were relatively low, indicating a high degree of absolute reliability, *SEM* = 7.21 s and 95% LOA = -0.78 s ± 19.97. It was concluded that practice is required before consistent performance can be achieved on the obstacle course.

15. SUBJECT TERMS
BALANCE BIOMECHANICS LOADS(FORCES) PERFORMANCE TESTS POSTURE(PHYSIOLOGY)
MISSIONS FIELD TESTS PHYSICAL TESTS TEST AND EVALUATION PROTECTIVE CLOTHING
MOBILITY FLEXIBILITY TASK ANALYSIS MANEUVERABILITY PERFORMANCE(HUMAN)
RUNNING MOVABILITY ARMY PERSONNEL RANGE OF MOTION PROTECTIVE EQUIPMENT
TRAINING LOCOMOTION DATA COLLECTION MAXIMAL EFFORT ENDURANCE(PHYSIOLOGY)
CRAWLING WALKING OBSTACLE COURSES CIE(CLOTHING AND INDIVIDUAL EQUIPMENT)

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			K. Blake Mitchell
U	U	U	SAR	92	19b. TELEPHONE NUMBER *(include area code)* 508-233-5326

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

LIST OF FIGURES

LIST OF TABLES

PREFACE

The study reported here was carried out between October 2014 and June 2015 by personnel of the Biomechanics and Engineering Team and the Human Factors and Engineering Team, Warfighter Directorate, U.S. Army Natick Soldier Research, Development and Engineering Center, Natick, MA. The purpose of the study was to analyze the reliability of times to complete a 10-station obstacle course that is part of a battery of tests to assess physical performance of military personnel as affected by clothing and body-borne equipment. This test battery is referred to as the Load Effects Assessment Program (LEAP).

The effort was conducted under Project 14-077 (Physical Performance Effects of Clothing and Individual Equipment: Marine Corps – Load Effects Assessment Program Related Task Metrics).

ACKNOWLEDGEMENTS

The authors are most grateful to Mr. Adam N. Adams, Ms. Ellen Bertschi, Dr. Hyeg Joo Choi, Dr. Todd N. Garlie, Mr. Charles Glover, Ms. Caroline Kingsley, and Ms. Christine Sniezek, all of the U.S. Army Natick Soldier Research, Development and Engineering Center (NSRDEC), for their assistance in the conduct of this study. Ms. Karen Gregorczyk, Ms. Rose Guerra, Dr. Leif Hasselquist, Ms. Kerri Loverro, and Dr. Jeffrey Schiffman, NSRDEC, are acknowledged for their participation in the planning of this effort. Mr. James S. Cahoon, Mr. Michael P. Coyne, Mr. Richard Crivello, and Mr. David Garcia, NSRDEC, are particularly recognized for their work in assembling the obstacle course used in the study. The authors thank, as well, Dr. Hyeg Joo Choi, NSRDEC, and Ms. Linda L. Bossi, Defence Research and Development Canada-Toronto, for their helpful comments on an earlier draft of this report.

Special recognition is due to the study volunteers, who were enlisted men and women assigned to Headquarters Research and Development Detachment, NSRDEC.

EXECUTIVE SUMMARY

Background and Purpose. Between October 2014 and June 2015, the U.S. Army Natick Soldier Research, Development and Engineering Center (NSRDEC) conducted a study of a 10-station obstacle course that is part of a new battery of physical tests for assessing the effects of clothing and equipment on the performance of military personnel. The battery, which is referred to as the Load Effects Assessment Program (LEAP), includes rifle firing, vertical jump, and weight transfer activities, in addition to the obstacle course. The activities involved in completing the 10 stations of the course include running, crawling, climbing, vaulting, pulling, and balancing.

The LEAP battery emanated from efforts undertaken by the U.S. Marine Corps. In 2009, the Marine Corps Systems Command determined that an objective method was needed for quantifying effects of a service member's load on execution of combat-relevant movements and tasks. Tasks were identified that were judged to be critical in the areas of mobility, lethality, and survivability of dismounted troops. The Marine Corps Systems Command, in association with HumanSystems® Inc. (HSI), devised a test battery aligned with the critical tasks. An initial LEAP system was fabricated by HSI and set up by the Marine Corps Systems Command. Additional LEAP systems were produced subsequently for Canadian and for Australian defense establishments. More recently, two systems were fabricated for the U.S. Army. One of these systems is at the U.S. Army Maneuver Center of Excellence, Fort Benning, GA, and the other is at NSRDEC, Natick, MA.

The organizations that have a LEAP are using or are planning to use the test battery for clothing and equipment assessments. The conduct of such studies by different military establishments, using the same test battery and common metrics, is a very positive feature of the LEAP and sets this test battery apart from others, which are generally unique to a specific organization. Use of the LEAP by the different organizations holds promise for the sharing of results in international fora on impacts of service members' body-borne loads on physical performance. International cooperation has the potential for accelerating identification of factors associated with military loads that negatively impact performance and identifying means for improving performance of encumbered service members. Scientific panels sponsored by the North Atlantic Treaty Organization and by The Technical Cooperation Program are already engaged in planning and carrying out data exchanges focused around testing with the LEAP.

Initial testing conducted by the U.S. Marine Corps and by the Canadian military on their respective LEAP systems showed that performance on the battery is affected by variations in clothing and equipment configurations and that the primary measure of obstacle course performance—total time to complete a run of the 10 stations—appears to be repeatable. At this point in LEAP development, it is appropriate to assess the reliability of the measurements taken on the test battery.

Test reliability is defined as the repeatability, consistency, or reproducibility of a measurement or of an individual's performance. The purpose of this study was to establish the reliability of course completion time, the principal dependent measurement of performance on the LEAP obstacle course. In the assessment of reliability of the course completion time

measure, the number of trials needed to achieve statistically stable performance was obtained and selected trials were subjected to statistical analyses of reliability. Analyses were also done to examine changes over trials in times to complete the segments comprising the total obstacle course. In addition, subjective data were acquired by administering the 15-category Borg Rating of Perceived Exertion (RPE) and two thermal scales to the participants before and after each run of the obstacle course. These data were analyzed to identify changes in ratings between the two administrations.

Method. In the fall of 2014, NSRDEC investigators conducted a round of testing on the LEAP with U.S. Army enlisted personnel: a group of seven men and one woman. Testing was suspended for the winter months because of extreme weather conditions and a lack of potential participants. A second round of testing occurred in the spring of 2015, during which data were obtained on another group of Army enlistees: 12 men and two women. Two of the three women withdrew from the study early in testing. The third woman and all 19 men completed all aspects of the study. After investigating for outliers, the data for that woman were removed from the dataset, and the data for the 19 men were retained for analysis. The 19 participants had recently completed Army Advanced Individual Training and were awaiting their first assignments to regular Army units. Their time in service was approximately 6 months.

During the study, a participant was scheduled for one session per day on 5 days. A participant reported to the test site at the same time each day, either in the morning or in the afternoon. Orientation occurred at the first session and testing was conducted at the remaining four sessions. Orientation included instructions for properly executing activities at each station on the obstacle course and traversal of the entire course at 50% and then at 75% of maximal effort. At each of the four testing sessions, a participant completed two traversals of the entire obstacle course at 100% of maximal effort, for a total of eight obstacle course trials.

The protocol employed at each of the four testing sessions was the same. Warm-up exercises were performed, which consisted of repetitions of exercises comprising a preparation drill prescribed as part of the Army's physical readiness training program. Following the warm-up, participants rested until heart rate was below 100 beats$\cdot$min^{-1}, as displayed on a heart rate monitor. The participants then completed the RPE and the Thermal Comfort and the Thermal Sensation Scales and began a run of the course. Participants were directed to traverse the course as quickly as possible expending 100% of maximal effort, adhering to proper procedures, and without risking injury. Timers located on the course were used to record time to complete the entire course and times to complete course segments. Immediately upon course completion, the RPE and the thermal scales were again administered. After a rest break of approximately 1 hr, participants performed repetitions of the warm-up exercises and then rested until heart rate as displayed on the heart rate monitor was below 100 beats$\cdot$min^{-1}. At that time, participants moved to the start of the obstacle course where they completed the RPE and the thermal scales and proceeded on a second run of the obstacle course. The participants were again instructed to expend 100% of maximal effort to complete the course quickly. After the second run, the RPE and the thermal scales were administered and the participants were released from testing. Throughout the study, participants wore Army duty uniforms, gloves, and ballistic protective helmets. The total weight of these items was approximately 10.5 lb (4.8 kg). Participants did not carry any additional equipment, such as a military fighting load, and a weapon was not used.

Results. Prior to carrying out statistical tests of reliability of the LEAP obstacle course completion time measure, a repeated measures analysis of variance (ANOVA) was done on the course completion times for the participants' eight trials. A significant Trials main effect was obtained ($p < .001$). Post hoc tests in the form of the Tukey honestly significant difference (Tukey *a*) procedure revealed a gradual decrease in completion times as participants progressed through the eight trials. Further, the post hoc tests indicated that Trials 5 and 6 were the two earliest trials that reflected stabilization of course completion times. Therefore, the data from these two trials were selected for analyses of reliability of the course completion time measure and the data for the other trials were set aside.

An intraclass correlation coefficient (*ICC*) was computed on the data for the two selected trials to estimate relative reliability (i.e., the consistency of the position or rank of individuals in the study sample relative to each other based on their respective obstacle course completion times). The value obtained, *ICC*(2,1) = .93, demonstrated a high degree of relative reliability of the course completion time measure. Further, the 95% confidence interval obtained for the *ICC* was 95% CI [.83, .97]. Absolute reliability (i.e., the consistency of an individual's scores from trial to trial) was also assessed by calculating the standard error of measurement (*SEM*) and the 95% limits of agreement (LOA). The *SEM* calculation yielded a value of 7.21 s, which is 3.13% of the grand mean of the times for the two selected trials. These relatively low values indicate a high degree of absolute reliability for the course completion time measure. The LOA value calculated was 19.97 s, which is an estimate of the random error component of the observed measurements. The 95% LOA obtained is expressed as -0.78 s ± 19.97. Thus, it can be predicted that, in 95% of the instances in which an individual performs two trials, course completion time on the retrial will not be more than 20.75 s (95% CI [-29.26, -12.25]) slower than the time on the previous trial and will not be more than 19.20 s (95% CI [10.70, 27.71]) faster than the time on the previous trial.

In addition to assessing the reliability of the time to complete the entire LEAP obstacle course, times to complete individual segments comprising the course were analyzed using ANOVAs, with Tukey post hoc tests being carried out as appropriate. The raw data were the participants' completion times for a given course segment on each of the eight trials. In the main, a course segment corresponded to an obstacle station on the course. The ANOVA performed on the Agility Run was the only analysis that failed to yield a significant main effect of Trials ($p > .05$). Analysis of the Tunnel & Hatch obstacle yielded a significant main effect of Trials, but the post hoc tests did not reveal significant differences ($p > .05$) among the trials. In all other analyses, significant differences among trials were obtained in the post hoc tests. In general, the analyses of course segments revealed that times were faster for the later trials than for the initial ones, reflecting the findings from the analysis of time to complete the entire obstacle course. Also reflecting the analysis of course completion time, the decrease in segment times was gradual, with a number of instances in which the times on the intermediate trials did not differ significantly ($p > .05$) from the times for the earlier or the later trials.

In the analysis of the RPE data, *t* tests for small, correlated samples yielded a significant difference ($p < .05$) between the pre-run and the post-run ratings for each trial. The mean ratings prior to the obstacle course run had low values (*No exertion at all*); the mean values after the run were higher, in the mid-range of the scale (*Somewhat hard* to *Hard*). For the Thermal Comfort and the Thermal Sensation Scales, Wilcoxon signed ranks tests also yielded significant differences ($p < .05$) between the pre-run and the post-run ratings for each trial. The medians for the five-point Thermal Comfort Scale increased from 1 (*Comfortable*) to 2 (*Slightly uncomfortable*). On the Thermal Sensation Scale, which ranged from 0.0 (*Unbearably cold*) to 8.0 (*Unbearably hot*) in increments of 0.5, the medians increased from 4 (*Neutral*) to 5 (*Warm*).

Conclusions. Practice is required before consistent performance can be achieved on the obstacle course portion of the LEAP. The data acquired in this study indicate that times to complete the course stabilized after four traversals under conditions in which participants were instructed to expend 100% of maximal effort to cover the entire course as quickly as possible. The data also indicate that, once stable performance is achieved, time to complete the obstacle course is a measurement with high relative and absolute reliability. Examination of times to traverse the individual segments comprising the obstacle course revealed performance improvements with repeated trials. Compared with the first trial, some obstacles were completed as much as 14% faster by the point at which times to complete the entire course had stabilized. The magnitudes of the changes that occurred in performance of individual course segments reinforce the need for subjects to practice running the course in order to achieve consistent times and to ensure that the high reliability of the obstacle course completion time measure is not compromised. Contrasts of RPE and thermal scale ratings given by the participants before and after each run of the obstacle course revealed increases in perceived exertion and decreases in thermal comfort after a course traversal.

Recommendations. A number of traversals of the LEAP obstacle course at 100% of maximal effort are required to achieve stable performance. In this study, statistically consistent course times were obtained after the fourth traversal. These findings emphasize the necessity for investigators using the LEAP obstacle course to conduct a similar number of practice trials before beginning formal testing of subjects. A very positive aspect of the LEAP test battery is its use by a number of organizations in this country and abroad. Formal exchanges of data under the sponsorship of the North Atlantic Treaty Organization and The Technical Cooperation Program should continue in order to take full advantage of findings from the organizations conducting testing with the LEAP.

RELIABILITY ANALYSIS OF TIME TO COMPLETE THE OBSTACLE COURSE PORTION OF THE LOAD EFFECTS ASSESSMENT PROGRAM (LEAP)

INTRODUCTION

This report documents a study carried out on a 10-station obstacle course that is part of a new battery of tests developed for assessing the effects of clothing and equipment on the performance of military personnel. The battery is referred to as the Load Effects Assessment Program (LEAP). The study was conducted at the U.S. Army Natick Soldier Research, Development and Engineering Center (NSRDEC), Natick, MA, between October 2014 and June 2015. The purpose of the study was to establish the reliability of course completion time, which is the principal dependent measurement of performance on the LEAP obstacle course. In the assessment of reliability of this measure, the number of trials needed to achieve statistically stable performance was obtained and selected trials were subjected to statistical analyses of reliability. Analyses were also done to examine changes between trials in times to complete the segments comprising the total obstacle course.

In 2009, the U.S. Marine Corps Systems Command (Quantico, VA), began work to develop a performance battery for testing the physical capabilities of Marines on a variety of mission-relevant tasks as affected by the military clothing and equipment being worn and carried. A major impetus for undertaking the work was the concern that all the clothing and equipment items comprising the heavy loads borne by dismounted military personnel operating in Iraq and Afghanistan—such as personal protective equipment (PPE), weapons, ammunition, rations, and water—were negatively affecting accomplishment of tactical maneuvers and mission critical activities (Task Force Devil Combined Arms Assessment Team, 2003). The PPE that was part of combat loads included ballistic protective vests with plates, collars, groin protectors, and extremity armor. Some service members maintained that, although the purpose of wearing the PPE was to decrease the likelihood of becoming a casualty, the protective equipment so impaired their mobility that vulnerability to enemy action actually increased.

The Marine Corps Systems Command determined that an objective method was needed for quantifying effects of a service member's load on execution of combat-relevant movements and tasks (Richter, 2014). It was envisioned that the method would be applied to current components of service members' loads to identify the relative contributions to performance impairment of the load characteristics of weight, bulk, and rigidity. It was also envisioned that the method would yield information to guide design of future clothing and equipment, particularly with regard to trade-offs between protection, performance, and survivability (Bossi et al., 2014).

Through an iterative process involving consultation with subject matter experts experienced in combat mobility requirements, tasks were identified that were judged to be critical in the areas of mobility, lethality, and survivability of dismounted troops (Kelly et al., 2014). The Marine Corps Systems Command, in association with HumanSystems® Inc. (HSI; Guelph, Ontario, Canada), devised a test battery aligned with the critical tasks. The battery is referred to as the LEAP. The LEAP consists of a 10-station obstacle course, along with rifle firing, vertical jump, and weight transfer stations (Figure 1). There is also an area, the questionnaire kiosk, which is set aside primarily for the administration of questionnaires to the subjects once they complete the LEAP. The components of the LEAP were designed to be transportable so that the system can be moved from one location to another.

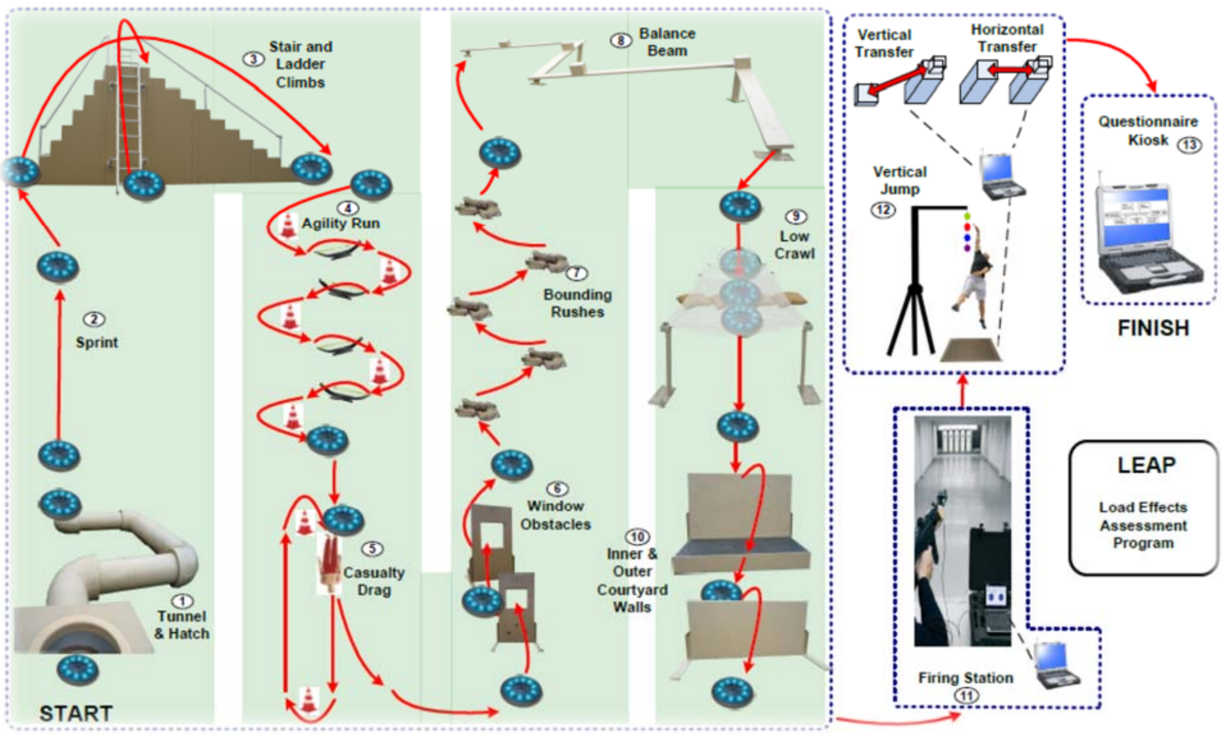


Figure 1. Layout of the LEAP. The discs indicate timer locations.

For the obstacle course portion of the LEAP, the individual being tested is directed to exert 100% of maximal effort to complete the course as quickly as possible, employing instructions in proper procedures while traversing the obstacles. Appendix A contains descriptions of the obstacles in the sequence in which they are traversed. Timers, as shown by the black and blue discs in Figure 1, are located at the start and the end of the course to obtain a measurement of total course completion time. Additional timers are placed along the course to obtain times to complete individual obstacles and times to transition between obstacles. The activities involved in traversing the 10 stations of the course include running, crawling, climbing, vaulting, pulling, and balancing.

At the rifle firing station (Figure 1), a ST-2000 Expert System weapon simulator from FN America, LLC (McLean, VA), is used to test and record firing accuracy. An E-type silhouette target is placed 150 ft (45.7 m) from the firing line. A FN optical unit, mounted on a rifle or a carbine, emits an infra-red LED light onto the target and the FN software converts this to scores upon trigger pull. As used in testing on the LEAP, an individual fires five shots in succession from each of three positions, the kneeling, the standing, and the prone unsupported firing positions (Department of the U.S. Army, 2008), for a total of 15 shots. A shot is taken upon a verbal command from the tester. The accuracy (i.e., distance from the target center) and the precision (i.e., shot dispersion) are calculated and recorded by the software.

The vertical jump station is equipped with a pressure sensor mat and a tall tripod (Figure 1). Tennis balls are hung at different heights above the ground from an arm extension attached to the tripod. The individual being tested is instructed to stand on the sensor mat and then to jump up as high as possible, reaching up above the head to touch a ball at the greatest height possible. Software is used to calculate jump height from time off the sensor mat. Three jumps are performed in succession, with the individual stepping off the mat between jumps.

Two tasks are performed at the weight transfer station (Figure 1). For the first task, the horizontal transfer, the individual being tested moves an ammunition can weighing 30 lb (13.6 kg), which is located on top of a 4-ft (1.20-m) high platform, to a platform at the same height located 4 ft (1.20 m) from the first and then returns the can to its initial position. On the second task, the vertical weight transfer, the individual being tested must lift the 30-lb (13.6-kg) ammunition can from about 6 in. (15 cm) above ground level, place it on a platform 5.7 ft (1.73 m) high and at a distance of 4 ft (1.20 m) from the first, and then return the can to its original location. Individuals being tested perform six trials on each transfer task. They are instructed to work continuously and as quickly as possible to complete all six trials. Time to complete each trial and cumulative time to perform a set of six trials are recorded.

The rifle firing, the vertical jump, and the weight transfer tasks can be performed individually and independently of the obstacle course portion of the LEAP to study effects of clothing and equipment. However, the principal purpose for their inclusion in the LEAP is to serve to assess fatiguing effects of completing the obstacle course portion under various clothing and equipment conditions. In this context, the rifle firing, the vertical jump, and the weight transfer tasks are performed immediately before and immediately following a run of the obstacle portion of the LEAP and the respective scores before and after a run of the obstacle course are compared.

After the LEAP was delivered to the U.S. Marine Corps, HSI fabricated systems, which are highly similar to the U.S. Marine Corps version, for Defence Research and Development Canada-Toronto, Canadian Department of National Defence, and for the Defence Science and Technology Organization, Australian Department of Defence. More recently, HSI fabricated two systems for the U.S. Army. Like the U.S. Marine Corps and the Canadian and Australian military, the U.S. Army is concerned about the effects of heavy body-borne loads on mission accomplishment and combat effectiveness. The Product Director Soldier Systems Integration of the U.S. Army’s Program Executive Office Soldier has formulated a program to identify factors associated with current Army clothing and equipment items that negatively impact Soldiers’ performance and to investigate changes that may result in improved performance. The LEAP was identified as a test battery for use in obtaining the quantitative data needed to carry out the program.

The obstacle course is a major portion of the LEAP test battery. There are other obstacle courses at a number of sites that have been used for research into the effects of military clothing and equipment, including load-carriage equipment. The designs and layouts of the courses differ.

Some of these courses are permanently installed outdoors (Brainerd & Bruno, 1985; Hasselquist et al., 2013); others are set up indoors and disassembled when not in use (Hasselquist, Bensel, Corner, & Gregorczyk, 2012; Pandorf et al., 2002; Stevenson, Reid, Bryant, Pelot, & Morin, 2001). The course length, the number and type of obstacles, and the distance between obstacles vary. Thus, the extent to which the findings from a study on one of these courses can be generalized to other courses is not known. Therefore, the use of the LEAP as a common test method by U.S. military services and by foreign military establishments is an important step. It holds promise for the sharing of results in international fora on impacts of service members' loads on performance and for the identification of means for reducing negative performance effects. Scientific panels sponsored by the North Atlantic Treaty Organization and by The Technical Cooperation Program are engaged in planning and carrying out such data exchanges (Billing & Fordy, 2015).

One of the two U.S. Army LEAP systems was delivered to the U.S. Army Maneuver Center of Excellence, Fort Benning, GA, and the other to NSRDEC[1]. These U.S. Army systems are again highly similar to the U.S. Marine Corps and the Canadian and Australian military versions, having the same components and layout and entailing the same activities by the subjects. However, as new systems have been fabricated, some changes have been made to increase durability of the components and to facilitate assembly and disassembly of the systems (Kelly et al., 2014).

Initial testing conducted by the U.S. Marine Corps (Tack, Kelly, Richter, & Bray-Miners, 2012) and by the Canadian military (Bray-Miners & Kelly, 2013) on their LEAP systems (i.e., MC-LEAP and CAN-LEAP, respectively) showed that performance on the battery is affected by variations in clothing and equipment configurations. In the Marine Corps and the Canadian tests, military volunteers completed the LEAP a number of times, wearing different components of protective clothing and equipment. For example, eight configurations, ranging in mass from 21.8 to 79.0 lb (9.9 to 36.0 kg), were included in the Marine Corps test (Tack et al., 2012). The lightest configuration consisted of field clothing, a helmet, and an M-16 rifle. The heaviest configuration also included a ballistic protective vest with front, back, and side plates and an assault load of ammunition, grenades, field radio, hydration system, and a ballistic protective undergarment.

The means for the eight clothing and equipment configurations tested by Tack et al. (2012) generally showed that total course completion time and individual obstacle times increased as the weight of the items borne on the body increased. Statistical analyses carried out on the data yielded a number of significant differences among the configurations (Tack et al., 2012). Analyses of rifle firing, vertical jump, and weight transfer data also yielded significant differences between pre- and post-obstacle course performance.

In addition to including test conditions that differed in the clothing and equipment items worn, Tack et al. (2012) included two conditions in the U.S. Marine Corps testing that were comprised of identical items in order to check whether these conditions would yield similar

[1]Each LEAP fabricated to date has been assigned a unique identifier: U.S. Marine Corps — MC-LEAP; Defence Research and Development Canada — CAN-LEAP; Defence Science and Technology Organization — AUS-LEAP; U.S. Army Maneuver Center of Excellence — LEAP-A1; NSRDEC — LEAP-A2.

course and individual obstacle completion times. These two conditions were not analyzed separately, but were included in the analyses performed on all eight configurations tested (Tack et al., 2012). Post hoc analyses did not reveal statistically significant differences between the two identical conditions. These results are suggestive of the repeatability of the obstacle course time measures.

DuCharme et al. (2014) tested both intra- and inter-individual repeatability of individual obstacle completion times using the CAN-LEAP. They obtained data on a group of Canadian military men who completed the obstacle course portion of the LEAP twice in one week wearing the same clothing and equipment configuration on both occasions. Expressing the time to complete each obstacle as a percentage of total course completion time, DuCharme et al. (2014) found that the percentages of time spent on the individual obstacles during the two trials were highly similar. They also tested two separate groups of Canadian military men. The two groups wore the same clothing and equipment as they ran the LEAP obstacle course. Again expressing the time to complete each obstacle as a percentage of total course completion time, DuCharme et al. (2014) found that the two groups were highly similar in the percentages of time spent on the individual obstacles.

As was the case in the results obtained by Tack et al. (2012), the work of DuCharme et al. (2014), using the measure of percentage of course completion time, suggests that performance on the LEAP obstacle course is repeatable. Further, results reported by Tack et al. (2012) on the U.S. Marine Corps testing and by Bray-Miners and Kelly (2013) on the Canadian military testing are promising insofar as they suggest that performance on the LEAP is sensitive to manipulations of the gear worn and carried. At this point in data acquisition using the LEAP, it is appropriate to assess the validity and the reliability of the measurements taken on the test battery. Validity is generally defined as the degree to which a test measures what it purports to measure (Brown, 1996). In devising the LEAP, emphasis was placed on including components in the test that experts judged to be critical to success in military operations (Kelly et al., 2014). Thus, it can be said that the face validity of the LEAP is high. Construct validity of the LEAP has yet to be established.

Test reliability is defined as the repeatability, consistency, or reproducibility of a measurement or of an individual's performance (Atkinson & Nevill, 1998; Hopkins, 2000). Atkinson and Nevill (1998) maintained that reliability of a new measurement tool should be assessed before its validity because, if values obtained from repeated administrations of the test are not adequately consistent, the tool will never be valid. Further, it is important to first establish that consistent baseline performance can be achieved before introducing experimental manipulations or interventions that are aimed at altering performance on the test.

Pandorf et al. (2003) and Spiering et al. (2012) conducted studies on the reliability of a number of physical performance tests devised to assess service members' readiness for occupational and combat-related duties. These authors maintained that, although elements of the military services in this country and in foreign military establishments use various test batteries to measure an individual's military readiness (Ayoub, Jiang, Smith, Selan, & McDaniel, 1987; Rayson, Holliman, & Belyavin, 2000; Stevenson et al., 1992), only some of the tests have been evaluated for reliability. Thus, in some instances, there may not be evidence that a test is a good measure of physical proficiency. The LEAP is coming into wider use as a test battery for obtaining quantitative measurements of the effects of loads on service members' physical performance. It is an appropriate time, therefore, to assess the reliability of LEAP performance measurements. The study reported here was conducted primarily to establish the test-retest reliability of the course completion time measurement taken on the obstacle course portion of the LEAP.

One of the tests investigated by Pandorf et al. (2003) was an obstacle course; the reliability of the course completion time measure was studied. According to Pandorf et al. (2003), their study was the first to address reliability of performance on an obstacle course. The course used by Pandorf et al. (2003) was different than the LEAP. It was shorter, being comprised of six stations, rather than 10, and it did not include some physical activities that are part of the LEAP course, such as climbing and descending stairs and a ladder and dragging a 180-lb (81.6-kg) manikin. Participants in the study conducted by Pandorf et al. (2003), who were 10 U.S. Army enlisted men, executed two runs of the course at each of four sessions. Only one session was conducted on any one day of testing. Analyses of course completion time indicated that performance improved significantly from the first to the second session and again from the second to the third session. However, there was no significant improvement from the third to the fourth session. Pandorf et al. (2003) concluded that achieving stable performance on the obstacle course they used required four runs of the course scheduled over two sessions, with two course traversals at a session and each session being conducted on a separate day.

The testing schedule established by Pandorf et al. (2003) was adopted in designing this current study. That is, participants completed two trials on the LEAP obstacle course at each of four sessions and there was only one session on any day of testing. In the assessment of the reliability of the course completion time measure, the number of trials needed to achieve statistically stable performance was obtained. This information is critical in designing future studies using the LEAP obstacle course to investigate the performance effects of the loads borne by military personnel. Times to complete individual obstacles along the course and the transitions between obstacles were also recorded in this testing. These data were obtained to document the elapsed times associated with course segments that comprised the total course completion times and to examine changes in times with repeated trials on the course.

Prior to the start of this study, personnel from HSI provided a week of training on the LEAP to NSRDEC investigators on site at the U.S. Army Soldier Systems Center, Natick, MA. The LEAP tests were set up and study investigators were trained on all aspects of administration of the LEAP test battery and the recording of data. Pilot subjects were also run through the LEAP tests and sample data were obtained for analysis. Personnel from HSI provided a LEAP manual (Kelly, 2015) that was used in the training, along with the personal instructions in course set up, administration, and scoring. In addition, a set of subjects' instructions, reflecting refinements of information in the HSI manual, was provided by researchers at Defence Research and Development Canada (L. Bossi, personal communication, September 19, 2014). In addition to involvement in designing and fabricating the LEAP, HSI personnel had previously participated in U.S. Marine Corps (Tack et al., 2012) and Canadian (Bray-Miners & Kelly, 2013) studies conducted on the LEAP. Given the knowledge and experience of the HSI personnel, there is some assurance that the NSRDEC investigators administered the LEAP in the manner intended by the developers and followed procedures applied in the earlier studies (Bray-Miners & Kelly, 2013; Tack et al., 2012).

In the fall of 2014, NSRDEC investigators conducted a round of testing on U.S. Army enlisted personnel to acquire data on reliability of the obstacle course portion of the LEAP. Testing was suspended for the winter months because of extreme weather conditions and a lack of potential participants. A second round of testing occurred in the spring of 2015, during which additional data for assessment of reliability were obtained on another group of U.S. Army enlistees. Each test participant executed eight trials on the 10-station LEAP obstacle course. Participants were not tested on the rifle firing, the vertical jump, or the weight transfer stations. The method used, the statistical analyses applied to the data, and the results of the reliability assessment are presented in this report.

METHOD

Participants

Participants in the study were recruited from among U.S. Army enlistees who serve as human research volunteers assigned to Headquarters Research and Development Detachment, NSRDEC. Recruitment was carried out prior to testing that was conducted in October 2014. A second recruitment effort was undertaken prior to another round of testing that was completed in June 2015.

The Soldiers had no previous experience with any aspects of the LEAP. They were informed of the purpose of the study, the nature of the test conditions, the risks associated with the testing, all procedures affecting a volunteer's well-being, and a volunteer's right to discontinue participation at any time without penalty.

The study was reviewed by the NSRDEC Human Subjects Research Determination Panel in accordance with the Federal Policy for the Protection of Human Subjects, U.S. Department of Defense, 32 Code of Federal Regulations Part 219. Approval for the study was granted under the NSRDEC Assurance for the Protection of Human Research Subjects, DoD A20124.

Before volunteers could begin participation in the study, they underwent medical screening, including a physical examination and clinical review of their medical records, with an emphasis on the musculoskeletal system. Individuals with a history of back problems, including herniated discs or previous orthopedic injuries that limit the range of motion about the shoulder, hip, knee, or ankle joint, were excluded from participation, as were any women who reported that they were pregnant.

In estimating the sample size for this study, the test-retest correlation coefficient obtained by Pandorf et al. (2003) from analysis of completion times on their obstacle course was used. The correlation coefficient equaled .92. The LEAP course and its completion time measure are new areas of investigation. Cohen (1988) recommended that a small effect size (d) be selected in these situations and, therefore, d was set equal to .2. Applying Hopkins' (2000) formula for estimating sample size, calculation of the number of participants needed in this study yielded a value of 16. After the two recruitment efforts and medical screening, 19 men and three women were available to serve as volunteers in the study. Allowing for the possibility of individuals terminating their participation before completing all aspects of testing and loss of data for other reasons, it was determined that the number of volunteers was sufficient and no additional recruitment efforts were undertaken.

Volunteers were tested in two groups, one group in October 2014 and the other in June 2015. The first group was comprised of seven men and one woman. The woman withdrew from the study early in testing; the seven men completed all aspects of the testing. There were 12 men and two women in the second group. One of the women withdrew from the study early in testing. The remaining woman and the 12 men completed the study. After investigating each of the two groups for outliers, the data of the remaining woman were removed from the dataset. This woman's times to complete the obstacle course were substantially slower than the men's times. No other outliers were detected.

After removal of the outlier, the obstacle course completion times for the seven men in the first group and the 12 men in the second group were contrasted to determine whether the times for the groups differed significantly ($p < .05$). To carry this out, a *t* test for small uncorrelated samples was performed on the data for each of the eight trials that participants executed. The Bonferroni adjustment was applied. The mean time for the second group was lower than the mean for the first group on each trial, but the differences between groups did not achieve significance ($p > .05$). Therefore, the data of the two groups were combined to obtain a total sample of 19 men.

At the time of testing, the participants had recently completed Army Advanced Individual Training and were awaiting their first assignments to regular Army units. Their time in service was approximately six months. The military occupational specialties (MOSs) of the participants were in combat service support areas, such as radio operator (MOS 25C), wheeled vehicle mechanic (MOS 91B), and supply specialist (MOS 92Y). Summary statistics for the demographic data on the participants are presented in Table 1. The mean stature and weight of the participants listed there are approximately equivalent to the 65th and the 47th percentiles, respectively, calculated from the measurements for 4082 men acquired in the most recent anthropometric survey of U.S. Army personnel (Gordon et al., 2014). Thus, the participants were, on average, somewhat taller and about the same weight as the averages obtained for the large sample of Army men. The information in Table 1 on the physical training (PT) test is self-reported. The most recent PT test listing includes the PT test score, which is the number of points earned out of a possible 300. The minimum number of points required to pass the PT test is 180.

Table 1. *Demographics of Test Participants (N = 19)*

Measure	Mean		*SD*		Minimum		Maximum	
Age, years	23.5		4.5		18.0		34.0	
Stature, in. (cm)	70.0	(177.8)	2.9	(7.4)	60.9	(154.7)	74.6	(189.5)
Weight, lb (kg)	185.9	(84.3)	21.8	(9.9)	157.5	(71.4)	236.0	(107.0)
Most Recent PT Test								
Score (points)	254.5		23.4		207		298	
Push-ups (number in 2 min)	58.1		11.6		42		80	
Sit-ups (number in 2 min)	68.7		8.8		52		84	
2-mi Run Time (min)	14.0		0.8		12.8		15.5	

Materials

Attire Worn for Testing

Throughout the study, participants wore Army duty uniforms, gloves, and ballistic protective helmets. The weights of the helmet and the components of the uniform are listed in Table 2. The weights are for size medium items. The investigators supplied duty gloves and the helmets; participants wore their own uniform items. Participants did not carry any additional equipment, such as a military fighting load, and a weapon was not used. It is expected that participants will carry a weapon for most testing done on the LEAP obstacle course. For the purpose of assessing LEAP obstacle course reliability, it was decided to avoid the arbitrary selection of a weapon and devising tentative instructions for the participants on the way to handle the weapon during obstacle traversals (e.g., when to sling the weapon, when to hold the weapon at the low ready). Therefore, a weapon was not used in this study.

Table 2. *Items Worn for Testing and Their Weights in a Size Medium*

	Weight	
Item	lb	kg
T-shirt and shorts	0.80	0.36
Combat boots and socks	4.21	1.91
Army Combat Uniform shirt and trousers	1.78	0.81
Duty gloves	0.24	0.11
Advanced Combat Helmet	3.48	1.58
Total	10.52	4.77

LEAP Obstacle Course and Course Equipment

The obstacle course portion of the LEAP consists of 10 stations. The version of the course used here (LEAP-A2) was fabricated by HSI (Guelph, Ontario, Canada). For this study, the course was set up outdoors on a flat, grassy area at the U.S. Army Soldier Systems Center, Natick, MA. The course layout is shown in Figure 2. The course was set up according to the directions provided by HSI (Kelly, 2015). The obstacles are described and illustrated in Appendix A.

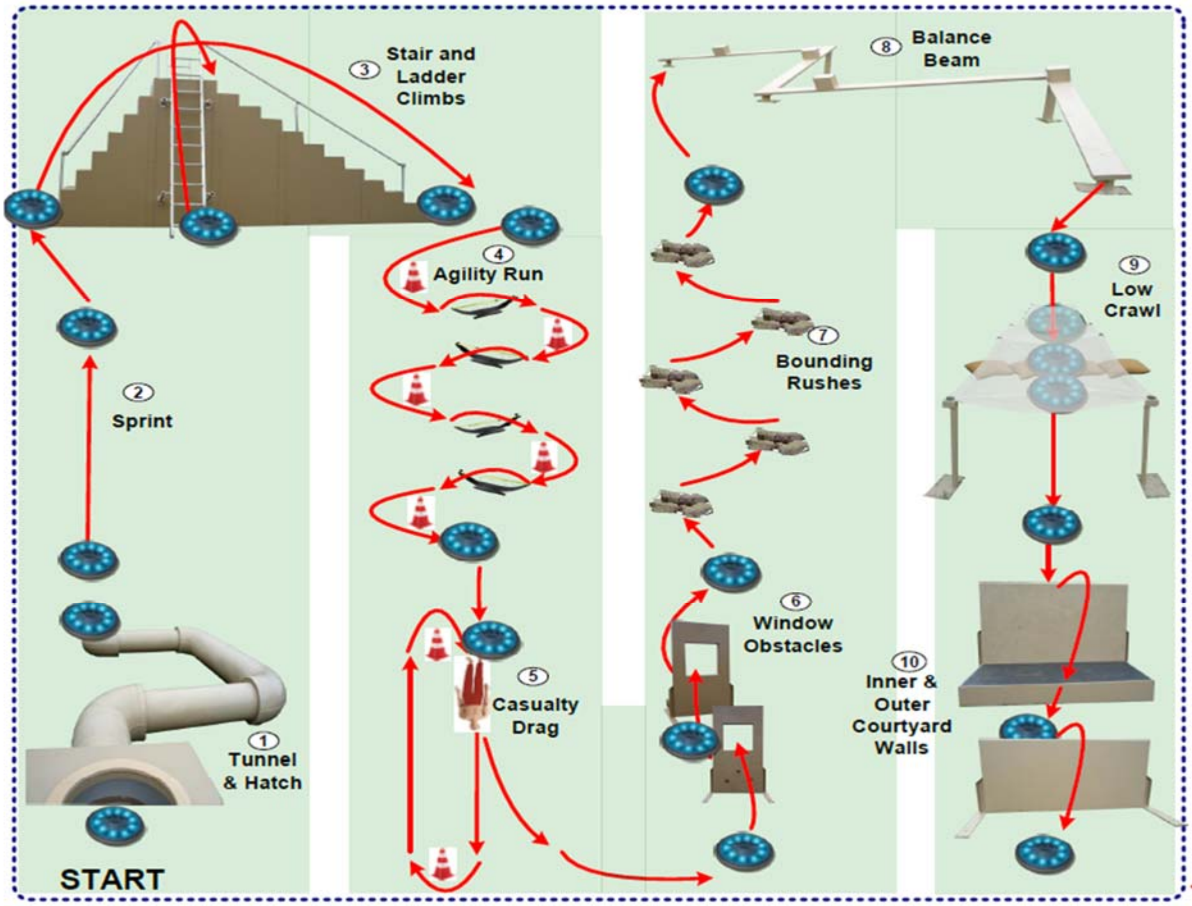


Figure 2. Layout of obstacle course portion of the LEAP. The discs indicate timer locations.

Timers (FitLight Timing System, FitLight Sports Corp., Aurora, Ontario, Canada) located at the beginning and the end of the obstacle course were used to obtain time to complete a run of the entire course. Additional timers (FitLight Timing System) were placed along the course to record times to complete individual obstacles and to transit between obstacles. A listing of the segments of the course for which times were obtained is in Appendix B. Each run of the course was also recorded on a video camera (Hero3, GoPro, Inc., San Mateo, CA). The clock on the Windows™ Media Player, in conjunction with the playing of recorded video, was used as a back-up if there was a timer malfunction.

The NSRDEC investigators followed the guidance and training provided by HSI (Kelly, 2015) and the subjects' instructions furnished by Defence Research and Development Canada (L. Bossi, personal communication, September 19, 2014) in carrying out LEAP testing in this study.

Ancillary Equipment

Body dimension measurements were taken on the test participants following the procedures employed in the most recent anthropometric survey of U.S. military personnel (Hotzman et al., 2011). Men wore uniform trousers and women wore trousers, a bra, and a t-shirt while the measurements were made.

Participants wore a Forerunner® 210 heart rate monitor (Garmin International, Inc., Olathe, KS) on their wrists throughout testing. The monitors were used by the investigators to check participants' heart rates between trials at each session. Heart rate data were not recorded in this study.

The 15-category Rating of Perceived Exertion (RPE) scale devised by Borg (1970, 1982) was administered to the participants immediately before and after they completed a run of the obstacle course to assess the perceived exertion associated with traversals of the course. A copy of the RPE is in Appendix C. The RPE is widely used by exercise physiologists in a variety of research and clinical settings (Noble, 1982; Pandolf, 1982) and has often been employed in studies of body-borne loads (Goslin & Rorke, 1986; Patton, Kaszuba, Mello, & Reynolds, 1990). In developing this instrument, Borg (1970) structured the 15-point scale such that ratings would increase linearly with the heart rate of healthy middle-aged men performing moderate to hard work on a bicycle ergometer. Borg (1982) reported high, positive correlations (r = .80 to .90) of heart rate with RPE.

The rating given by an individual using the RPE scale is considered to represent a value along a perceptual continuum of effort and to be an estimate of the individual's physical work capacity (Borg & Noble, 1974). It has been suggested that local factors, related to feelings of strain in the exercising muscles and joints, and a central factor, related primarily to sensations from the cardiopulmonary systems, are reflected in an individual's rating. There is evidence that the characteristics of the physical activity determine whether local or central factors dominate an individual's overall rating of exertion (Pandolf, 1978; Robertson et al., 1982). Based upon investigations in which individuals walked with external loads on the body, researchers suggested that local physiological factors, not central factors, dominate the overall perception of exertion when body-borne loads are being carried (Goslin & Rorke, 1986; Patton et al., 1990). These could include feelings of increased muscle tension or sensations from skin, tendons, and ligaments.

Immediately before and after completing a run of the obstacle course participants were also asked to rate their thermal status using two category rating scales, which are presented in Appendix D. One of these rating instruments was the Thermal Comfort Scale. This is a 5-point scale, ranging from 1 (*Comfortable/OK*) to 5 (*Intolerable*). Respondents select one number to indicate how comfortable or uncomfortable they feel. The second of these rating instruments was the Thermal Sensation Scale (Young, Sawka, Epstein, Decristofano, & Pandolf, 1987), a scale on which respondents select one number, ranging from 0.0 (*Unbearably cold*) to 8.0 (*Unbearably hot*) in increments of 0.5, to indicate how hot or cold they feel.

Procedure

A participant was scheduled for one session per day on five days. For most participants, the sessions were conducted on consecutive days. However, several participants missed one session and their test days were interrupted by a weekend break. Orientation occurred at the first session and testing was conducted at the remaining four sessions. Participants were directed to abstain from heavy and moderate exercise and alcohol consumption throughout their term in the study. A participant was tested at the same time each day, either in the morning or in the afternoon. At the orientation, participants completed a demographic questionnaire and were briefed on the study. The body measurements were taken and the appropriate size helmet and gloves were identified for each participant. Participants were also familiarized with the LEAP obstacle course.

For course familiarization, participants first walked through the course with a U.S. Army non-commissioned officer who presented instructions in accordance with guidance provided by HSI (Kelly, 2015) and refined by researchers at Defence Research and Development Canada (L. Bossi, personal communication, September 19, 2014). The non-commissioned officer demonstrated the proper technique in traversing each station on the course. During this course orientation, participants completed the Window #1 & Window #2 station and the Outer & Inner Courtyard Walls station three times to have the opportunity to devise procedures (e.g., foot placement, hand placement) for successfully traversing these obstacles. Participants also performed the Stair & Ladder Climbs to learn the proper order of execution of the activities involved in this station. In addition, participants had the opportunity to do the Balance Beam once. The orientation session concluded with participants traversing the entire course at 50% and then at 75% of maximal effort.

The protocol employed at each of the four testing sessions was the same. Warm-up exercises were performed, which consisted of 10 repetitions of exercises comprising a preparation drill prescribed as part of the Army's physical readiness training program (Department of the U.S. Army, 2012). The 10 exercises, which the participants were experienced in performing as part of their daily physical training, are described in Appendix E. Following the warm-up, participants rested until heart rate was below 100 beats$\cdot$min^{-1}, as displayed on the Garmin monitor. They donned the helmet and gloves and moved to the start of the obstacle course. The participants then completed the RPE (Appendix C) and the Thermal Comfort and the Thermal Sensation Scales (Appendix D) and began a run of the course. The participants were directed to traverse the course as quickly as possible expending 100% of maximal effort, without risking injury, and following instructions in proper procedures while traversing the obstacles.

Immediately upon course completion, the RPE and the thermal scales were administered. After a rest break of approximately 1 hr, participants performed three to five repetitions of the warm-up exercises (Appendix E) and then rested until heart rate as displayed on the Garmin monitor was below 100 beats$\cdot$min^{-1}. At that time, participants donned the helmet and gloves and moved to the start of the obstacle course where they completed the RPE and the thermal scales. A second course run was then completed, with the participants again being instructed to expend 100% of maximal effort to complete the course quickly. After the second run, the RPE and the thermal scales were administered and the participants were released from testing.

Guidelines for protection of the health and safety of participants were established prior to initiation of testing. These included the presence of study investigators on the obstacle course who were available to provide immediate assistance to participants, if required. The number of participants on the course at any one time was limited to one. Investigators also checked the course throughout the testing sessions, removing leaves and other debris that could pose a slipping or tripping hazard.

The determination was made that testing would be cancelled and rescheduled in the case of rain or wet grass. Further, given that testing was scheduled for June and October, months in which hot weather can occur in eastern Massachusetts, U.S. Army guidance for heat stress control was followed (Departments of the U.S. Army and U.S. Air Force, 2003). Specifically, it was determined that testing would be cancelled in the event that wet-bulb globe temperature (WBGT) equaled or exceeded 90 °F (32.2 °C). Personnel of the Safety Office, NSRDEC, monitored WBGT at the test site on days during which high WBGT readings were a possibility. Participants were also provided water and encouraged to drink when not engaged in a run of the obstacle course. During these breaks, participants could not use tobacco products or take caffeine or energy drinks.

During the October 2014 and the June 2015 testing, there were no occurrences of precipitation that required rescheduling of test sessions and the grass surfaces remained dry. Also, WBGT readings did not reach the critical level at which cancellation of a session would have been required.

Aside from the monitoring of WBGT as needed, ambient environmental conditions were not recorded. However, historical weather data were obtained in the form of hourly readings of temperature and percentage of relative humidity (%RH) for the periods of each day during which testing was conducted. These data are summarized in Table 3 and presented in detail in Appendix F. Temperatures increased from morning to afternoon and were lower during the October testing period than during the June testing, typical patterns for the geographic area.

Table 3. *Summary of Climatic Conditions During Test Sessions*

	Session 1		Session 2		Session 3		Session 4	
Time	Temp. (°F)	%RH	Temp. (°F)	%RH	Temp. (°F)	%RH	Temp. (°F)	%RH
				October 2014				
0900 - 1100	51.6 (5.2)	78.0 (8.0)	59.9 (3.9)	96.0 (4.2)	51.7 (3.5)	73.0 (8.8)	46.7 (2.8)	79.0 (5.4)
1300 - 1500	69.9 (1.7)	59.0 (2.2)	72.4 (0.6)	66.3 (3.7)	56.3 (1.1)	55.0 (1.4)	51.5 (1.0)	65.7 (0.9)
				June 2015				
0900 - 1100	71.2 (2.0)	83.7 (3.3)	72.8 (3.1)	61.0 (10.4)	74.8 (2.6)	75.0 (1.4)	74.7 (2.4)	59.7 (5.2)
1300 - 1500	74.9 (1.6)	80.3 (2.9)	81.7 (0.2)	40.0 (2.4)	86.1 (0.4)	56.0 (1.6)	86.5 (2.1)	45.3 (0.5)

Note. Data are the means (with *SD*s in parentheses) obtained from hourly readings of temperature and percentage of relative humidity (%RH) for the time periods indicated. The weather station was located about 0.7 mi (1.1 km) from the test site.

Statistical Analyses

Obstacle Course Completion Time

The approach taken to investigate the reliability of participants' times to traverse the obstacle course portion of the LEAP was that proposed by Atkinson and Nevill (1998). As the first step, a one-way analysis of variance (ANOVA) with repeated measures was carried out [Trials (t = 1-8) x Subjects (n = 1-19)] to identify whether there were differences in course completion time among the trials. A finding of a significant Trials main effect ($p < .05$) was followed by post hoc tests in the form of the Tukey honestly significant difference (HSD; Tukey *a*) procedure, with the significance level again set at $p < .05$. The Tukey HSD procedure entailed pairwise comparisons between trials to determine whether completion times differed significantly.

The results of the Tukey HSD tests were used to identify data for further analysis. The data selected for further analysis were the earliest two trials that the Tukey procedure revealed did not differ significantly ($p > .05$) from each other or from subsequent trials. The absence of a significant difference between the two trials and between these trials and subsequent trials was taken as an indicator that obstacle course completion times had stabilized by this point in testing. Analyses of the reliability of the obstacle course completion time measure were performed on the data for these two trials, and the data for the other six trials that participants carried out were set aside.

Underlying the analytical techniques applied here to assess reliability of the obstacle course completion time measure is the fact that any single observed measurement has two components: a true measure and an error. Summed together, the true measurement and the error equal the observed measurement. The true measure is the genuine value of the thing measured (Guilford, 1956). The error component of a measure is divided into two categories: systematic bias and random error (Nunnally, 1967).

Systematic bias may be positive or negative and tends to be constant. It is sometimes associated with the measuring instrument itself. In this study, for example, course completion time was obtained from a timer positioned at the start of the obstacle course and another at the end. Even if the outputs of the two timers were recorded correctly, there could be systematic bias due to the internal calibrations of the timers themselves. Other possible sources of systematic bias are changes in an individual's motivation or physical state (Hopkins, 2000). Systematic bias is not considered very important in making measurements related to human performance because it is assumed to contribute equally to the mean score of all participants in a study (Nunnally, 1967). Random errors, on the other hand, are important. They are inconsistent, unpredictable, unknown, and due to chance. Thus, it is not possible to know the amount of any observed measurement that is attributable to the true measure and the amount that is attributable to error. Although this is the case for any observed measurement, the situation is particularly germane in establishing reliability of a measurement because reliability concerns the consistency or reproducibility of the observed measurement when the measurement is made repeatedly (Atkinson & Nevill, 1998; Bruton, Conway, & Holgate, 2000; Hopkins, 2000). The statistical approach taken to address the impossibility of obtaining the true reliability of a measurement is to estimate reliability based on the variance of the data.

A long-standing and frequently used statistical technique for estimating reliability is the intraclass correlation coefficient (*ICC*). An *ICC* was calculated for the data obtained in this study. An *ICC* is an estimate of the relative reliability of the measurement being analyzed. It is referred to as an estimate of relative reliability because it addresses the consistency of the position or rank of individuals in the study sample relative to each other based on their respective scores (Weir, 2005). In the context of this study, the *ICC* would reflect the degree to which participants maintain their relative positions from trial to trial in terms of their respective obstacle course completion times. If a number of participants have faster times than the remaining participants on one trial, but slower times on the other trial, it would indicate a lack of consistency and low relative reliability of the course completion time measurement, and low values of the *ICC* would be expected.

For calculation of the *ICC* on the two selected trials, a two-way random effects, single-measure model was used. Following established protocol, this model is designated in the following manner: *ICC*(2,1) (Shrout & Fleiss, 1979). To obtain the *ICC* value, a one-way ANOVA with repeated measures was first carried out on the data for the two selected trials [Trials (t = selected trial 1-2) x Subjects (n = 1-19)]. Mean square terms from the ANOVA were then entered into the formula for the *ICC* two-way random effects model (Weir, 2005). The formula is:

$$ICC(2{,}1) = \frac{MS_{SS} - MS_E}{MS_{SS} + (k-1)MS_E + \frac{k(MS_T - MS_E)}{n}} \quad (1)$$

where,
MS_{SS} is the subjects mean square;
MS_E is the error mean square;
MS_T is the trials mean square;
k is the number of trials;
n is the number of subjects.

An *ICC* typically takes on a value between 0 and 1. The value of the *ICC*, and the relative reliability of the observed measurement, increases as the observed measurement approaches the true measurement (Bruton et al., 2000). A rule of thumb for interpreting *ICC* values is: An *ICC* value close to 1 indicates “excellent” reliability of the measure being analyzed; a value greater than .9 “high” reliability; a value from .8 to .9 “good” reliability; and a value from .7 to .8 “questionable” reliability (Atkinson & Nevill, 1998). Recognizing that the *ICC* value is an estimate of the reliability of a measurement, a 95% confidence interval (95% CI) is often constructed for the *ICC*. The 95% CI is a means of providing a range of values about the *ICC* that 95 times out of 100 is likely to include the true *ICC* value for the population (Eliasziw, Young, Woodbury, & Fryday-Field, 1994).

In recent years, there has been increasing emphasis in the measurement reliability literature on applying statistical tests to estimate absolute reliability, in addition to estimating relative reliability using the *ICC*. Some contributors to the reliability literature have maintained that use of only a single estimate of reliability, such as the *ICC*, leaves gaps in the assessment of how consistent a measurement actually is (Atkinson & Nevill, 1998; Bland & Altman, 1986; Bruton et al., 2000). Absolute reliability reflects the degree to which an individual's scores vary from trial to trial (Weir, 2005). Thus, absolute reliability concerns the consistency of an individual participant's scores, whereas relative reliability concerns the consistency of the position or ranking of an individual's scores from trial to trial relative to the position of the scores of all other individuals in the study sample. In addition to calculating an *ICC* on the data for this study, calculations were carried out using two methods that yield estimates of absolute reliability. These are the standard error of measurement (*SEM*) and the 95% limits of agreement (LOA).

The data entered into the calculation of the *SEM* were for the same two selected trials used in obtaining the *ICC*. The value of the *SEM* was calculated as the square root of the mean square error term from the ANOVA that was used to obtain the *ICC* in Equation 1 (Eliasziw et al., 1994). Specifically,

$$SEM = \sqrt{MS_E} \tag{2}$$

where,
MS_E is the error mean square also used in Equation 1.

Unlike values of *ICC*s, which are dimensionless, *SEM*s are expressed in the actual units of measurement and also can be expressed as a percentage of the grand mean of the scores of all individuals in the study. The *SEM* is the random variation in scores about the true measurement that would be expected when one individual is tested repeatedly under the same conditions. The smaller the *SEM*, the more reliable the measure (Atkinson & Nevill, 1998). Unlike the *ICC*, there are no particular values of the *SEM* that denote "high" reliability or "questionable" reliability of a measure. The investigator must make the decision regarding whether the *SEM* obtained is small enough to declare that the measure is sufficiently reliable. The decision is typically based on considerations related to the situations in which the measure will be used.

The 95% LOA, a statistical method introduced by Bland and Altman (1986) as an indicator of absolute reliability, was calculated in this study from the same two selected trials used in calculation of the *ICC* and the *SEM*. Bland and Altman (1986) applied their method to assess agreement between two different techniques for making blood pressure measurements. Other investigators have extended Bland and Altman's (1986) statistical approach to assessment of test-retest reliability (Atkinson & Nevill, 1998; Bruton et al., 2000; Hopkins, 2000; Pandorf et al., 2003).

Before proceeding to obtain the 95% LOA on the two trials selected from the data of this study, a Pearson product-moment correlation coefficient (r) was obtained to determine whether the selected data were heteroscedastic (i.e., the amount of random error increased as obstacle course completion times increased) or homoscedastic (i.e., no relationship existed between the random error and the values of the completion time measures; Atkinson & Nevill, 1998). The raw input data for calculating r were each participant's mean completion time on the two selected trials and the absolute value of the difference between these trials.

Bland and Altman (1986) maintained that, if data are heteroscedastic, as evidenced by a significant *r* value ($p < .05$), the 95% ratio LOA should be computed. They further maintained that, if data are homoscedastic, as evidenced by an *r* value that does not achieve significance ($p > .05$), the 95% LOA should be computed. The respective formulas are:

$$95\%\ ratio\ LOA = \left[1.96\left(\frac{SD\ of\ the\ difference\ scores}{average\ of\ the\ mean\ values}\right)\right] x\ 100 \qquad (3)$$

$$95\%\ LOA = 1.96(SD\ of\ the\ difference\ scores) \qquad (4)$$

Both LOA approaches (Equations 3 and 4) involve obtaining the difference between an individual's scores and then calculating the *SD* of the differences over all individuals in the study sample. Thus, the LOA is focused on the test-retest differences evidenced by the individual participants. The mean of the differences is also obtained. This is done by taking the difference between an individual's scores and then calculating the mean of the differences over all individuals in the study sample. Bland and Altman (1986) interpreted the LOA as an estimate of the random error component of the observed measurement. The mean of the differences was interpreted by Bland and Altman (1986) as the systematic bias, or the general tendency of measurements to be different between repeated trials in either a positive or a negative direction.

A Bland-Altman plot was prepared as part of the LOA analysis carried out on the data from this study. This type of plot is a simple way to examine the data of each individual participating in a study. It is a graphical representation of the difference between an individual's observed measurements plotted against the mean of that individual's measurements. According to Bland and Altman (1986), given that the true component of a measurement cannot be known, the mean of the individual's observed measurements is the best estimate.

A Bland-Altman plot is also a visual presentation of the findings from the LOA analysis. The findings presented in the plot include a bias line, which represents the value of the systematic bias, and random error lines. The random error lines in the plot represent the 95% LOA expressed as the bias value ± 2 *SD*s. Thus, the 95% LOA method yields the range of values within which the differences in an individual's scores from trial to trial would fall most (95%) of the time, or 95 times out of 100 (Bland & Altman, 1986). Following from this, an interpretation of the high and low values forming the range of the 95% LOA is that, given the number of trials and number of individuals tested in a particular study, when another individual is tested and retested, the retest score has five chances in 100, or a 5% chance, of differing from the score on the first test by an amount exceeding either the high or the low value forming the range (Hopkins, 2000).

Bland and Altman (1986) pointed out that values of bias and the 95% LOA values are estimates and, thus, a different sample of participants may yield different values. They proposed that standard errors and confidence intervals be obtained to determine how precise the estimates are. Therefore, calculations were performed on the data from this study to obtain the standard error of the bias value and the 95% CIs of the high and the low values of the 95% LOA.

Like the *SEM*, there are no particular values of the 95% LOA that denote “high” reliability or “questionable” reliability of a measure. The narrower the range of values, the more reliable the measure is, but, again, the investigator must make the decision regarding whether the range of values is small enough to declare that the measure is sufficiently reliable.

Completion Times for Obstacle Course Segments

The data for obstacle course segments of the LEAP were analyzed to determine whether there were differences in segment completion time among the eight trials completed by each participant. For this, a one-way repeated measures ANOVA was carried out [Trials (t = 1-8) x Participants (n = 1-19)] on each segment. A finding of a significant Trials main effect ($p < .05$) was followed by completion of post hoc tests in the form of the Tukey HSD (Tukey *a*) procedure to determine pair-wise differences between trials, with the significance level again set at $p < .05$.

The RPE and the Thermal Rating Scales

For the RPE data, the participants’ ratings given prior to each of the eight trials on the obstacle course were compared with the ratings given immediately upon course completion by applying *t* tests for small, correlated samples. The significance level was set at $p < .05$ and the Bonferroni correction was applied. The two thermal rating scales were analyzed using the Wilcoxon signed ranks test, a nonparametric test (Siegel & Castellan, 1988). The participants’ ratings given prior to a run of the obstacle course were compared with the ratings given immediately upon course completion. The significance level was again set at $p < .05$ and the Bonferroni correction was applied.

RESULTS

The findings related to analysis of the reliability of time to complete the obstacle course portion of the LEAP are presented first. These are followed by results of the analyses carried out on times to complete segments of the obstacle course. Finally, results of the RPE and the thermal rating scales analyses are presented.

Many of the figures and the tables for obstacle course completion times and course segment times contain results of the Tukey HSD statistical difference tests. Results of these post hoc tests are indicated by upper case letters (e.g., A, B, and C). Trials that do not share the same letter differed significantly in the HSD tests ($p < .05$). Conversely, those trials that share the same letter were not significantly different ($p > .05$). Using Figure 3 as an example of this scheme, one letter, “A”, is associated with Trials 1 and 2; one letter, “B”, is associated with Trial 4. These designations indicate that there was no significant difference in the data for Trials 1 and 2, but that the data for both these trials differed significantly from the data for Trial 4. Similarly, the letters “AB” are associated with Trial 3, which indicates that Trial 3 did not differ significantly from any trials designated with the letter “A” or with the letter “B,” that is, Trial 3 did not differ significantly from Trials 1 and 2 or from Trials 4 through 7.

Obstacle Course Completion Time

The results of the one-way repeated measures ANOVA carried out to identify changes in obstacle course completion time over the eight trials accomplished by each of the 19 participants yielded a significant main effect of Trials, $F(7, 126) = 12.25$, $p < .001$. The trial means calculated over participants, the standard error of the means (*SE*), and the results of the Tukey HSD post hoc tests are presented in Figure 3.

Visual inspection of the means plotted in Figure 3 indicates that course completion time generally decreased from Trials 1 through 8. An exception to the decrease in time from trial to trial was Trial 6, which had a mean that was higher, although by less than 1 s, than the mean for Trial 5 (Figure 3). Comparison of the completion times for the first and the last trials revealed that the mean time for Trial 8 was 10% faster than the mean time for Trial 1. Mean times for Trials 5 and 6 were approximately 7% faster than the mean time for Trial 1.

The Tukey HSD post hoc test results (Figure 3) reflect a gradual decrease in times as the participants progressed through the trials. The slowest course completion times, the times for Trials 1 and 2, were not significantly different from each other or from the time for Trial 3 ($p > .05$), but Trial 1 and Trial 2 times did differ significantly from the times for Trials 4 through 8 ($p < .05$). Trial 3 also did not differ significantly from Trials 4 through 7 ($p > .05$). The time for Trial 8, the fastest trial, did not differ significantly from the times for Trials 5 through 7 ($p > .05$), but was significantly faster than times for Trials 1 through 4 ($p < .05$).

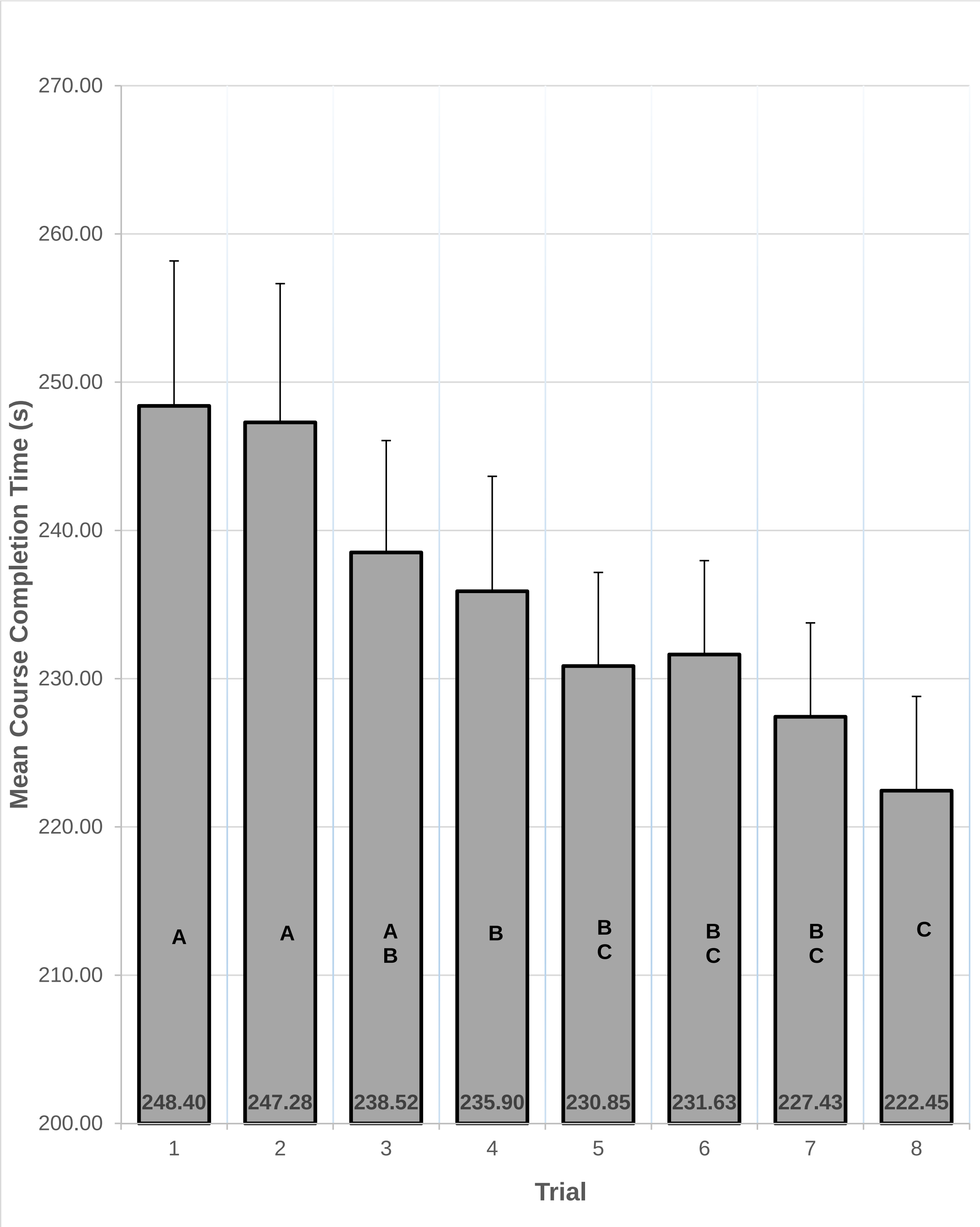


Figure 3. Mean obstacle course completion time for each of the eight trials. The error bars indicate +1*SE*. Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$, $N = 19$).

Based on the findings from the Tukey HSD tests that course completion times from Trial 5 through Trial 8 did not differ significantly ($p > .05$), Trials 5 and 6 were identified as the two, earliest trials that reflected stabilization of completion times. Therefore, the data from these two trials were selected for analyses of reliability of the course completion time measure.

Calculation of the intraclass correlation coefficient [*ICC*(2,1)] on Trial 5 and Trial 6 data yielded a value of .93, demonstrating a high degree of relative reliability of the course completion time measure. Further, the confidence interval obtained for the *ICC* was 95% CI [.83, .97]. The value of the lower limit of the confidence interval, .83, equates with good reliability and the value of the upper limit, .97, equates with excellent reliability.

With regard to the *SEM*, one of the two calculations performed to assess the absolute reliability of the course completion time measure, the Trial 5 and Trial 6 data yielded a value of 7.21 s, which is 3.13% of the grand mean of the times for these two trials. These relatively low values indicate a high degree of absolute reliability for the course completion time measure.

Figure 4 is the Bland-Altman plot of the data on Trials 5 and 6 for the 19 participants. Each data point is the time difference between Trials 5 and 6 for a participant plotted against the participant's mean for these two trials. Each participant's mean over the two trials is cited by Bland and Altman (1996) as the best estimate of the true value of a participant's course completion time.

In addition to calculating the *SEM*, the 95% LOA was calculated on the data for Trials 5 and 6 as another means of assessing the absolute reliability of the course completion time measure. A Pearson product-moment correlation coefficient was first obtained for the data of the two trials to determine whether these data were heteroscedastic or homoscedastic. The correlation coefficient was not significant, $r(17) = .40$, $p > .05$, indicating that the data were homoscedastic. Therefore, the 95% LOA was computed using Equation 4, which entailed multiplying the *SD* of the difference between participants' course completion times on Trials 5 and 6 by 1.96.

Application of Equation 4 yielded a value for the LOA of 19.97 s, which is an estimate of the random error component of the observed measurement (Bland and Altman, 1996). The mean of the difference between participants' completion times on Trials 5 and 6, the value referred to by Bland and Altman (1996) as the systematic bias component of the observed measurement, equaled -0.78 s (95% CI [-5.69, 4.13]). The 95% LOA is, therefore, expressed as -0.78 s ± 19.97. Thus, it can be predicted that, in 95% of the instances in which an individual performs two trials, course completion time on the retrial will not be more than 20.75 s (95% CI [-29.26, -12.25]) slower than the time on the previous trial and will not be more than 19.20 s (95% CI [10.70, 27.71]) faster than the time on the previous trial. The bias line and the 95% LOA are presented in Figure 4, along with the data for the individual participants.

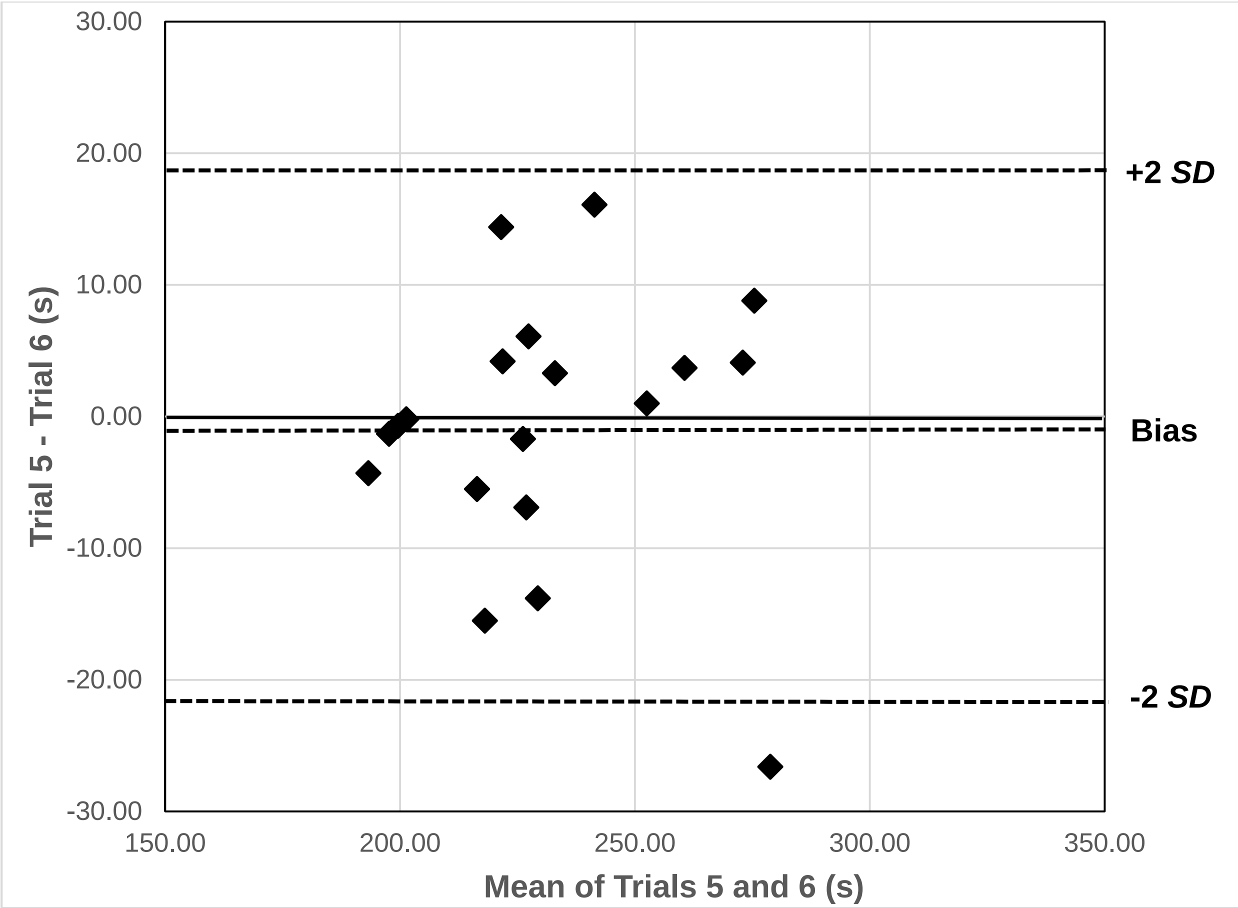


Figure 4. Bland-Altman plot for repeated measures of obstacle course completion time showing each participant's data (*N* = 19) for the difference between Trials 5 and 6 plotted against the mean of these two trials. The calculated bias line and the 95% LOA are also shown (95% LOA = Bias value ± 2 *SD*).

As indicated in Figure 4, the data point for one participant fell below -2 *SD*s (i.e., below -20.75 s). This participant had a course completion time on Trial 6 that was 27 s longer than his time on Trial 5. The data points for the remaining 18 participants were within -0.78 s ± 19.97, or -20.75 to 19.20 s. Given that the value for systematic bias, the range of values corresponding to ± 2 *SD*s, and the 95% confidence intervals for these values are narrow, the 95% LOA, like the *SEM*, indicates a high degree of absolute reliability for the course completion time measure.

Completion Times for Obstacle Course Segments

Twenty-five segments of the obstacle course were individually timed. The complete list of the segments and the corresponding segment numbers are in Appendix B. For analysis purposes, the times for some of the shorter segments were combined. Some of the timed segments comprising a single obstacle were also combined. Table 4 is a list of the single and the combined segments that were analyzed. Each row in the table represents one analysis and identifies the segment or segments included in the analysis. The last entry in the table, Interobstacle Transitions, refers to seven short-duration segments that were located between obstacles.

Table 4. *Obstacle Course Segments Analyzed*

Segment Number	Name
1	Tunnel & Hatch
3	Sprint
5-9	Stair & Ladder Climbs
11	Agility Run
13	Casualty Drag
15-16	Window #1 & Window #2
17	Bounding Rushes
18	Balance Beam
20-22	Crawls
24-25	Outer & Inner Courtyard Walls
2, 4, 10, 12, 14, 19, 23	Interobstacle Transitions

The results of the individual ANOVAs performed on the obstacle course segments to assess the effects of the Trial variable are summarized in Table 5. The segments analyzed were those listed in Table 4. Table 5 contains the values of the *F* ratios and their significance levels. The means for each trial and the findings from the Tukey HSD post hoc tests are also included in the table. The trial means for each course segment are presented graphically in Appendix G.

The ANOVA performed on the Agility Run was the only analysis that failed to yield a significant Trials main effect ($p > .05$). Analysis of the Tunnel & Hatch yielded a significant main effect of Trials, but the post hoc tests did not reveal significant differences ($p > .05$) among the trials. In all other analyses, significant differences among trials were obtained in the post hoc tests (Table 5). Reflecting the findings from the analysis of time to complete the entire obstacle course, the analyses of course segments generally revealed that times were faster for the later trials than for the initial ones. The Sprint was an exception: On that segment, the fastest time was achieved on Trial 1. Also reflecting the analysis of course completion time, the decrease in segment times was gradual, with a number of instances in which the times for the intermediate trials did not differ significantly ($p > .05$) from the times for the earlier or the later trials.

Table 5. *Trial Means (s) and Results of Analyses of Course Segments*

		Trial									
Seg. No.	Name	1	2	3	4	5	6	7	8	F^a	p^b
1	Tunnel & Hatch	25.47 A	25.79 A	24.28 A	24.45 A	24.19 A	24.29 A	23.48 A	23.52 A	2.27	< .05
3	Sprint	3.78 A	4.00 AB	3.91 AB	4.02 AB	4.07 B	4.04 B	3.96 AB	3.88 AB	2.88	< .025
5-9	Stair & Ladder Climbs	49.65 AB	50.14 A	48.59 ABC	48.32 ABCD	47.26 BCDE	46.83 CDE	45.86 DE	45.25 E	9.33	< .001
11	Agility Run	14.71 A	15.08 A	14.75 A	14.78 A	14.77 A	14.70 A	14.59 A	14.58 A	1.39	>.05
13	Casualty Drag	31.81 AB	34.51 A	30.29 AB	30.21 AB	27.54 B	28.94 AB	28.66 AB	28.02 AB	2.18	< .05
15-16	Window #1 & Window #2	11.76 A	11.07 AB	10.79 BC	10.24 BC	10.14 BC	10.62 BC	10.01 C	9.88 C	8.38	< .001
17	Bounding Rushes	29.43 A	28.63 AB	28.09 ABC	27.33 BC	27.27 BC	27.56 BC	26.67 C	26.55 C	5.02	< .001
18	Balance Beam	16.46 A	16.60 A	16.42 A	16.49 A	15.70 AB	15.41 ABC	15.04 BC	14.39 C	7.86	< .001
20-22	Crawls	26.94 A	25.89 AB	25.52 AB	25.17 AB	24.55 B	24.84 AB	24.16 B	23.93 B	3.48	< .005
24-25	Outer & Inner Walls	11.58 A	11.36 A	11.02 AB	10.51 AB	10.66 AB	10.37 AB	11.81 A	9.75 B	3.60	< .005
2, 4, 10, 12, 14, 19, 23	Interobstacle Transitions	26.78 A	26.35 AB	24.82 ABC	24.33 ABC	24.72 ABC	24.03 ABC	23.22 BC	22.71 C	3.79	< .005

Note. Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$, $N = 19$).
[a]For all F ratios, $df = 7{,}126$. [b]p value of the F ratio.

Durations of Obstacle Course Segments as Percentages of Course Completion Time

Calculations were done to obtain an overview of the contributions that course segment times made to total course completion time. For this, the proportions of total obstacle course time that the participants spent on the various segments, as listed in Table 4, were obtained and a mean was calculated over participants. The calculations were done for two trials: Trial 1, the slowest trial, and Trial 8, the fastest and last trial. The findings, which are presented in Figure 5, are the durations of the segments expressed as percentages of course completion time for each of these two trials.

As displayed in Figure 5, about 20% of the time to complete the entire obstacle course was spent on the Stair & Ladder Climbs obstacle, whereas the other elements each contributed 12% or less to the total course completion time. The distribution of time spent executing the various course segments changed only slightly between Trials 1 and 8. There were small differences on the Tunnel & Hatch, Sprint, Stair & Ladder Climbs, and Agility Run segments. In these instances, the proportion of Trial 8 time spent on a segment was slightly higher than the proportion of Trial 1 time. The greatest difference was on the Interobstacle Transitions segment. Here, the Trial 8 proportion was lower by about 2 percentage points than the Trial 1 percentage, indicating that participants spent less of their total course time on these transitions during Trial 8 than during Trial 1.

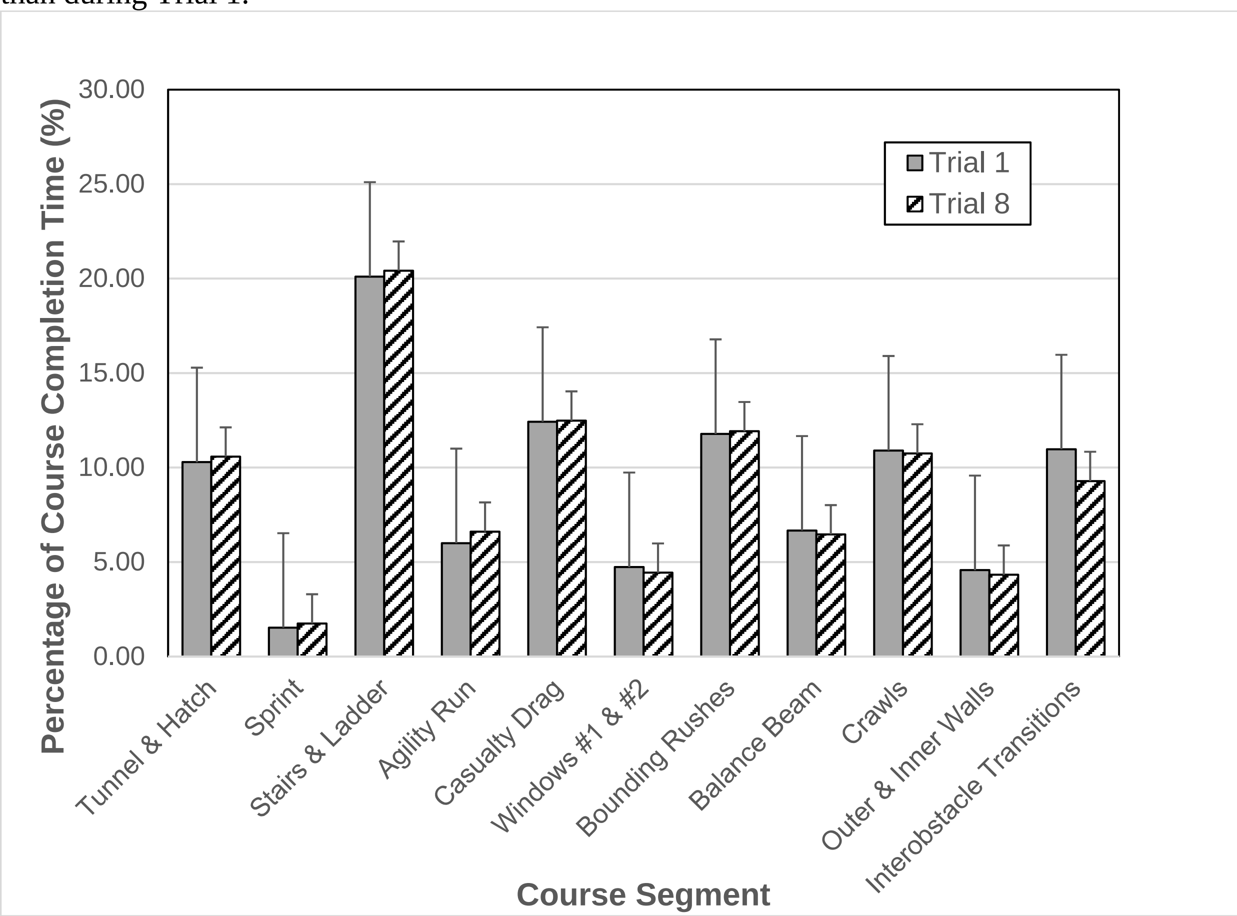


Figure 5. Mean segment durations as percentages of obstacle course completion times on Trial 1 (Mean completion time = 248.40 s) and Trial 8 (Mean completion time = 222.45 s). The error bars indicate +1*SE*.

The RPE and the Thermal Rating Scales

The *t* tests for correlated samples performed on the Borg RPE ratings yielded a significant difference between the pre-run and the post-run ratings for each trial. The values ranged from $t(18) = 10.86$, $p < .001$, for Trial 7 to $t(18) = 20.25$, $p < .001$, for Trial 1. The mean RPE ratings are presented in Figure 6. The means prior to the obstacle course run had values of approximately 6 (*No exertion at all*); the mean values after the run were in the range of 14 to 15 (*Somewhat hard* to *Hard*).

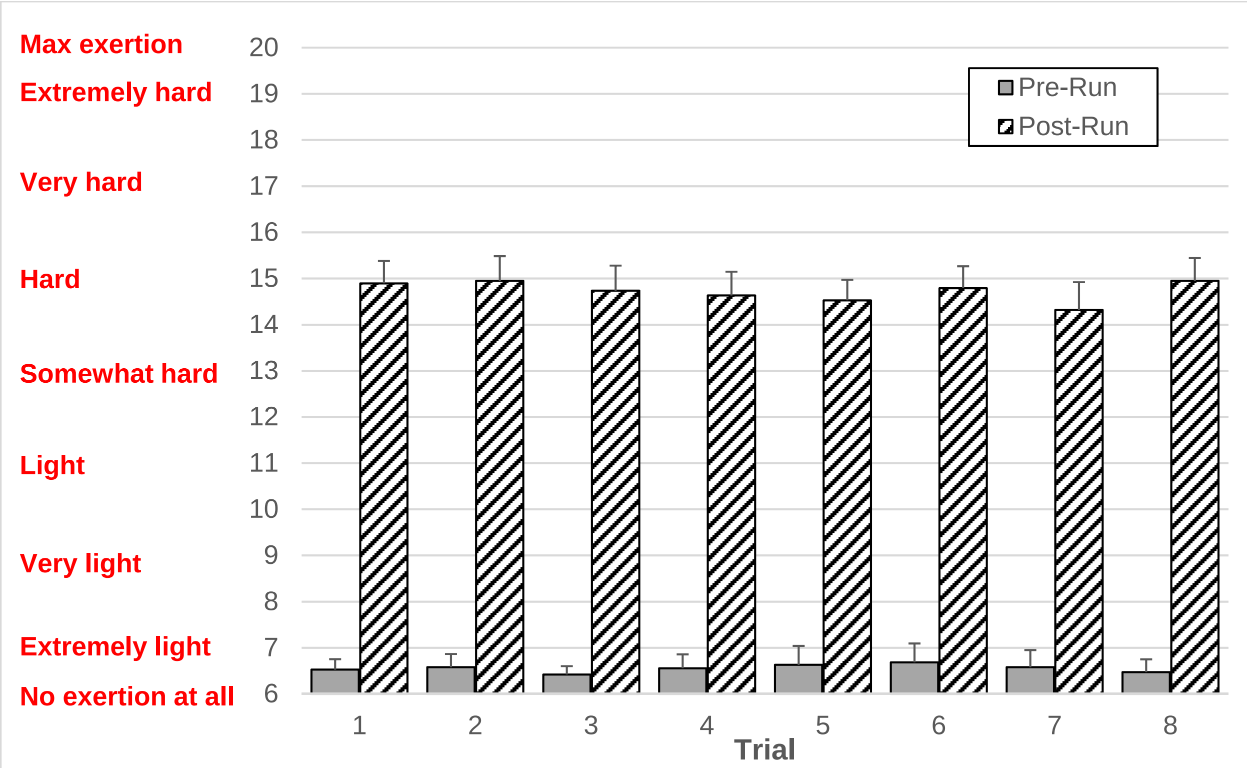


Figure 6. Mean ratings on the RPE given before and after each run on the obstacle course (*N* = 19). The error bars indicate +1*SE*.

The median ratings obtained before and immediately after each run of the obstacle course are presented in Figures 7 and 8 for the Thermal Comfort and the Thermal Sensation Scales, respectively. The median ratings on these two scales were higher after than before the course runs. The medians for the Thermal Comfort Scale increased from 1 (*Comfortable*) to 2 (*Slightly uncomfortable*). The medians on the Thermal Sensation Scale increased from 4 (*Neutral*) to 5 (*Warm*). The Wilcoxon signed ranks tests performed on the Thermal Comfort Scale data yielded significant differences between the pre- and the post-run ratings for each trial. The *z* values ranged from $z(18) = 3.11$, $p < .001$, for Trial 2 to $z(16) = 3.46$, $p < .0003$, for Trial 1. For the Thermal Sensation Scale data, the Wilcoxon tests again yielded significant differences between the pre- and the post-run ratings for each trial. The *z* values obtained from analyses of the Thermal Sensation Scale ratings ranged from $z(17) = 3.34$, $p < .0005$, for Trial 4 to $z(19) = 3.77$, $p < .00011$, for Trial 1.

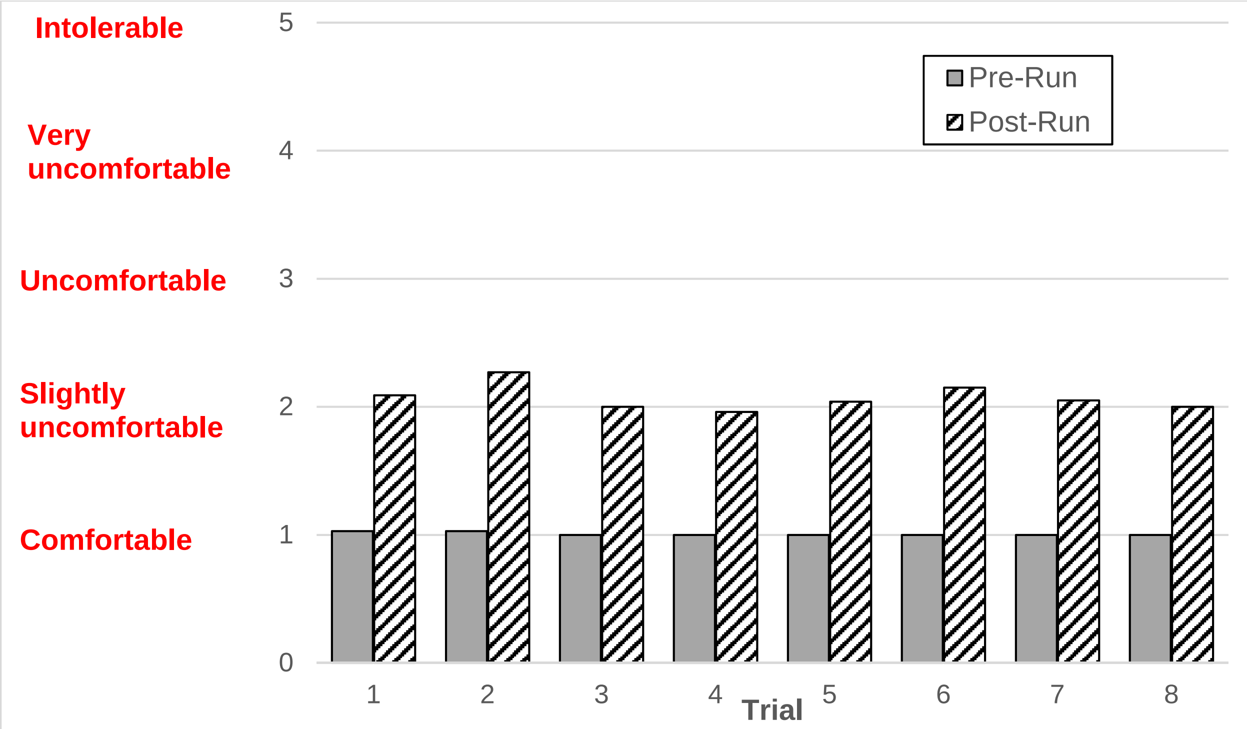


Figure 7. Median ratings on the Thermal Comfort Scale given before and after each run on the obstacle course (*N* = 19).

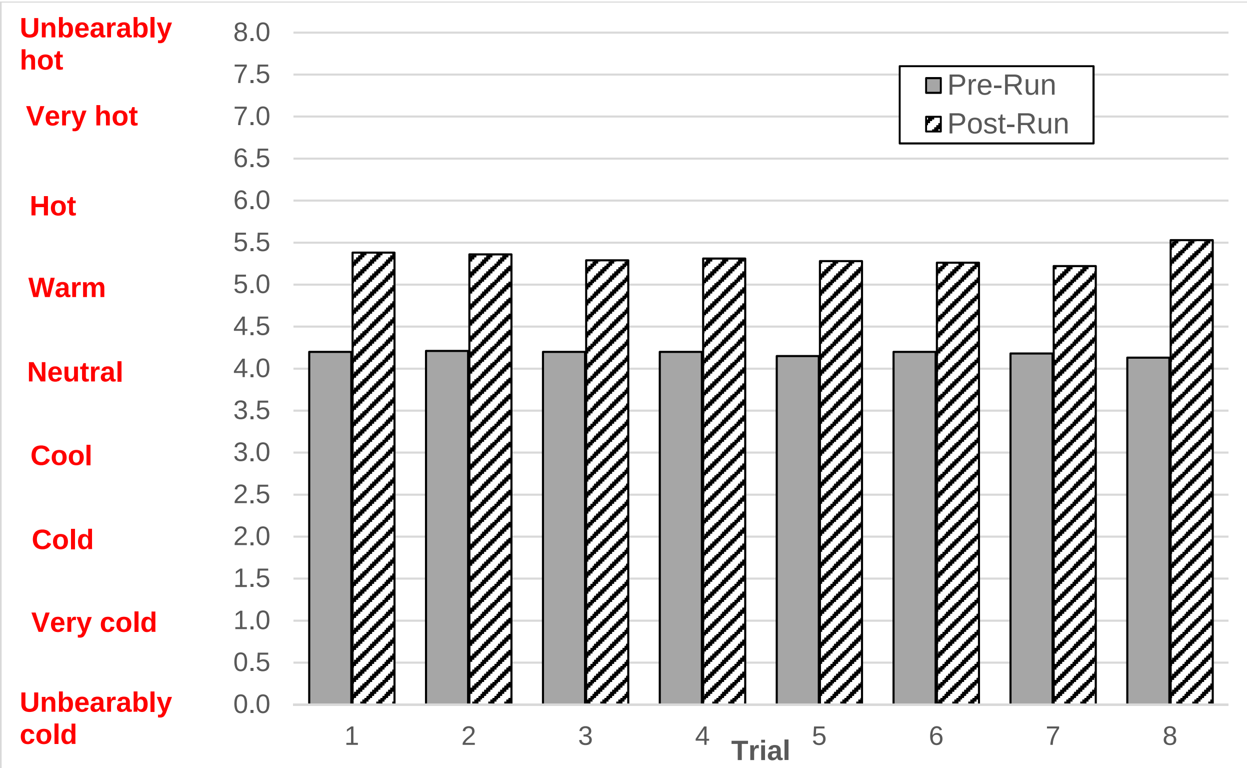


Figure 8. Median ratings on the Thermal Sensation Scale given before and after each run on the obstacle course (*N* = 19).

To investigate whether responses on the thermal rating scales were related to ambient temperatures during testing sessions (Table 3, Appendix F), median ratings were calculated separately for the morning and the afternoon participants in the October 2014 testing group and in the June 2015 group. Ratings given upon completion of an obstacle course run were used and the medians were obtained over participants' eight trials. The data are in Table 6. The differences in the medians associated with time of day of testing were small. The medians were not higher for the afternoon than for the morning sessions and thus did not reflect the higher afternoon temperatures. The differences in the medians associated with time of year of testing were also small, but the median ratings for the June testing were slightly higher than those for the October testing.

Table 6. *Median Ratings on Thermal Scales for Each Testing Group at Completion of Obstacle Course Trials During Morning and Afternoon Sessions*

	Session Time of Day	
Testing Group	Morning	Afternoon
	Thermal Comfort Scale	
October 2014[a]	1.88	1.77
June 2015[b]	2.58	2.02
	Thermal Sensation Scale	
October 2014[a]	5.44	5.23
June 2015[b]	5.73	5.47

Note. The medians were calculated over all eight trials for the morning participants in each testing group and for the afternoon participants in each testing group.
[a]For the morning participants, n = 32. For the afternoon participants, n = 24.
[b]For the morning participants, n = 48. For the afternoon participants, n = 48.

DISCUSSION

The interest of the U.S. Army and Marine Corps, as well as of military establishments of other nations, in using the LEAP as a test battery for assessing the effects on performance of clothing and equipment highlights the importance of establishing the reliability of this new test battery. As presently planned, time to complete the entire obstacle portion of the LEAP and completion times of individual course segments will factor into decisions regarding whether developmental items of clothing and equipment should be adopted for military use, replacing items currently being issued to military personnel. Decisions to introduce new clothing and protective equipment into the military supply system represent large expenditures of funds and, more importantly, have a direct impact on the combat effectiveness and survivability of service members. Before performance on the LEAP is factored into these decisions, it is imperative to establish that this testing tool has high test-retest reliability. The current study was undertaken primarily to assess the reliability of the time to complete the obstacle course portion of the LEAP. Times to complete individual segments along the obstacle course were also recorded and analyzed to examine changes with repeated trials.

Reliability of Time to Complete the Obstacle Course Portion of the LEAP

A basic element in calculating the reliability of a performance testing tool is determination of the number of iterations required before scores achieve statistically stable values. Failure to obtain consistent scores after repeated iterations indicates low test reliability. If consistent scores are achieved, but only after a large number of iterations, investigators may judge that the testing tool is impractical to use in light of study and participant scheduling constraints. In the current study, mean times to complete the first and the second runs of the obstacle course were highly similar, differing by only about 1 s. However, the faster times on the third and subsequent trials indicated that the approximately equal times on the first two trials were not evidence that stable performance had been achieved in the earliest runs of the course. Reliability calculations performed on the LEAP data revealed that time for obstacle course completion did stabilize and that statistically consistent course times were obtained after the fourth traversal of the course.

Participants in the study by Pandorf et al. (2003) executed eight trials of a six-station, indoor obstacle course. As in the current study of the LEAP, participants' schedules entailed two course traversals on each of four days. Pandorf et al. (2003) reported that stable performance was achieved after four runs on the course they used, the same number of iterations required for stable performance in the current study. Aside from the current study, the effort conducted by Pandorf et al. (2003) appears to be the only investigation of reliability of measurement of completion time on an obstacle course. It is premature on the basis of only two studies to generalize this finding on the number of runs required to obtain statistically stable completion times to other obstacle courses.

The computations done to assess reliability of the time to complete the obstacle portion of the LEAP were carried out on the two earliest trials showing statistically stable performance. There were the fifth and the sixth of the eight trials. The data for the four initial trials, which showed significant learning effects, and the last two trials, which did not differ significantly from

Trials 5 and 6, were not used in the reliability calculations. An *ICC* value of .93 was obtained, indicating that times on the obstacle course demonstrated high relative reliability. Thus, the position or ranking of a participant's completion times from trial to trial was highly similar relative to the position of the times of all other participants. A participant who had a fast completion time relative to other participants on one trial also tended to have a fast time on the other trial, and participants who were slow on one trial tended to be slow on the other.

It is interesting to note that Pandorf et al. (2003) reported a similar *ICC* value, .92, for the completion time measure on the obstacle course they used. Again, given the dearth of data on reliability of obstacle course completion time measures, it is premature to generalize the findings of the current study and the work of Pandorf et al. (2003) to conclude that all obstacle courses will reveal high relative reliability once statistically consistent performance has been achieved.

In light of the limitations associated with calculating only relative reliability that a number of authors have pointed out (Atkinson & Nevill, 1998; Bland & Altman, 1986; Bruton et al., 2000), absolute reliability was also assessed in the current study. Unlike relative reliability, which concerns the consistency of an individual's scores relative to the scores of other individuals in the study sample, absolute reliability is focused on the extent to which an individual's scores vary from trial to trial. Two methods were used to calculate absolute reliability. For one of these, the *SEM*, high absolute reliability was obtained when again analyzing the two earliest trials showing statistically stable performance: Trials 5 and 6. The value of the *SEM* was 3.13% of the grand mean of the times for all participants on these two trials. The low *SEM* value suggests that the obstacle course completion time measure can precisely differentiate between different participants.

The second method for determining absolute reliability, the 95% LOA, also yielded high reliability in analyses again performed on Trials 5 and 6. The value obtained for systematic bias was -0.78 s and the value of the random error component was 19.97 s. Therefore, it can be expected that an individual's completion time on a retrial would lie between 19.20 s and -20.75 s of the previous trial 95% of the time. Stated another way, if an individual performs two trials, the time on the second trial has 5 chances in 100 (a 5% chance) of being more than 20.75 s slower than the first trial or more than 19.20 s faster than the first trial. Like the low value obtained for the *SEM*, the narrow range of the 95% LOA suggests that the obstacle course completion time measure has high absolute reliability and can precisely differentiate between different participants.

The test for heteroscedasticity that preceded the calculation of the 95% LOA revealed that the obstacle course completion time measure did not demonstrate heteroscedasticity. Therefore, the magnitude of the between-trial variation in completion times for an individual does not depend on how fast or slow the individual was on the course.

The calculations carried out on the data from the current study indicate that the LEAP obstacle course completion time measure has high relative and absolute reliability. Therefore, investigators can use the obstacle course portion of the LEAP with some assurance that consistent completion time scores will be obtained. However, course familiarization is required before stable performance can be expected. Furthermore, the findings on the reliability of the completion time measure apply only to situations in which the obstacle course is laid out as it was for the current study, with the same stations in the same locations relative to each other. In addition, the applicability of the reliability results is contingent upon the testing procedures being the same ones used in the current study. That is, participants traversed the course individually and were instructed to expend 100% of maximal effort to traverse the entire course, from beginning to end, as quickly as possible.

The military organizations that have LEAP systems are conducting or plan to conduct tests of current and prototype military clothing and equipment using the LEAP battery. A consideration in carrying out these tests is the protocol to employ during orientation sessions. Practice on the obstacle course on one or more days preceding days of formal testing is needed if participants are to achieve stable completion times on the test trials. There is also the additional element of practice on the rifle firing, the vertical jump, and the weight transfer activities. Exposure to traversing the course in the clothing and equipment that will subsequently be worn during the formal testing is also desirable. However, study scheduling constraints, participant availability, maintenance of participants' motivation, and related considerations are likely to limit the number of pre-test practice sessions and trials that can be undertaken.

Investigators will develop protocols for orientation sessions that are compatible with their particular circumstances and the number and composition of clothing and equipment conditions to be included in a particular study. Given the results of the current study, however, it appears advisable to conduct three or four practice trials with participants wearing basic clothing on days prior to the initiation of testing. Investigators may choose to follow these initial trials with one or two practice trials in which participants are outfitted in configurations of clothing and equipment being tested in the study. Familiarity with completing the course while wearing, for example, the heaviest configuration under study and a lighter configuration is likely to aid participants in devising procedures for traversing the obstacles while encumbered.

Changes in Times to Traverse Obstacle Course Segments

In addition to obtaining times to complete traversals of the entire LEAP obstacle course, times to complete segments of the course were examined in the current study. Examination of the timing data for each individual course segment generally revealed faster times on the later trials than on the initial ones, as would be expected given the decreases in times to complete the entire obstacle course with repeated trials. However, there were exceptions. On the Agility Run, the mean times for the fastest trials were only about 3% lower than the mean time for the slowest trial and, on the Sprint, the fastest time occurred on the first trial. For the remaining segments, there were improvements in times before obstacle course completion time stabilized, but the extent of the improvements varied among obstacles.

Pandorf et al. (2003) also obtained differences among the obstacles they used in the extent to which performance improved with repeated trials. The greatest decreases in times on the course used by Pandorf et al. (2003) occurred on a low crawl activity, where a 12% decrease in time was found relative to time on the first trial. This was followed by a 9% improvement on an obstacle that entailed shimmying along a horizontal pipe.

In the current study, the greatest improvements in times before performance stabilized were associated with the Casualty Drag and the Window #1 & Window #2 obstacles. By Trial 5, mean completion times on these segments had decreased by approximately 14% relative to Trial 1 mean times. There were also substantial improvements on the Bounding Rushes, the Crawls, the Outer & Inner Courtyard Walls, and the Interobstacle Transitions. For these segments, mean times for Trial 5 were 7 to 9% faster than Trial 1 mean times. Changes in obstacle traversal times of these magnitudes highlight the need for practice runs on the LEAP course prior to undertaking formal testing. In addition, they suggest that the LEAP obstacles differ in terms of the development of strategies and skills required of the participants.

Another approach examined the proportions of total course time that participants spent on the individual segments comprising the course. Thus, the possibility could be examined that the proportions of the course completion time spent on the various segments changed as participants repeatedly ran the obstacle course, due perhaps to motor learning or strategy development (Hopkins, 2000).

Calculations were carried out on the data for Trials 1 and 8, the slowest and the fastest trials, respectively, to examine the distribution of time among course segments. The two trials yielded similar results. There was no evidence, for example, that a smaller proportion of total course time during Trial 8 was spent on complex obstacles, such as the Stair & Ladder Climbs and the Balance Beam, than was spent during Trial 1. The greatest difference between the two trials was the proportion of time used on the Interobstacle Transitions, which were seven short-duration segments located between obstacles. The Trial 8 percentage was about 2 points lower than the Trial 1 percentage. It can be posited that participants spent a smaller proportion of their total course completion times traversing the short segments because, with repeated trials, they developed the skill of approaching an upcoming obstacle in a proper position for traversing it. They may then have been able to adopt the strategy of proceeding quickly along these short segments between obstacles.

Participants' Subjective Ratings Before and After Obstacle Course Traversals

When the LEAP test battery is used for assessing effects of clothing and equipment on performance, it is likely that investigators will administer questionnaires to obtain participants' opinions of the impact of the items on successful completion of the LEAP, acceptability of the items for military use, and other matters specific to the items under test. In the current study, the questionnaires were limited to obtaining participants' responses on the RPE and their ratings of thermal status. The study provided an opportunity to obtain data on the extent to which the RPE and the thermal ratings before and immediately after a run of the LEAP obstacle course differed when participants were wearing only basic clothing and equipment.

Contrasts of participants' pre- and post-run responses indicated that, although the participants were not burdened with PPE, reported exertion was significantly higher after each run. The mean RPE rating before undertaking the obstacle course was about a 6 (*No exertion at all*), whereas the mean rating upon completion of the course was in the range of 14 to 15 (*Somewhat hard* to *Hard*). The ratings of thermal comfort and thermal sensation given before and after the obstacle course run also differed significantly. Median ratings of thermal comfort increased one scale point from a level of *Comfortable* before each run to a level of *Slightly uncomfortable* after the run. Similarly, median ratings of thermal sensation increased one scale point from *Neutral* to *Warm*.

It is likely that most future studies using the LEAP test battery will be focused on clothing and equipment assessments. Investigators will want to devise questionnaires to obtain participants' opinions of the items being tested. It would seem advisable to administer the RPE (Borg, 1970, 1982), as well. This instrument gives insight into how hard participants judge that they are working and, thus, RPE ratings would be expected to reflect the participants' sensations of the exertional burden imposed by the clothing and equipment under test. The RPE also has the benefit of being easy and quick to administer and to score. In addition, it would be informative to have RPE data available as a common metric from the different sites at which tests using the LEAP are being conducted. Whether the Thermal Comfort and the Thermal Sensation Scales should be used in future studies is likely to be a study-specific determination. It may be informative to have such rating data when certain types of clothing and equipment are being tested or somewhat extreme ambient environmental conditions are anticipated.

CONCLUSIONS AND RECOMMENDATIONS

This study has established that practice is required before consistent performance can be achieved on the obstacle course portion of the LEAP. The analysis indicates that times to complete the course stabilized after four traversals under conditions in which participants were instructed to expend 100% of maximal effort to cover the entire course as quickly as possible. This study has also established that, once stable performance is achieved, time to complete the obstacle course is a measure with high relative and absolute reliability.

Examination of times to traverse the individual segments comprising the obstacle course revealed performance improvements with repeated trials. Compared with the first trial, some obstacles were completed as much as 14% faster by the point at which times to complete the entire course had stabilized. The magnitudes of the changes that occurred in performance of individual course segments reinforce the need for subjects to practice running the course in order to achieve consistent times and to ensure that the high reliability of the obstacle course completion time measure is not compromised.

The findings from this study regarding changes in performance with repeated trials on the LEAP obstacle course and the reliability of the course completion time measure apply only to testing that replicates the methods used here, including layout of the course and instructions to participants. In conducting the study, the investigators followed the training and instructions they received from personnel of HSI, the company involved with designing the LEAP test battery. Company personnel also participated in the conduct of previous studies using the LEAP battery. Therefore, there is some assurance that this study was carried out in accordance with procedures applied in earlier studies.

A number of U.S. and foreign military organizations are already using or plan to use the LEAP for clothing and equipment assessments. The conduct of such studies by different military establishments, using the same test battery and common metrics, is a very positive feature of the LEAP and sets this test battery apart from others, which are generally unique to an organization. The opportunity to share findings through international scientific organizations is being pursued. International cooperation has the potential for accelerating identification of factors associated with military loads that negatively impact performance and means of improving performance of encumbered service members.

As with any new test, there is a need to acquire information on the LEAP test battery itself, its strengths and its weaknesses. The current study of reliability of the LEAP obstacle course completion time measure is one contribution to that effort. A principal intended use of the LEAP is as an instrument to evaluate the effects of military clothing and equipment on military service members' performance. Therefore, sensitivity of LEAP metrics to manipulation of relevant parameters such as weight, bulk, and rigidity of body-borne items must be established. Investigators at NSRDEC are engaged in a study of weight effects on LEAP performance, but additional basic studies of the LEAP as a testing instrument are needed, to include investigating the construct validity of LEAP metrics.

With experience using new tests, it is not unusual for investigators to recognize a need to revise equipment, administrative procedures, and other aspects of the test methodology. There is the potential for an organization that uses the LEAP battery to decide that a change in some aspect of the test is needed to better serve their purposes, such as modifying subjects' instructions, adding new obstacles, or reordering existent obstacles. More subtle changes are also possible, especially when investigators begin to use the LEAP for the first time. Investigators in the current study drew on the training and guidance provided by HSI in order to learn and apply the same methods used in previous studies. However, despite best efforts, there is some possibility that the methods employed in the current study were not identical to those used in earlier testing. The topic of intentionally or unintentionally changing some aspect of the LEAP is important because changing the test can have the effect of creating a new test, thereby jeopardizing the benefits that can be derived from having a number of organizations using the same test and metrics.

At present, studies are being conducted using the LEAP test battery and additional investigations are in the planning phase. Recommendations are put forward here for consideration in planning future efforts.

- This study has established that, once stable performance is achieved, the time to complete the obstacle course portion of the LEAP is a highly reliable measure. However, a number of traversals of the course at 100% of maximal effort are required to achieve stable performance. In this study, statistically consistent course times were obtained after the fourth traversal. These findings emphasize the necessity for investigators using the LEAP obstacle course to conduct a similar number of practice trials before beginning formal testing of subjects.

- A very positive aspect of the LEAP test battery is its use by a number of organizations in this country and abroad. Formal mechanisms for exchange of data on the LEAP and other obstacle courses used to assess clothing and equipment have been put in place by the North Atlantic Treaty Organization and by The Technical Cooperation Program. Efforts have been initiated under the sponsorship of these organizations to share guidance, recommendations, and lessons learned. These formal exchanges of data should continue in order to take full advantage of findings from the organizations conducting testing with the LEAP.

- To ensure that different organizations employ the same methods in conducting studies with the LEAP, and thus acquire comparable data, a highly detailed manual covering all aspects of LEAP testing, including set-up, administration, and data handling, is needed. Although HSI has prepared an extensive user manual, a finer level of detail regarding test execution should be added as organizations gain experience in administration of the LEAP. Preparation and maintenance of such a document might best be accomplished by representatives of the organizations using the LEAP.

- As experience with the LEAP increases, one or more organizations may revise the test battery to better suit their needs. A mechanism is needed for making all users aware of the revisions to avoid datasets being assumed to be comparable when they are not. Further, other organizations may find the revisions in the LEAP methods worthwhile and adopt them.

This document reports research undertaken at the U.S. Army Natick Soldier Research, Development and Engineering Center, Natick, MA, and has been assigned No. NATICK/TR- 17/002 in a series of reports approved for publication.

REFERENCES

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26*, 217-238.

Ayoub, M. M., Jiang, B. C., Smith, J. L., Selan, J. L., & McDaniel, J. W. (1987). Establishing a physical criterion for assigning personnel to U.S. Air Force jobs. *American Industrial Hygiene Association Journal, 48*, 464-470.

Billing, D., & Fordy, G. (2015). *Development of a standardized data capture and evaluation plan for the assessment of soldier agility and mobility* (Rep. TTCP Subgroup HUM Joint Panel (JP1) Human Systems Performance–Land, Collaborative Project entitled “Effects of Soldier Combat System (SCS) Burden on Soldier Performance"). Melbourne, Australia: Defence Science and Technology Organization for The Technical Cooperation Program.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet,* Feb(8), 307-310.

Borg, G. A. V. (1970). Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine, 2,* 92-98.

Borg, G. A. V. (1982). Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise, 14,* 377-381.

Borg, G. A. V., & Noble, B. J. (1974). Perceived exertion. In J. H. Wilmore (Ed.), *Exercise and Sport Sciences Reviews: Vol. 2* (pp. 131-153). New York: Academic Press.

Bossi, L., Richter, M., Tack, D., Kelly, A., Patterson, M., & LaFiandra, M. (2014). *Load Effects Assessment Program (LEAP): A systematic multinational approach to understand and address soldier physical burden.* Poster presented at the 3rd International Congress on Soldiers’ Physical Performance, Boston, MA.

Brainerd, S. T., & Bruno, R. S. (1985). *Human factors evaluation of a prototype load-carrying system* (Tech. Memo 15-85). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.

Bray-Miners, J., & Kelly, A. (2013). *CAN-LEAP summary of results — Fall 2012 experimentation series* (Contractor Report submitted by HumanSystems Inc., in partial fulfillment of PWGSC Contract No. W8486-094085/001/TOR). Toronto, CA: Defence Research and Development Canada.

Brown, J. D. (1996). *Testing in language programs.* Upper Saddle River, NJ: Prentice Hall Regents.

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy, 86*(2), 94-99.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Department of the U.S. Army. (2012). *Army physical readiness training* (Familiarization Manual 7-22). Washington, DC: Author.

Department of the U.S. Army. (2008). *Rifle marksmanship M16-/M4-series weapons* (Familiarization Manual 3-22.9). Washington, DC: Author.

Departments of the U.S. Army and U.S. Air Force. (2003). *Heat stress control and heat casualty management* (Tech. Bulletin Medical 507/Air Force Pamphlet 48-152(1)). Washington, DC: Author.

DuCharme, M. B., Jones, M. L. H., Terhaar, P., Pavlovic, N., Kelly, A., Wojtarowicz, D., & Bossi, L. M. (2014). *Intra- and inter-individual reproducibility of the CAN-LEAP obstacle course.* Poster presented at the 3rd International Congress on Soldiers' Physical Performance, Boston, MA.

Eliasziw, M., Young, S. L., Woodbury, M. G., Fryday-Field, K. (1994). Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. *Physical Therapy, 74,* 777-788.

Gordon, C. C., Blackwell, C. L., Bradtmiller, B., Parham, J. L., Barrientos, P., Paquette, S. P., . . . Kristensen, S. (2014). *2012 Anthropometric survey of U.S. Army personnel: Methods and summary statistics* (Tech. Rep. NATICK/TR-15/007). Natick, MA: U.S. Army Natick Soldier Research, Development and Engineering Center.

Goslin, B. R., & Rorke, S. C. (1986). The perception of exertion during load carriage. *Ergonomics, 29,* 677-686.

Guilford, J. P. (1956). *Fundamental statistics in psychology and education* (3rd ed.). New York: McGraw-Hill.

Hasselquist, L., Bensel, C. K., Brown, M. L., O'Donovan, M. P., Coyne, M., Gregorczyk, K. N., . . . Kirk, J. (2013). *Physiological, biomechanical, and maximal performance evaluation of medium rucksack prototypes* (Tech. Rep. NATICK/TR-13/023). Natick, MA: U.S. Army Natick Soldier Research, Development and Engineering Center.

Hasselquist, L., Bensel, C. K., Corner, B., & Gregorczyk, K. N. (2012). *An investigation of three extremity armor systems: Determination of physiological, biomechanical, and physical performance effects and quantification of body area coverage* (Tech. Rep. NATICK/TR-12/014). Natick, MA: U.S. Army Natick Soldier Research, Development and Engineering Center.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine, 30,* 1-15.

Hotzman, J., Gordon, C. C., Bradtmiller, B., Corner, B. D., Mucher, M., Kristensen, S., . . . Blackwell, C. (2011). *Measurer's handbook: US Army and Marine Corps anthropometric surveys, 2010-2011* (Tech. Rep. NATICK/TR-11/017). Natick, MA: U.S. Army Natick Soldier Research, Development and Engineering Center.

Kelly, A. (2015). *Load Effects Assessment Program — US Army (LEAP-A) operation manual.* Guelph, Ontario, Canada: HumanSystems Inc.

Kelly, A., Richter, M., Tack, D., Ueno, K., Terhaar, P., Kramkowski, E., . . . Bossi, L. (2014). *Load Effects Assessment Program (LEAP): Creation, evolution, and lessons learned.* Poster presented at the 3rd International Congress on Soldiers' Physical Performance, Boston, MA.

Noble, B. J. (1982). Clinical application of perceived exertion. *Medicine and Science in Sports and Exercise, 14,* 397-405.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Pandolf, K. B. (1978). Influence of local and central factors in dominating rated perceived exertion during physical work. *Perceptual and Motor Skills, 46,* 683-698.

Pandolf, K. B. (1982). Differential ratings of perceived exertion during physical exercise. *Medicine and Science in Sports and Exercise, 14,* 377-381.

Pandorf, C. E., Harman, E. A., Frykman, P. N., Patton, J. F., Mello, R. P., & Nindl, B. C. (2002). Correlates of load carriage and obstacle course performance among women. *Work, 18,* 179-189.

Pandorf, C. E., Nindl, B. C., Montain, S. J., Castellani, J. W., Frykman, P. N., Leone, C. D., & Harman, E. A. (2003). Reliability assessment of two militarily relevant occupational physical performance tests. *Canadian Journal of Applied Physiology, 28*(1), 27-37.

Patton, J. F., Kaszuba, J., Mello, R. P., & Reynolds, K. L. (1990). *Physiological and perceptual responses to prolonged treadmill load carriage* (Tech. Rep. T11-90). Natick, MA: U.S. Army Research Institute of Environmental Medicine.

Rayson, M., Holliman, D., & Belyavin, A. (2000). Development of physical selection procedures for the British Army. Phase 2: Relationship between physical performance tests and criterion tasks. *Ergonomics, 43,* 73-105.

Richter, M. (2014). *United States Marine Corps Load Effects Assessment Program (MC-LEAP): An emerging mobility assessment metric.* Poster presented at the 3rd International Congress on Soldiers' Physical Performance, Boston, MA.

Robertson, R. J., Caspersen, C. J., Allison, T. G., Skrinar, G. S., Abbott, R. A., & Metz, K. F. (1982). Differentiated perceptions of exertion and energy cost of young women while carrying loads. *European Journal of Applied Physiology, 49,* 69-78.

Spiering, B. A., Walker, L. A., Hendrickson, N. R., Simpson, K., Harman, E. A., Allison, S. C., & Sharp, M. A. (2012). Reliability of military-relevant tests designed to assess soldier readiness for occupational and combat-related duties. *Military Medicine*, *177,* 663-668.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Stevenson, J. M., Bryant, J. T., Andrew, G. M., Smith, J. T., French, S. L., Thomson, J. M., & Deakin, J. M. (1992). Development of physical fitness standards for Canadian Forces younger personnel. *Canadian Journal of Sport Science*, *17*(3), 214-221.

Stevenson, J. M., Reid, S. A., Bryant, J. T., Pelot, R. P., & Morin, E. L. (2001). Biomechanical assessment of the Canadian integrated load carriage system using objective assessment measures. In *Soldier mobility: Innovations in load carriage system design and evaluation* (Rep. RTO-MP-056, pp. 21:1-21:12). Neuilly-sur-Seine, France: NATO Research and Technology Organization.

Tack, D., Kelly, A., Richter, M., & Bray-Miners, J. (2012). *Preliminary results of MC-LEAP testing of U.S. Marine Corps combat load order configurations.* Quantico, VA: U.S. Marine Corps Systems Command.

Task Force Devil Combined Arms Assessment Team (2003). *The modern warrior's combat load: Dismounted operations in Afghanistan* (Report Draft 8-14-2003). Fort Leavenworth, KS: U.S. Army Center for Army Lessons Learned.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient. *Journal of Strength and Conditioning Research, 19,* 231-240.

Young, A. J., Sawka, M. N., Epstein, Y., Decristofano, B., & Pandolf, K. B. (1987). Cooling different body surfaces during upper and lower body exercise. *Journal of Applied Physiology*, *63*(3), 1218-1223.

APPENDIX A

Description of the Obstacle Course Portion of the LEAP

The 10 stations of the LEAP obstacle course are described below in the order in which they are undertaken. The descriptions are based on information presented by HSI (Kelly, 2015), the company that designed and has fabricated all the LEAP systems delivered to date.

A schematic diagram of the course indicating the layout of the stations and the distances between them, as well as the location of timers, is presented in Figure A-1. The course layout used in the current study is identical to the layout recommended by HSI. Administration of the LEAP followed guidance provided by HSI (Kelly, 2015) and by Defence Research and Development Canada-Toronto, Canadian Department of National Defence (L. Bossi, personal communication, September 19, 2014). Subjects are directed to exert 100% of maximal effort to complete the course as quickly as possible, without risking injury, and while following instructions in proper procedures to traverse the obstacles.

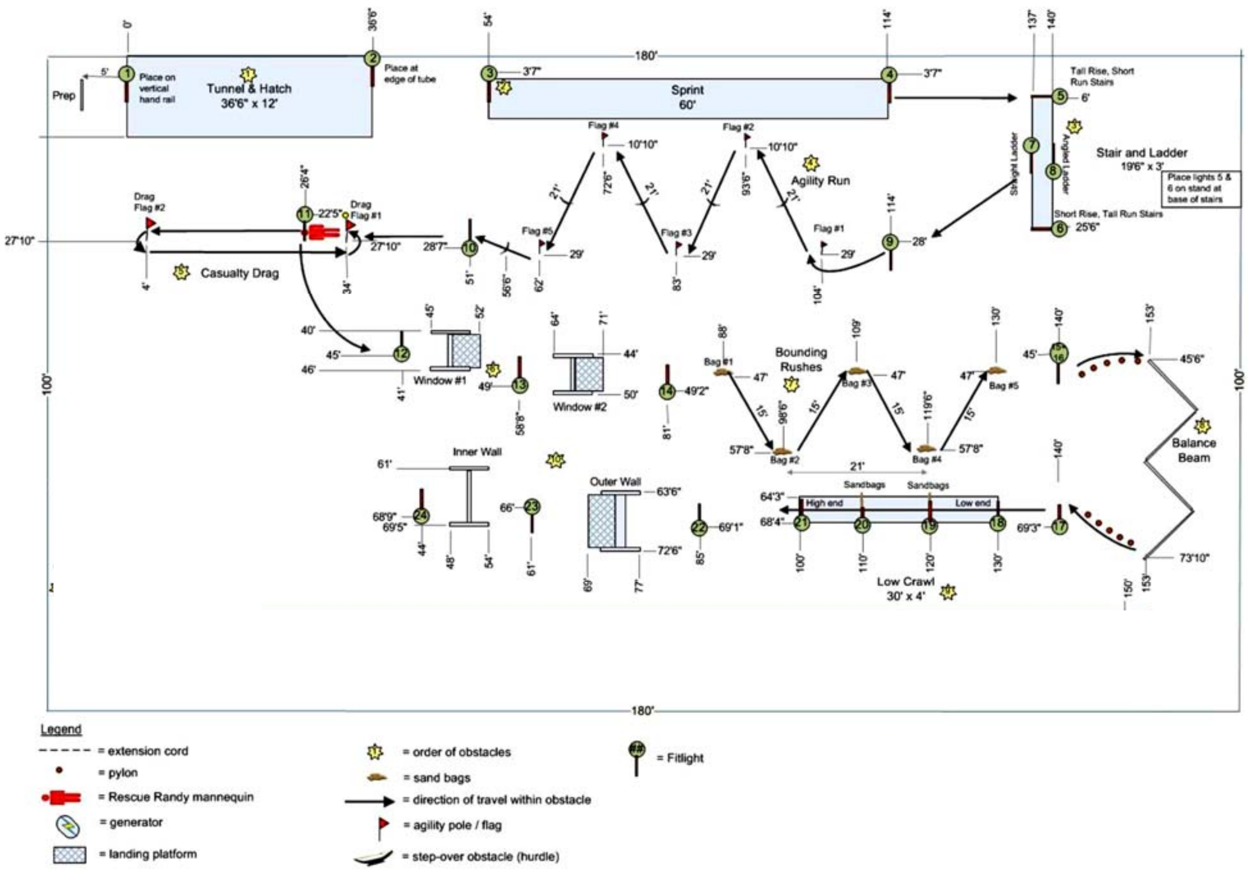
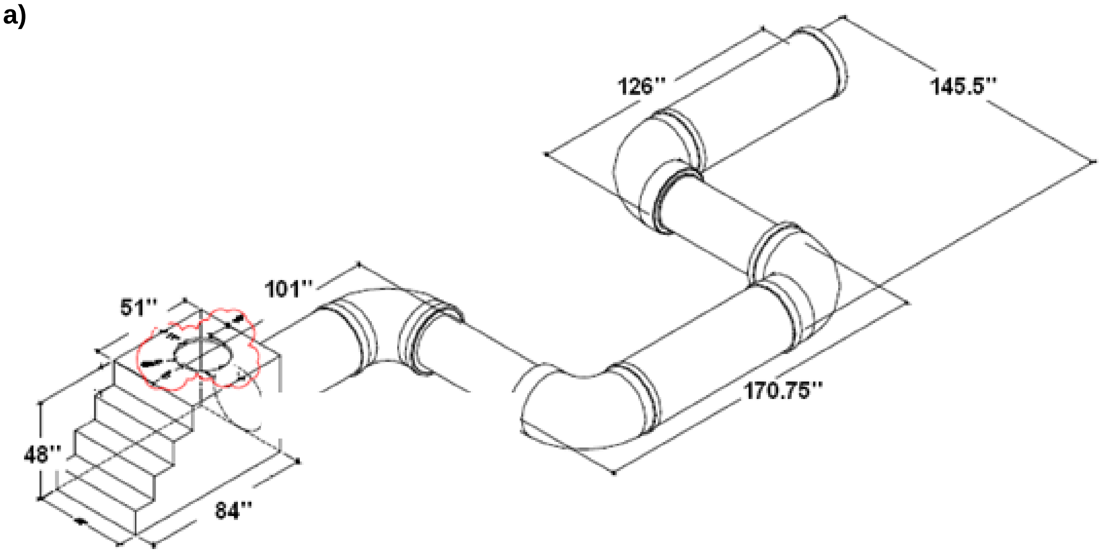


Figure A-1. Layout and dimensions (in feet) of the obstacle course portion of the LEAP.

Tunnel & Hatch

This station consists of stairs, a hatch, and a tunnel (Figure A-2a). There are four steps from the ground to the floor at the top of the stairs, where the hatch is located. The diameter of the hatch opening is 24 in. (61 cm). The tunnel is laid out in a "U" shape. The diameter of the tunnel varies between segments, with the smaller diameter segments measuring 24 in. (61 cm) and the larger diameter segments measuring 30 in. (76 cm). The subject approaches the stair portion of the obstacle (Figure A-2a) and climbs the stairs one step at a time. The subject lowers him/herself feet first into the hatch opening (Figure A-2b) and, getting into a crouched position, enters the opening of the tunnel on all fours. The subject may assume a prone position to pass through the narrower portions of the tunnel. The subject stands upon emerging at the end of the tunnel.

a)



b)

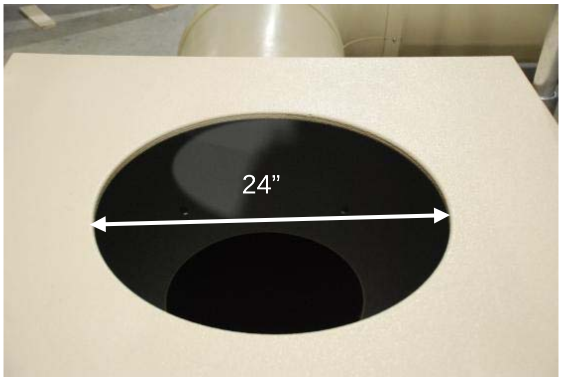


Figure A-2. Tunnel & Hatch. ***a)*** Layout and dimensions (in inches) of the Tunnel & Hatch. ***b)*** The hatch opening.

Sprint

The Sprint station is a straight run for a distance of 60 ft (18.3 m; Figure A-3).

Figure A-3. Layout of the Sprint.

Stair & Ladder Climbs

The Stair & Ladder Climbs station consists of two sets of stairs, one with a short run and high rise (i.e., steep stairs) and the other with a long run and low rise (i.e., shallow stairs). There is a platform at the top and a ladder on each side. One ladder is angled and the other is vertical (Figure A-4).

The subject progresses through this obstacle in the following order:

1) climb up the steep stairs
2) climb down the shallow stairs
3) climb up the shallow stairs
4) climb down the steep stairs
5) run to the base of the vertical ladder
6) climb up the vertical ladder
7) climb down the angled ladder
8) climb up the angled ladder
9) climb down the vertical ladder

The subject must ascend and descend the stairs one step at a time and hold on to at least one of the two railings (Figure A-4). The railings cannot be used to pull the body up the steps. When climbing the ladders, the subject may skip the rung closest to the ground, but must step on the second rung up. The subject must continue to step on each rung until reaching the last two rungs at the top of the ladder (Figure A-4). These two rungs can be skipped. Upon descending the ladders, the two top rungs can be skipped. The subject must step on all other rungs and must place both feet on the ground upon reaching the bottom of each ladder.

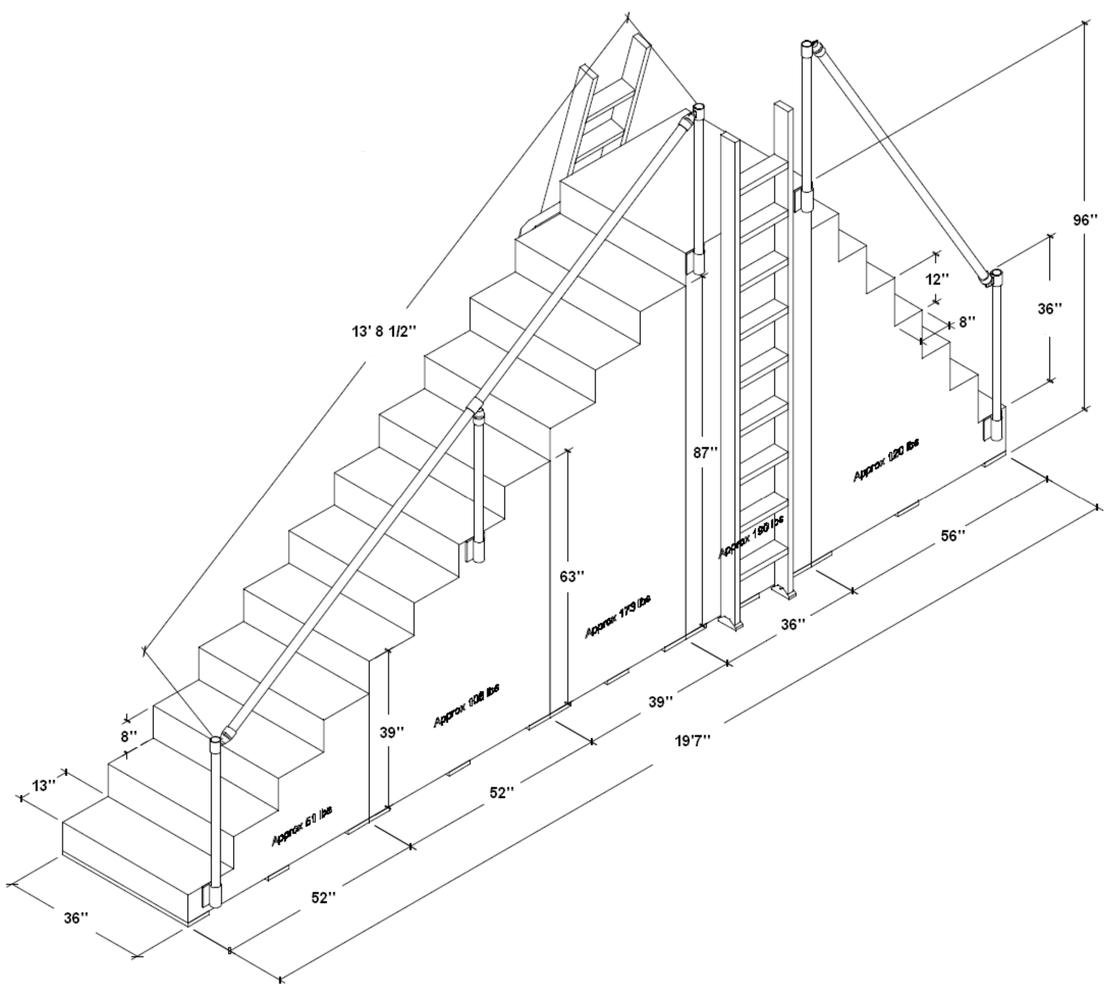


Figure A-4. Layout and dimensions (in inches) of the Stair & Ladder Climbs.

Agility Run

The layout and dimensions of the Agility Run station are shown in Figure A-5. The run consists of a series of five spring-loaded poles with flags inserted into the ground in a zigzag formation. There is a distance of 21 ft (6.4 m) between each pole. Two identical low hurdles, 12 in. (30.5 cm) high, are placed halfway between each set of poles, along a straight leg. The subject must jump over the hurdles. A sample of a low hurdle is shown in Figure A-6.

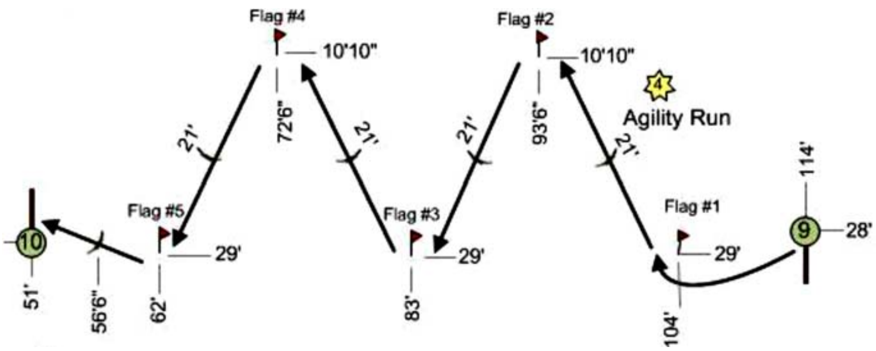


Figure A-5. Layout and dimensions of the Agility Run (in feet), showing the low hurdles placed midway along each straight leg.

Figure A-6. Sample of a 12-in. (30.5-cm) high hurdle.

To execute the Agility Run, the subject runs toward the first pole, makes a tight cut around the outside of the pole, and heads in the opposite direction toward the second pole, jumping over the hurdles placed midway along the straight leg. This continues for the sets of five poles and five hurdles that are placed in the formation shown in Figure A-5.

Casualty Drag

The subject drags a "Rescue Randy" manikin from which the lower legs have been removed along the ground following a prescribed route (Figures A-7a and A-7b). The manikin is outfitted with a ballistic protective vest. An Improved Outer Tactical Vest (IOTV), without ballistic plates, was used in the present testing. The manikin is initially positioned on the ground as shown in Figure A-7b. The subject grasps the casualty extraction strap or the shoulders of the vest in order to drag the manikin (Figure A-7a). The combined weight of the manikin and the IOTV was 180 lb (81.6 kg) in the current study.

To begin this station, the subject runs 7.8 ft (2.4 m) along a straight path to the location of the manikin and lifts the manikin by the extraction strap or the vest shoulders (Figure A-7a). The subject drags the manikin along a straight path for a distance of 30 ft (9.1 m), passes around the outside of a pole, and heads back 30 ft (9.1 m), passing around the outside of a second pole to return the manikin to its original position (Figure A-7b). The obstacle is completed when the manikin is placed back in its original position. The subject then runs off to the next station.

a)

b)

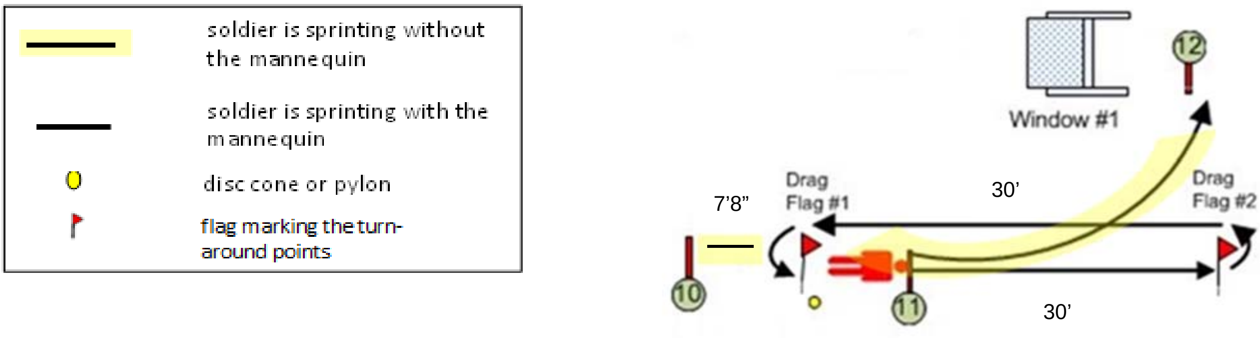


Figure A-7. Casualty Drag. ***a)*** Subject dragging the manikin by the casualty extraction strap on the vest. ***b)*** Layout and dimensions (in feet) of the Casualty Drag station.

Window #1 & Window #2

The two window structures at the Window #1 & Window #2 station are 12 ft (3.6 m) apart (Figure A-8a). Both structures measure 5 ft (1.5 m) wide x 10 ft (3 m) high x 8 in. (20.3 cm) deep. The window opening cutout in each structure is 36 in. x 36 in. (91 cm x 91 cm) and both window structures have a landing platform on the exit side (Figure A-8a). Each landing platform is 4 ft (1.2 m) long and 5 ft (1.5 m) wide and is 1 ft (0.3 m) above ground level. The bottom ledge of the opening of the first window is 5 ft (1.5 m) above the ground and the bottom ledge of the opening of the second window is 4 ft (1.2 m) off the ground. The first window has three toe holds on the approach side (Figure A-8b) and second window has a smooth surface, without toe holds (Figure A-8c).

Figure A-8. Window #1 & Window #2. ***a)*** Layouts of Window #1 & Window #2. Window #1 is in the foreground. ***b)*** Surface and dimensions (in feet) of Window #1. ***c)*** Surface and dimensions (in feet) of Window #2.

To complete this station, the subject must pass through the opening in the first window, land on the platform, and proceed to the second window, again passing through the window opening and landing on the platform (Figure A-8a). Subjects are free to choose whether or not to use the toe holds on the first window (Figure A-8b). They must land on their feet on the landing platforms. Thus, subjects are instructed not to dive or roll through the windows. Subjects are also instructed not to reach beyond the window openings in order to hold onto the edges of the 5-ft (1.5-m) wide window structures.

Bounding Rushes

The route for the Bounding Rushes station is laid out in a zigzag fashion (Figure A-9). Two sandbags are placed at each of five locations in a staggered pattern along the route. The two sandbags, which are side by side, represent a firing position. The subject runs to the first pair of sandbags, drops to a prone firing position, and takes a sight picture. The subject then rises, runs along a straight leg to the next pair of sandbags, and drops to a prone firing position. The subject repeats the actions of running and dropping to a prone firing position. After taking a sight picture at the fifth pair of sandbags, the subject rises to a standing position and runs 10 ft (3.0 m) to complete the station.

Typically, subjects would be carrying a weapon as they traversed the obstacle course, but a weapon was not carried in this study. Therefore, subjects were instructed to raise their arms to imitate holding a weapon and getting a sight picture while completing this station.

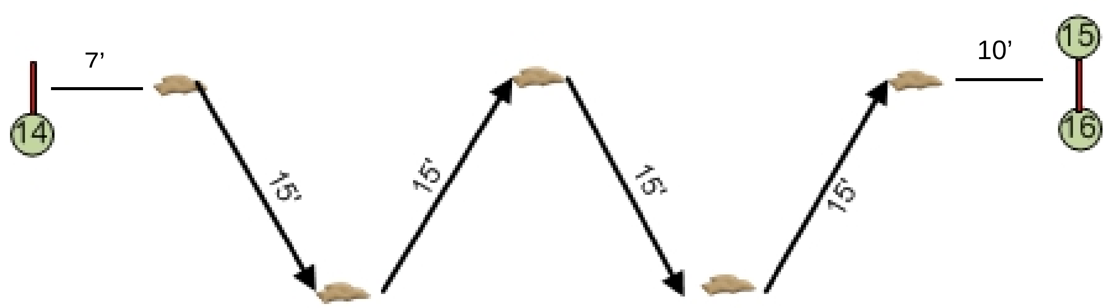


Figure A-9. Layout and dimensions (in feet) of the Bounding Rushes station showing the pairs of sandbags placed along the route. Each pair of sandbags represents a firing position.

Balance Beam

The Balance Beam station consists of a series of four sloped metal plank segments connected together at right angles (Figure A-10). Each of the segments is 10 ft (3 m) long and 8 in. (20.3 cm) wide. The first segment starts at a height of 6 in. (15.2 cm) off the ground and slopes upward at approximately 15°, reaching a maximum height of approximately 31 in.

(78.7 cm) from ground level. The second segment slopes downwards, reaching a height of 6 in. (15.2 cm) from the ground at the end. The third segment slopes upwards and the fourth downwards to the same specifications as the first two planks. In addition to being placed at a slope along their length, each plank segment is angled at approximately -8° to the horizontal. Four box-shaped obstacles are located on top of the planks, one on each segment (Figure A-10). The box-shaped obstacles measure 8 in. x 8 in. x 8 in. (20.3 cm x 20.3 cm x 20.3 cm). They are permanently affixed to the plank segments at locations of 41 in. (104.1 cm), 40 in. (101.6 cm), 28 in. (71.1 cm), and 12 in. (30.5 cm) in from the edge of the first, second, third, and fourth segments, respectively.

Figure A-10. The Balance Beam station showing the angled planks and the box-shaped obstacle located on each plank segment.

To complete the Balance Beam station, the subject steps up onto the end of the first plank segment. Jumping up onto the beam from the side is not permitted. The subject walks the entire length of each plank segment. Taking a long step onto the next plank, before walking the length of a plank segment, is not permitted. The subject must step over each box-shaped obstacle (Figure A-10); stepping on top of the box obstacle or swinging either foot around the side of the box obstacle is not permitted. Should subjects lose balance and step off before reaching the end of the fourth and last beam, they climb back on to the beam at the point at which they stepped off.

Crawls

The Crawls station is comprised of seven pairs of poles set up to form a straight path 30 ft (9.1 m) long and 4 ft (1.2 m) wide (Figure A-11a). The poles support a length of nylon fabric. Each pair of poles is connected by an arch to keep the poles from collapsing toward the center. The first five sets of poles are 20 in. (50.8 cm) high and are 5 ft (1.5 m) apart, making up a 20-ft (6.1-m) segment. Rows of sandbags are located at the 10-ft (3-m) mark and the 20-ft (6.1-m) mark. A pair of transition poles is also located at the 20-ft (6.1-m) mark (Figure A-11b). Here, the height of the poles is changed from 20 in. (50.8 cm) to 26 in. (66 cm). The remaining two pairs of poles are placed 5 ft (1.5 m) apart and form the final 10 ft (3 m) of the Crawls station.

a)

b)

Figure A-11. Crawls. ***a)*** Layout of the Crawls station showing the location of the sandbags and the height change at the transition point. ***b)*** Close-up of the height transition poles.

At the Crawls station, the subject executes a low crawl, a back crawl, and a high crawl. The subject assumes a prone position and crawls on the belly, with a cheek to the ground (i.e., low crawl), for 10 ft (3 m; to the first pair of sandbags). The subject crawls over the sandbags, rolls onto the back, and executes a back crawl for the next 10 ft (3 m; to the second pair of sandbags). The subject crawls over these sandbags and performs a high crawl, on hands and knees, for the remaining 10 ft (3 m) of the Crawls station (Figure A-11a). The subject then returns to a standing position and proceeds to the next station.

Outer & Inner Courtyard Walls

At the Outer & Inner Courtyard Walls station, the walls are 14 ft (4.3 m) apart (Figure A-12a). The outer wall structure measures 8 ft (2.4 m) wide x 6 ft (1.8 m) high x 18 in. (45.7 cm) deep. The outer wall contains nine toe holds (five protruding and five recessed) on the approach side and has a landing platform on the exit side (Figure A-12b). The inner wall structure is 8 ft (2.4 m) wide x 4 ft (1.2 m) high x 6 in. (15.2 cm) deep. The surface of this wall is smooth, without toe holds (Figure A-12c).

a)

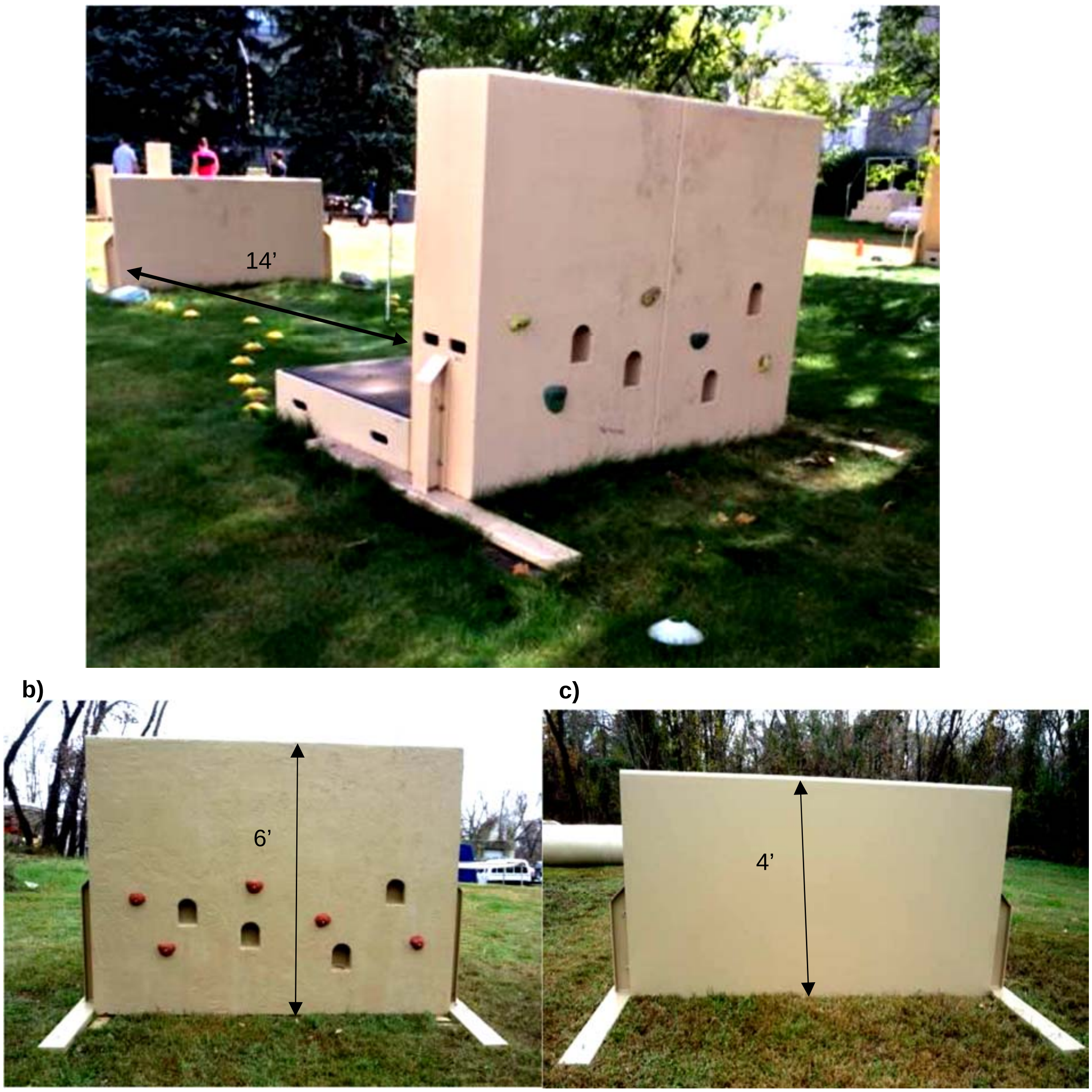


Figure A-12. Outer & Inner Courtyard Walls. ***a)*** Layout of outer and inner walls. The outer wall is in the foreground. ***b)*** Surface and dimensions (in feet) of the outer wall, showing toe holds. ***c)*** Surface and dimensions (in feet) of the inner wall, showing smooth surface without toe holds.

The subject begins this station by climbing over the outer wall. The subject then travels to the inner wall and climbs over it. Subjects are free to choose whether or not to use the toe holds on the outer wall (Figure A-12b). They must land on their feet on the landing platforms or landing areas. Thus, subjects are instructed not to dive or roll over the walls and to keep a low body profile as they pass over the tops of the walls. Subjects are also instructed not to reach around the 8-ft (2.4-m) wide wall structures to hold onto the edges.

Completion of the Outer & Inner Courtyard Walls station completes the timed portion of the LEAP.

APPENDIX B

Timed Segments of the LEAP Obstacle Course

Table B-1. *List of Segments of the Obstacle Course for Which Completion Times Were Obtained*

Number	Segment
1.	Tunnel & Hatch
2.	Tunnel & Hatch to Sprint Transition
3.	Sprint
4.	Sprint to Stair & Ladder Climbs Transition
5.	Steep stairs up, shallow stairs down
6.	Shallow stairs up, steep stairs down
7.	Stairs to ladder Transition
8.	Vertical ladder up, angled ladder down
9.	Angled ladder up, vertical ladder down
10.	Stair & Ladder Climbs to Agility Run Transition
11.	Agility Run
12.	Agility Run to Casualty Drag Transition
13.	Casualty Drag
14.	Casualty Drag to Window #1 & Window #2 Transition
15.	Window #1
16.	Window #2
17.	Bounding Rushes
18.	Balance Beam
19.	Balance Beam to Crawls Transition
20.	Low crawl
21.	Back crawl
22.	High crawl
23.	Crawls to Outer & Inner Courtyard Walls Transition
24.	Outer wall
25.	Inner wall

APPENDIX C

Sample of the Borg Rating of Perceived Exertion Scale

(Reprint of original)

Volunteer Number:______ Date:_________ Test Condition:____________

Borg Scale

RPE	**Exertion**
6	No exertion at all
7	Extremely light
8	
9	Very light
10	
11	Light
12	
13	Somewhat hard
14	
15	Hard (heavy)
16	
17	Very hard
18	
19	Extremely hard
20	Maximal exertion

Instructions for Borg Rating of Perceived Exertion (RPE) Scale

While doing physical activity, we want you to rate your perception of exertion. This feeling should reflect how heavy and strenuous the exercise feels to you, combining all sensations and feelings of physical stress, effort, and fatigue. Do not concern yourself with any one factor such as leg pain or shortness of breath, but try to focus on your total feeling of exertion.

Look at the rating scale below while you are engaging in an activity; it ranges from 6 to 20, where 6 means "no exertion at all" and 20 means "maximal exertion." Choose the number from below that best describes your level of exertion. This will give you a good idea of the intensity level of your activity, and you can use this information to speed up or slow down your movements to reach your desired range.

Try to appraise your feeling of exertion as honestly as possible, without thinking about what the actual physical load is. Your own feeling of effort and exertion is important, not how it compares to other people's. Look at the scales and the expressions and then give a number.

9 corresponds to "very light" exercise. For a healthy person, it is like walking slowly at his or her own pace for some minutes

13 on the scale is "somewhat hard" exercise, but it still feels OK to continue.

17 "very hard" is very strenuous. A healthy person can still go on, but he or she really has to push him- or herself. It feels very heavy, and the person is very tired.

19 on the scale is an extremely strenuous exercise level. For most people this is the most strenuous exercise they have ever experienced.

Borg RPE scale

APPENDIX D

Samples of the Thermal Comfort and the Thermal Sensation Scales

THERMAL COMFORT				
Comfortable / OK	**Slightly Uncomfortable**	**Uncomfortable**	**Very Uncomfortable**	**Intolerable**
1	**2**	**3**	**4**	**5**

THERMAL SENSATION		
	0.0	Unbearably Cold
	0.5	
	1.0	Very Cold
	1.5	
	2.0	Cold
	2.5	
	3.0	Cool
	3.5	
	4.0	Neutral
	4.5	
	5.0	Warm
	5.5	
	6.0	Hot
	6.5	
	7.0	Very Hot
	7.5	
	8.0	Unbearably Hot

APPENDIX E

Warm-up Exercises Performed Prior to LEAP Obstacle Course Runs

Participants in this study performed warm-up exercises prior to each run of the obstacle course. The exercises carried out comprise a preparation drill prescribed as part of the Army's physical readiness training program (Department of the U.S. Army, 2012). The 10 exercises, which the participants were experienced in performing as part of their daily Army physical training, are described here. Before the first of the two obstacle course runs at a session, study participants performed 10 repetitions of each of the 10 exercises. Before beginning the second run of the course, participants performed three to five repetitions of each of the exercises.

The information that follows is reprinted from: Chapter 6, *Army physical readiness training* (Familiarization Manual 7-22; 2012). Washington, DC: Department of the U.S. Army.

EXERCISE 1: BEND AND REACH

Purpose: This exercise develops the ability to squat and reach through the legs. It also serves to prepare the spine and extremities for more vigorous movements by moving the hips and spine through full flexion (Figure 6-56).
Starting Position: Straddle stance with arms overhead.
Cadence: SLOW
Count:
1. Squat with the heels flat as the spine rounds forward to allow the straight arms to reach as far as possible between the legs.
2. Return to the starting position.
3. Repeat count 1.
4. Return to the starting position.

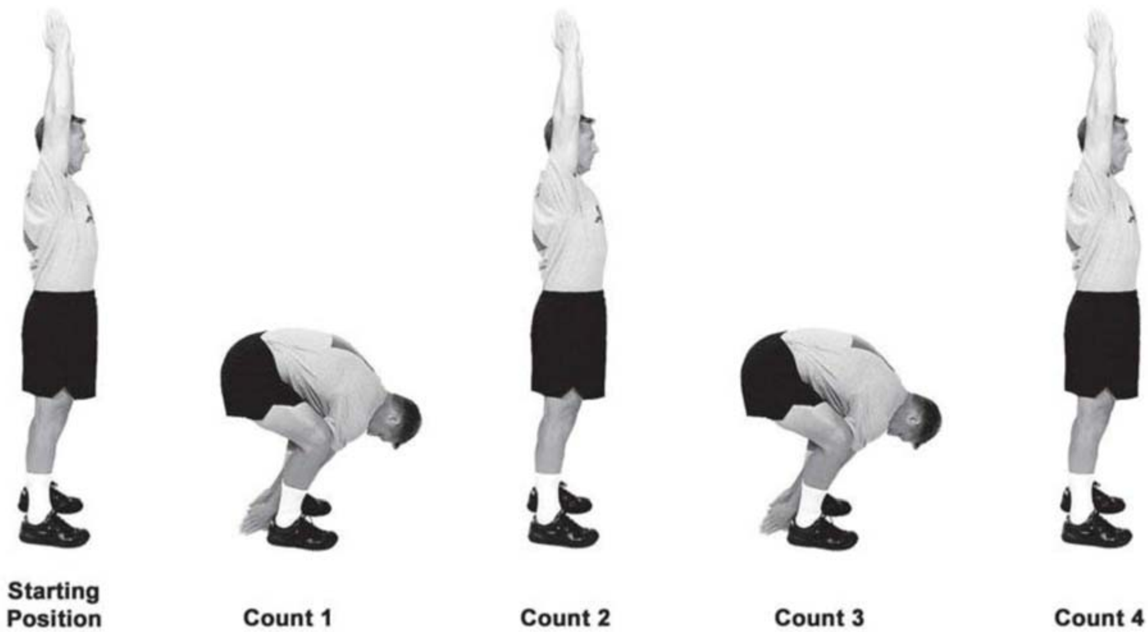


Figure 6-56. Bend and reach

Check Points:
- From the starting position, ensure that Soldiers have their hips set, their abdominals tight, and their arms fully extended overhead.
- The neck flexes to allow the gaze to the rear; this brings the head in line with the bend of the trunk.
- The heels and feet remain flat on the ground.
- On counts 2 and 4, they do not go past the starting position.

Precautions: This exercise is always performed at a slow cadence. To protect the back, move into the count 1 position in a slow, controlled manner. Do not bounce into or out of this position, as this may place an excessive load on the back.

EXERCISE 2: REAR LUNGE

Purpose: This exercise promotes balance, opens up the hip and trunk on the side of the lunge, and develops leg strength (Figure 6-58).
Starting Position: Straddle stance with hands on hips.
Cadence: SLOW
Count:
1. Take an exaggerated step backward with the left leg, touching down with the ball of the foot.
2. Return to the starting position.
3. Repeat count 1 with the right leg.
4. Return to the starting position.

Starting Position | Count 1 | Count 2 | Count 3 | Count 4

Figure 6-58. Rear lunge

Check Points:
- Maintain straightness of the back by keeping the abdominal muscles tight throughout the motion.
- After the foot touches down, allow the body to continue to lower. This promotes flexibility of the hip and trunk.
- On counts 1 and 3, step straight to the rear, keeping the feet directed forward. When viewed from the front, the feet maintain their distance apart both at the starting position and at the end of counts 1 and 3.
- Keep the rear leg as straight as possible but not locked.
- Ensure the heel of the rear foot does not touch the ground.

Precautions: This exercise is always performed at a slow cadence. On counts 1 and 3, move into position in a slow, controlled manner. If the cadence is too fast, it will be difficult to go through a full range of motion.

EXERCISE 3: HIGH JUMPER

Purpose: This exercise reinforces correct jumping and landing, stimulates balance and coordination, and develops explosive strength (Figure 6-60).
Starting Position: Forward leaning stance.
Cadence: MODERATE
Count:
1. Swing arms forward and jump a few inches.
2. Swing arms backward and jump a few inches.
3. Swing arms forward vigorously overhead while jumping forcefully.
4. Repeat count 2. On the last repetition, return to the starting position.

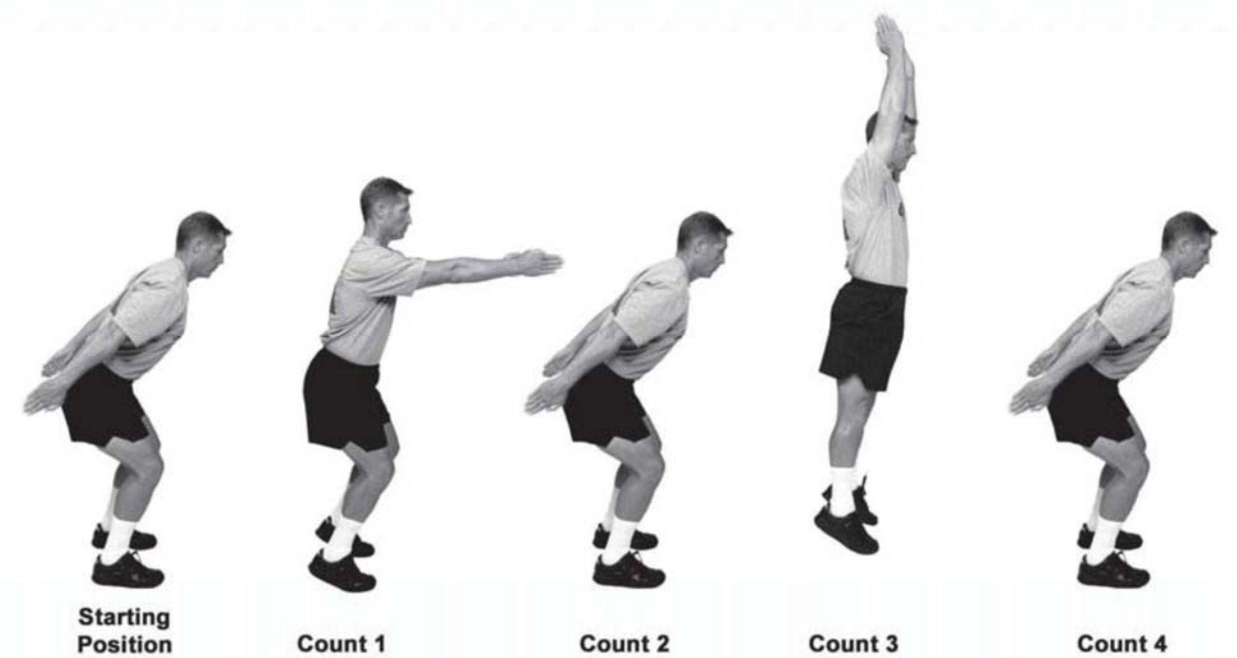


Figure 6-60. High jumper

Check Points:
- At the starting position, the shoulders, the knees, and the balls of the feet should form a straight vertical line.
- On count 1, the arms are parallel to the ground.
- On count 3, the arms should be extended fully overhead. The trunk and legs should also be aligned.
- On each count the Soldier is jumping. On counts 1, 2, and 4 the jumps are 4-6 inches off the ground.
- On count 3, the Soldier jumps higher (6-10 inches) while maintaining the posture pictured in Figure 6-60.
- On each landing, the feet should be directed forward and maintained at shoulder distance apart. The landing should be "soft" and proceed from the balls of the feet to the heels. The vertical line from the shoulders through the knees to the balls of the feet should be demonstrated on each landing.

Precautions: N/A

EXERCISE 4: ROWER

Purpose: This exercise improves the ability to move in and out of the supine position to a seated posture. It coordinates the action of the trunk and extremities while challenging the abdominal muscles (Figure 6-62).

Starting Position: Supine position, arms overhead and feet together, and pointing upward. The chin is tucked and the head is 1-2 inches above the ground. Arms are shoulder-width, palms facing inward, with fingers and thumbs extended and joined.

Cadence: SLOW

Count:

1. Sit up while swinging arms forward and bending at the hip and knees. At the end of the motion, the arms will be parallel to ground, palms facing inward.
2. Return to the starting position.
3. Repeat count 1.
4. Return to the starting position.

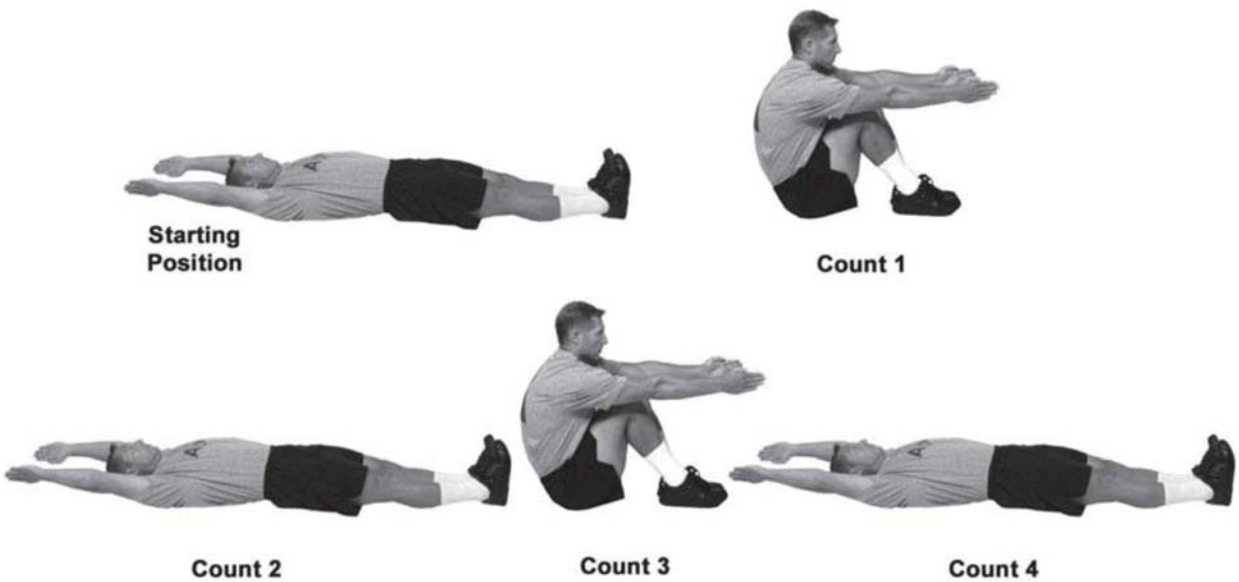


Figure 6-62. Rower

Check Points:

- At the starting position, the low back must not be arched excessively off the ground. To prevent this, tighten the abdominal muscles to tilt the pelvis and low back toward the ground.
- At the end of counts 1 and 3, the feet are flat and pulled near the buttocks. The legs stay together throughout the exercise and the arms are parallel to the ground.

Precautions: This exercise is always performed at a slow cadence. Do not arch the back to assume counts 1 and 3.

EXERCISE 5: SQUAT BENDER

Purpose: This exercise develops strength, endurance, and mobility of the lower back and lower extremities (Figure 6-65).
Starting Position: Straddle stance with hands on hips.
Cadence: SLOW
Count:
1. Squat while leaning slightly forward at the waist with the head up and extend the arms to the front, with arms parallel to the ground and palms facing inward.
2. Return to the starting position.
3. Bend forward and reach toward the ground with both arms extended and palms inward.
4. Return to the starting position.

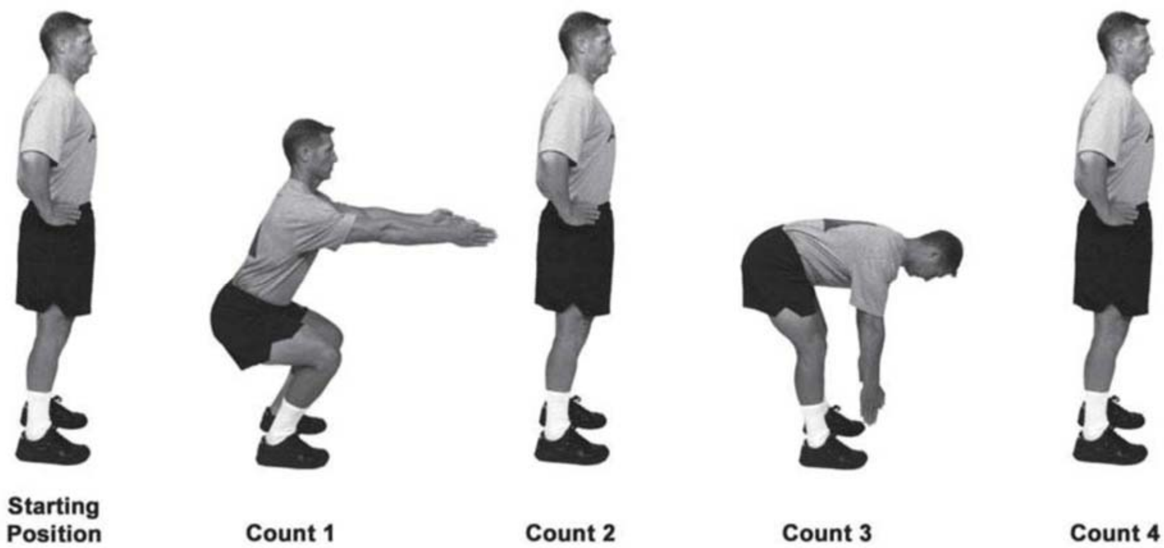


Figure 6-65. Squat bender

Check Points:
- At the end of count 1, the shoulders, knees and the balls of the feet should be aligned. The heels remain on the ground and the back is straight.
- On count 3, bend forward, keeping the head aligned with the spine and the knees slightly bent. Attempt to keep the back flat and parallel to the ground.

Precautions: This exercise is always performed at a slow cadence. Allowing the knees to go beyond the toes on count 1 increases stress to the knees.

EXERCISE 6: WINDMILL

Purpose: This exercise develops the ability to safely bend and rotate the trunk. It conditions the muscles of the trunk, legs, and shoulders (Figure 6-67).
Starting Position: Straddle stance with arms sideward, palms facing down, fingers and thumbs extended and joined.
Cadence: SLOW
Count:
1. Bend the hips and knees while rotating to the left. Reach down and touch the outside of the left foot with the right hand and look toward the rear. The left arm is pulled rearward to maintain a straight line with the right arm.
2. Return to the starting position.
3. Repeat count 1 to the right.
4. Return to the starting position.

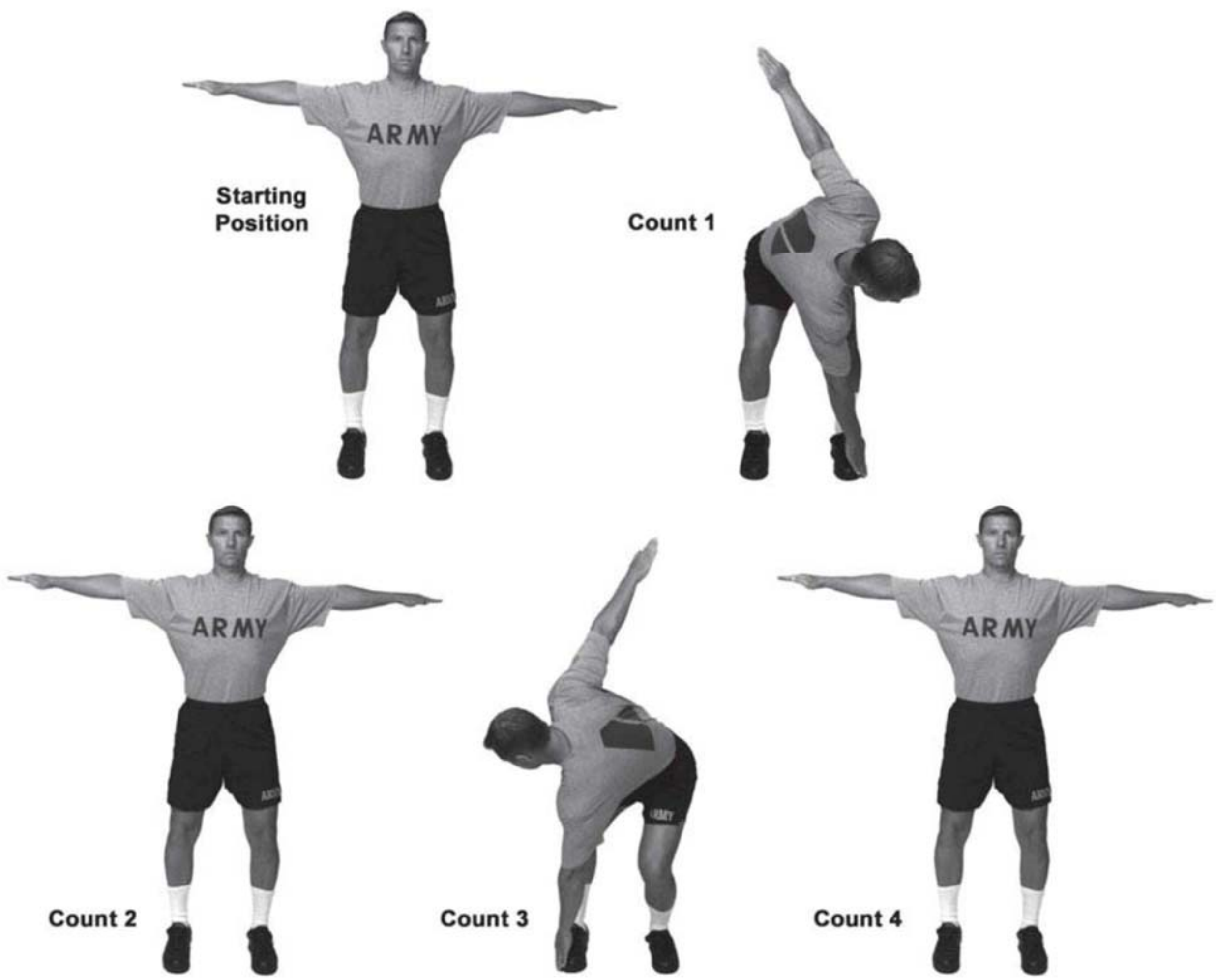


Figure 6-67. Windmill

Check Points:

- From the starting position, feet are straight ahead, arms are parallel to the ground, hips set, and abdominals are tight.

- On counts 1 and 3, ensure that the knees bend during the rotation. Head and eyes are directed to the rear on counts 1 and 3.

Precautions: This exercise is always performed at a slow cadence.

FORWARD LUNGE

Purpose: This exercise promotes balance and develops leg strength (Figure 6-71).
Starting Position: Straddle stance with hands on hips.
Cadence: SLOW
Count:
1. Take a step forward with the left leg (the left heel should be 3-6 inches forward of the right foot). Lunge forward, lowering the body and allow the left knee to bend until the thigh is parallel to the ground. Lean slightly forward, keeping the back straight.
2. Return to the starting position.
3. Repeat count 1 with the right leg.
4. Return to the starting position.

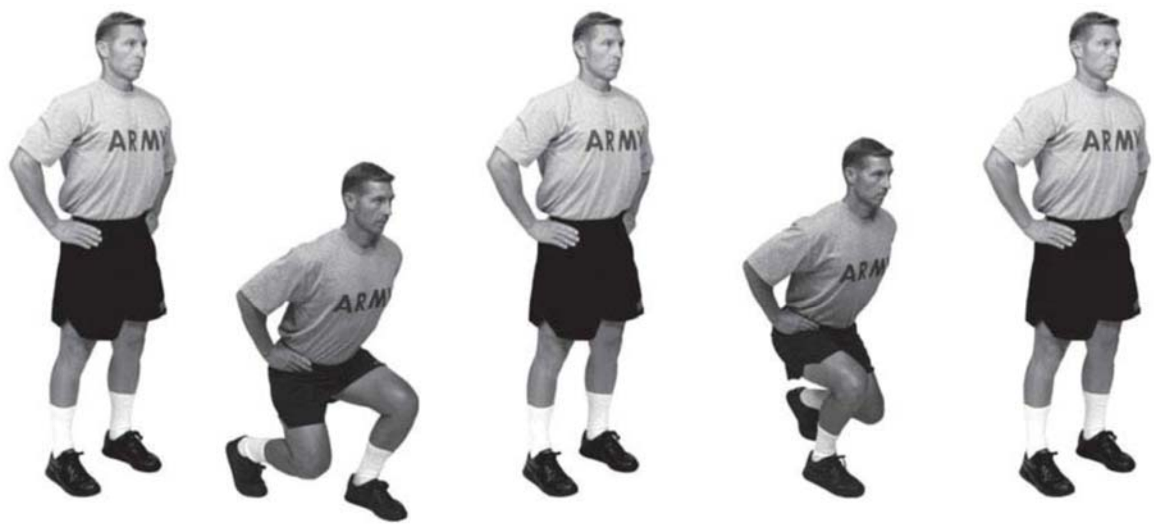


Figure 6-71. Forward lunge

Check Points:

- Keep the abdominal muscles tight throughout the motion.
- On counts 1 and 3, step straight forward, keeping the feet directed forward. When viewed from the front, the feet maintain their distance apart both at the starting position and at the end of counts 1 and 3.
- On counts 1 and 3, the rear knee may bend naturally, but does not touch the ground. The heel of the rear foot should be off the ground.

Precautions: This exercise is always performed at a slow cadence. On counts 1 and 3, move into position in a controlled manner. Spring off of the forward leg to return to the starting position. This avoids jerking the trunk to create momentum.

EXERCISE 8: PRONE ROW

Purpose: This exercise develops strength of the back and shoulders (Figure 6-73).
Starting Position: Prone position with the arms overhead, palms facing downward 1-2 inches off the ground, and toes pointed to the rear.
Cadence: SLOW
Count:
1. Raise the head and chest slightly while lifting the arms and pulling them rearward. Hands make fists as they move toward the shoulders.
2. Return to the starting position.
3. Repeat count 1.
4. Return to the starting position.

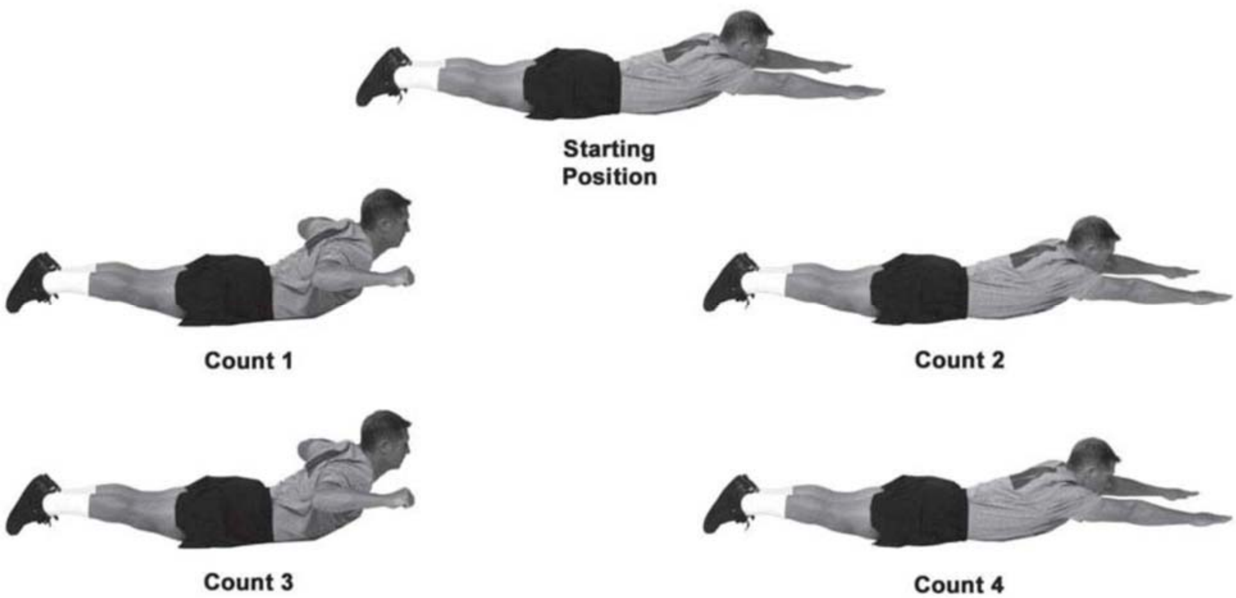


Figure 6-73. Prone row

Check Points:
- At the starting position, the abdominal muscles are tight and the head is aligned with the spine.
- On counts 1 and 3, the forearms are parallel to the ground and slightly higher than the trunk.
- On counts 1 and 3, the head is raised to look forward but not skyward.
- Throughout the exercise, the legs and toes remain in contact with the ground.

Precautions: This exercise is always performed at a slow cadence. Prevent overarching of the back by maintaining contractions of the abdominal and buttocks muscles throughout the exercise.

EXERCISE 9: BENT-LEG BODY TWIST

Purpose: This exercise strengthens trunk muscles and promotes control of trunk rotation (Figure 6-76).
Starting Position: Supine position with the hips and knees bent to 90 degrees, arms sideward, palms down with fingers spread. Knees and feet are together, and head is raised two or three inches off the ground with the chin slightly tucked.
Cadence: SLOW
Count:
1. Rotate the legs to the left while keeping the upper back and arms in place.
2. Return to the starting position.
3. Repeat count 1 to the right.
4. Return to the starting position.

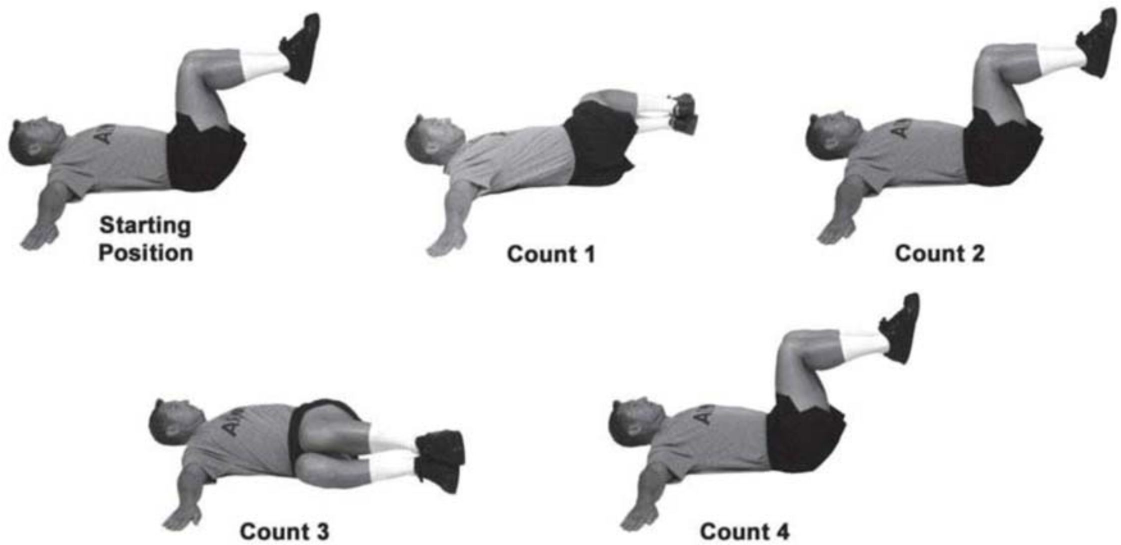


Figure 6-76. Bent-leg body twist

Check Points:
- Tighten the abdominal muscles in the starting position and maintain this contraction throughout the exercise.
- The head should be off the ground with the chin slightly tucked.
- Ensure that the hips and knees maintain 90-degree angles.
- Keep the feet and knees together throughout the exercise.
- Attempt to rotate the legs to about 8-10 inches off the ground. The opposite shoulder must remain in contact with the ground.

Precautions: This exercise is always performed at a slow cadence. Do not rotate the legs to a point beyond which the arms and shoulders can no longer maintain contact with the ground.

EXERCISE 10: PUSH-UP

Purpose: This exercise strengthens the muscles of the chest, shoulders, arms, and trunk (Figure 6-79).
Starting Position: Front leaning rest position.
Cadence: MODERATE
Count:
1. Bend the elbows, lowering the body until the upper arms are parallel with the ground.
2. Return to the starting position.
3. Repeat count 1.
4. Return to the starting position.

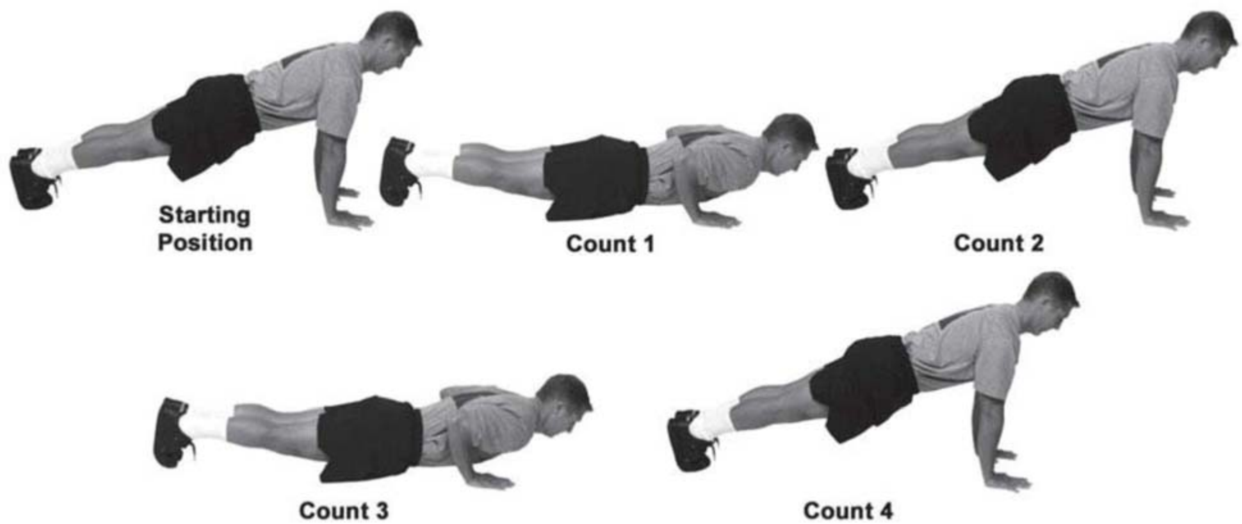


Figure 6-79. Push-up

Check Points:
- The hands are directly below the shoulders with fingers spread (middle fingers point straight ahead).
- On counts 1 and 3 the upper arms stay close to the trunk, elbows pointing rearward.
- On counts 2 and 4 the elbows straighten, but do not lock.
- The trunk should not sag. To prevent this, tighten the abdominal muscles while in the starting position and maintain this contraction throughout the exercise.

Precautions: N/A

APPENDIX F

Climatic Conditions During Testing

Table F-1. *Ambient Climatic Conditions During Test Sessions*

		October 2014				June 2015			
Test Sess.	Time	Temp. (°F)	Dew Pt. (°F)	%RH	Wind Speed ($mi \cdot hr^{-1}$)	Temp. (°F)	Dew Pt. (°F)	%RH	Wind Speed ($mi \cdot hr^{-1}$)
1	0900	45.4	42.0	89.0	0.0	69.2	65.0	86.0	1.0
	1000	51.3	44.0	75.0	0.0	70.4	66.0	86.0	1.0
	1100	58.2	48.0	70.0	1.0	74.0	67.0	79.0	3.0
	1300	67.5	54.0	62.0	2.0	74.3	68.0	80.0	1.0
	1400	70.6	55.0	58.0	1.0	73.4	68.0	84.0	1.0
	1500	71.6	55.0	57.0	0.0	77.1	69.0	77.0	3.0
2	0900	55.3	55.0	99.0	0.0	69.4	60.0	72.0	4.0
	1000	59.7	60.0	99.0	0.0	72.0	59.0	64.0	0.0
	1100	64.8	62.0	90.0	1.0	76.9	55.0	47.0	0.0
	1300	71.6	62.0	71.0	1.0	81.5	53.0	37.0	1.0
	1400	72.4	60.0	66.0	2.0	82.0	55.0	40.0	2.0
	1500	73.1	59.0	62.0	1.0	81.6	57.0	43.0	1.0
3	0900	47.1	43.0	85.0	2.0	72.4	64.0	76.0	0.0
	1000	52.6	43.0	70.0	0.0	73.7	66.0	76.0	1.0
	1100	55.5	43.0	64.0	4.0	78.4	69.0	73.0	2.0
	1300	57.9	41.0	54.0	1.0	85.6	69.0	58.0	5.0
	1400	55.4	40.0	57.0	0.0	86.3	68.0	54.0	3.0
	1500	55.7	39.0	54.0	4.0	86.5	69.0	56.0	6.0
4	0900	43.1	39.0	86.0	1.0	71.7	60.0	67.0	0.0
	1000	46.9	40.0	78.0	0.0	74.9	59.0	57.0	1.0
	1100	50.0	42.0	73.0	0.0	77.6	60.0	55.0	1.0
	1300	52.8	41.0	65.0	1.0	83.9	60.0	45.0	1.0
	1400	51.4	40.0	65.0	1.0	86.6	63.0	45.0	1.0
	1500	50.4	40.0	67.0	4.0	89.0	66.0	46.0	2.0

Note. Weather station was located about 0.7 mi (1.1 km) from the test site.

APPENDIX G

Trial Means and Results of Post Hoc Analyses for LEAP Obstacle Course Segments

Twenty-five segments of the LEAP obstacle course were individually timed. The complete list of the segments and the corresponding segment numbers are in Appendix B. For analysis purposes, the times for some of the shorter segments were combined. Some of the timed segments comprising a single obstacle were also combined. The segment numbers cited in this appendix indicate the segment times that were included in calculating the time data presented in the figures that follow.

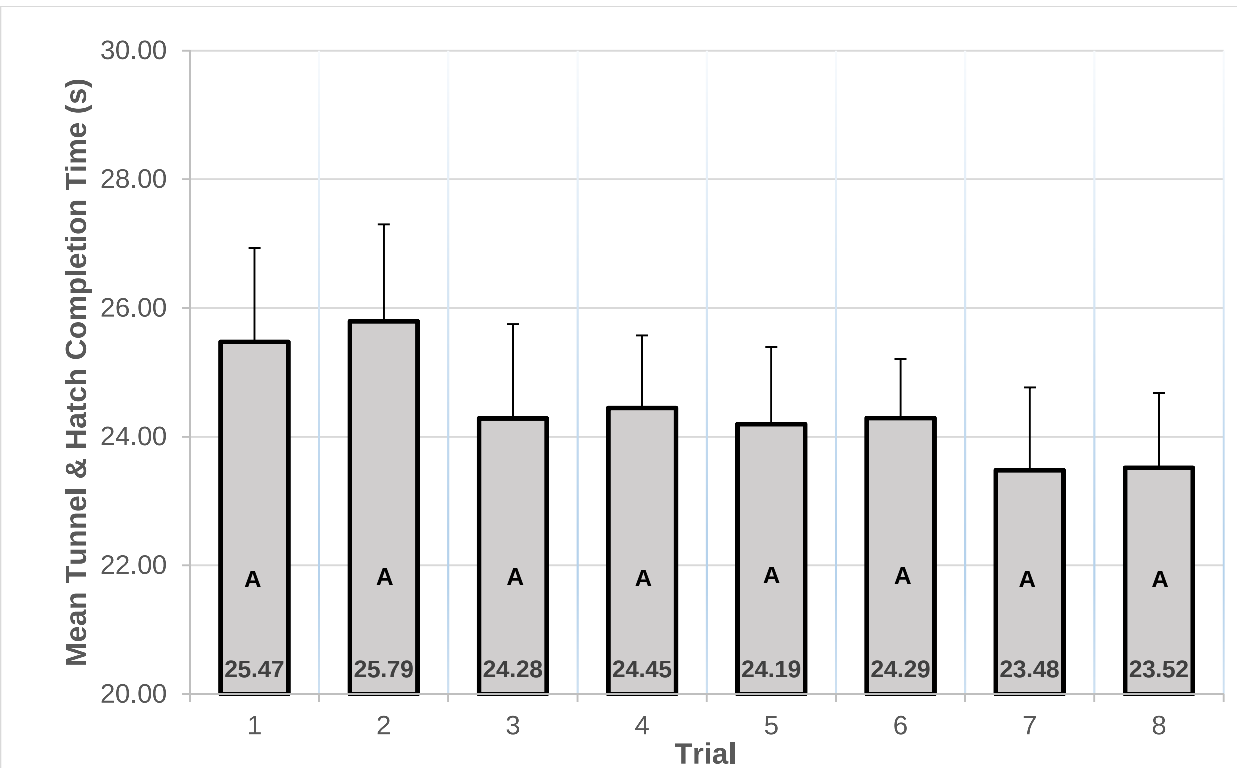


Figure G-1. Trial means (+1*SE*) and results of post hoc tests for Tunnel & Hatch obstacle (Segment 1). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

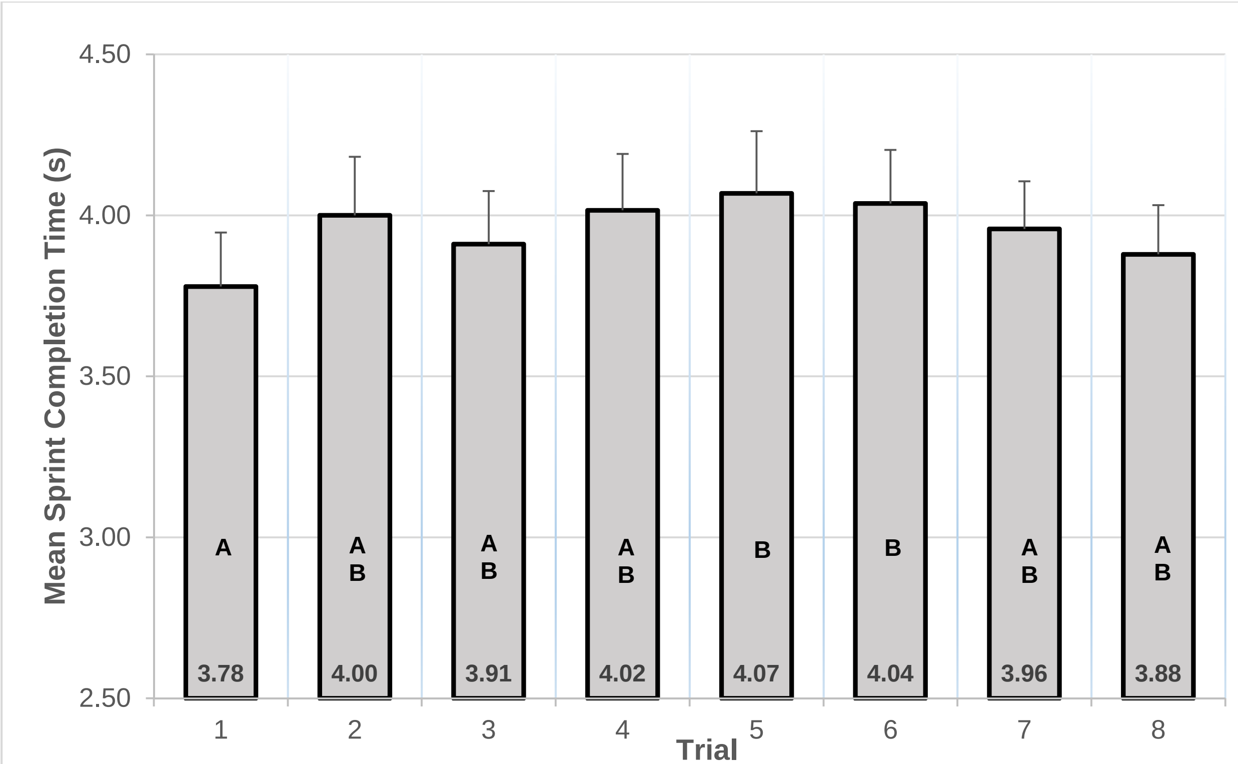


Figure G-2. Trial means (+1*SE*) and results of post hoc tests for Sprint obstacle (Segment 3). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

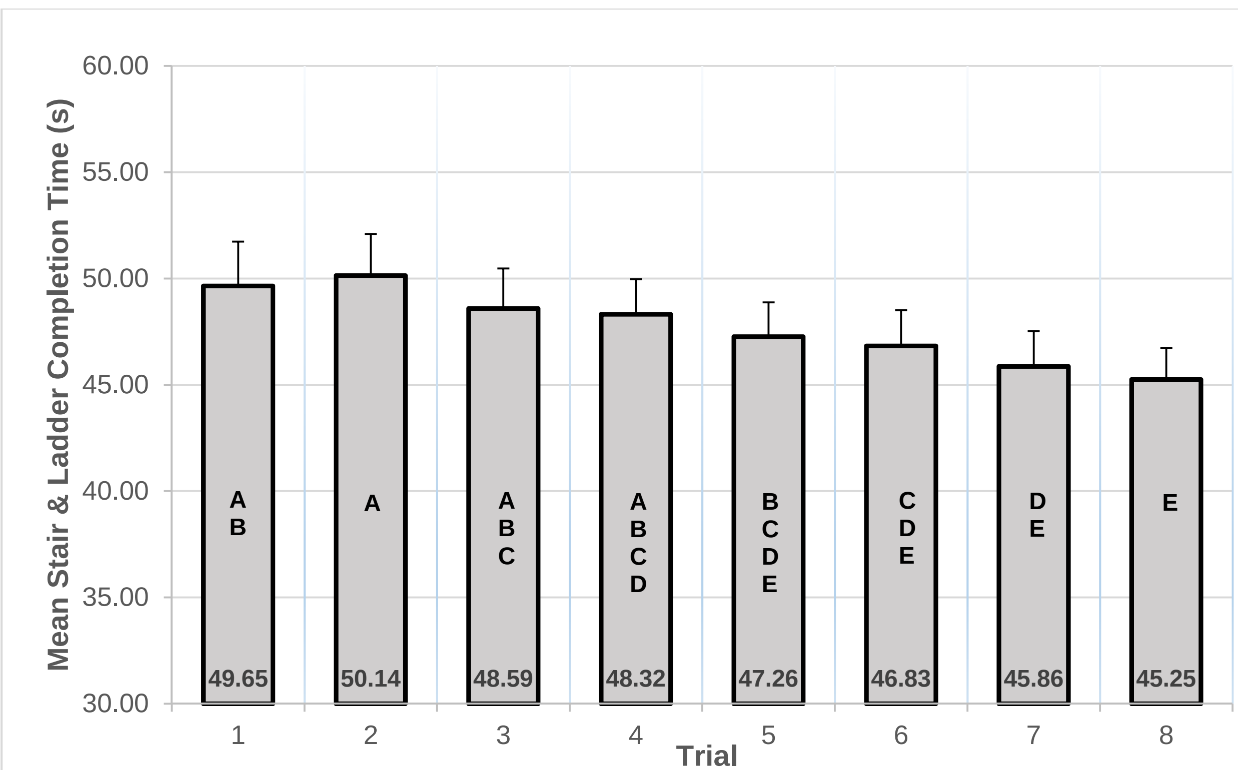


Figure G-3. Trial means (+1*SE*) and results of post hoc tests for Stair & Ladder obstacle (Segments 5-9). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

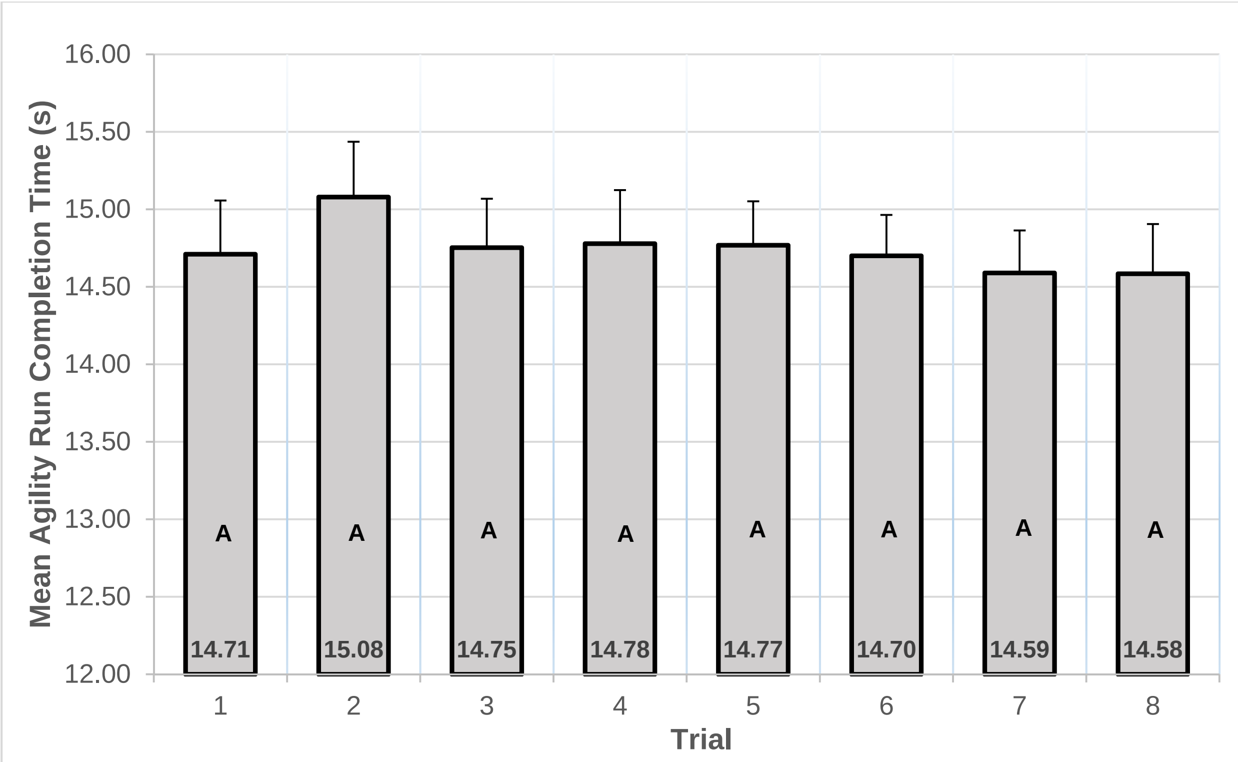


Figure G-4. Trial means (+1*SE*) and results of post hoc tests for Agility Run obstacle (Segment 11). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

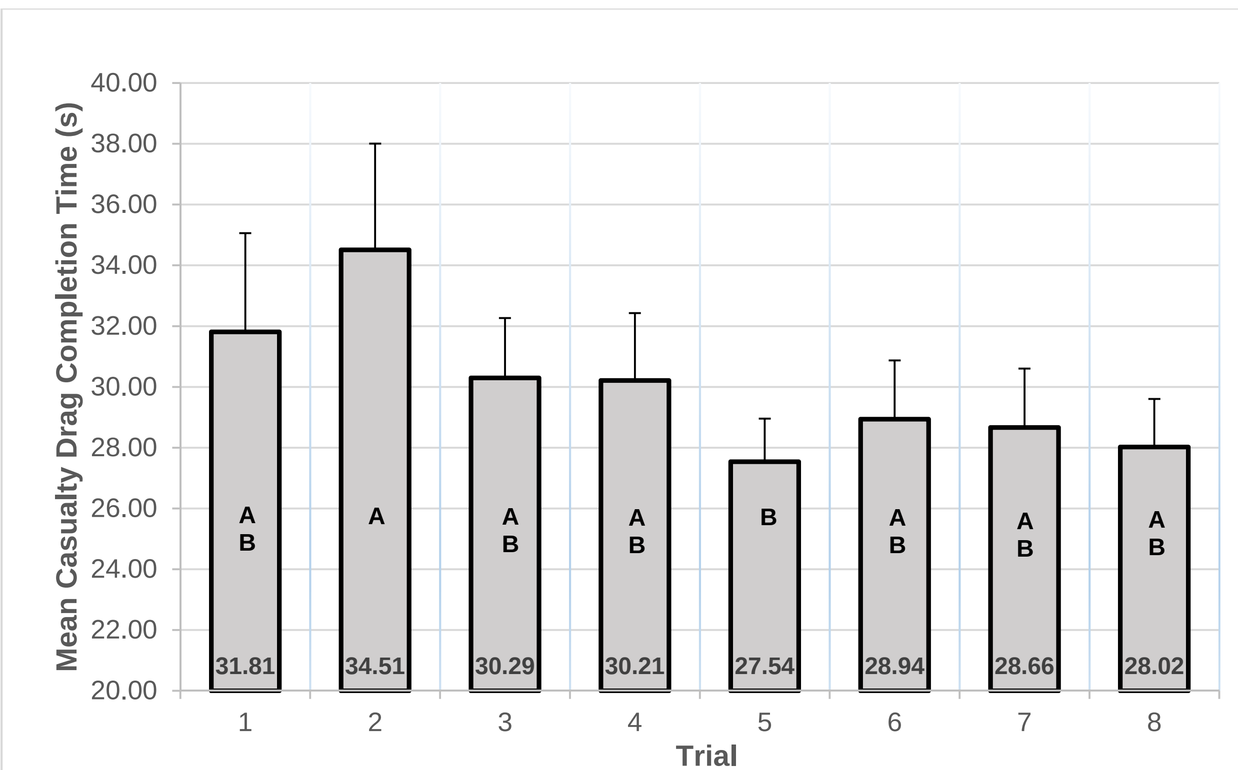


Figure G-5. Trial means (+1*SE*) and results of post hoc tests for Casualty Drag obstacle (Segment 13). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

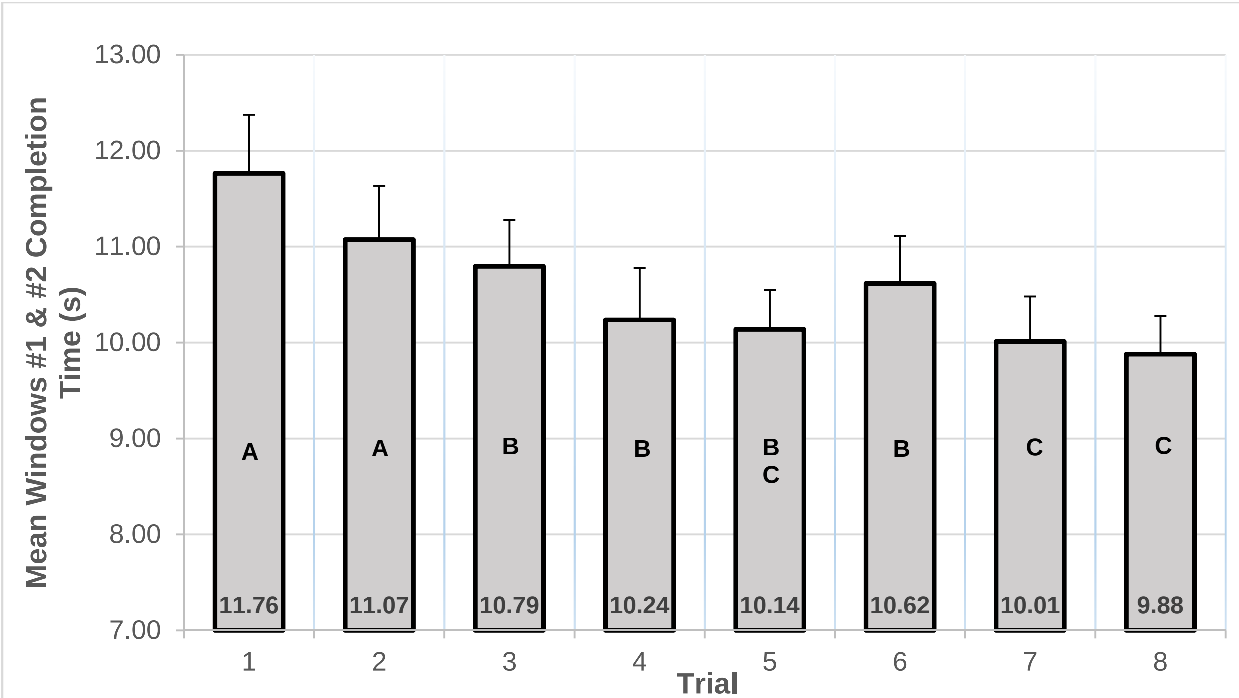


Figure G-6. Trial means (+1*SE*) and results of post hoc tests for Window #1 & Window #2 obstacle (Segments 15-16). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

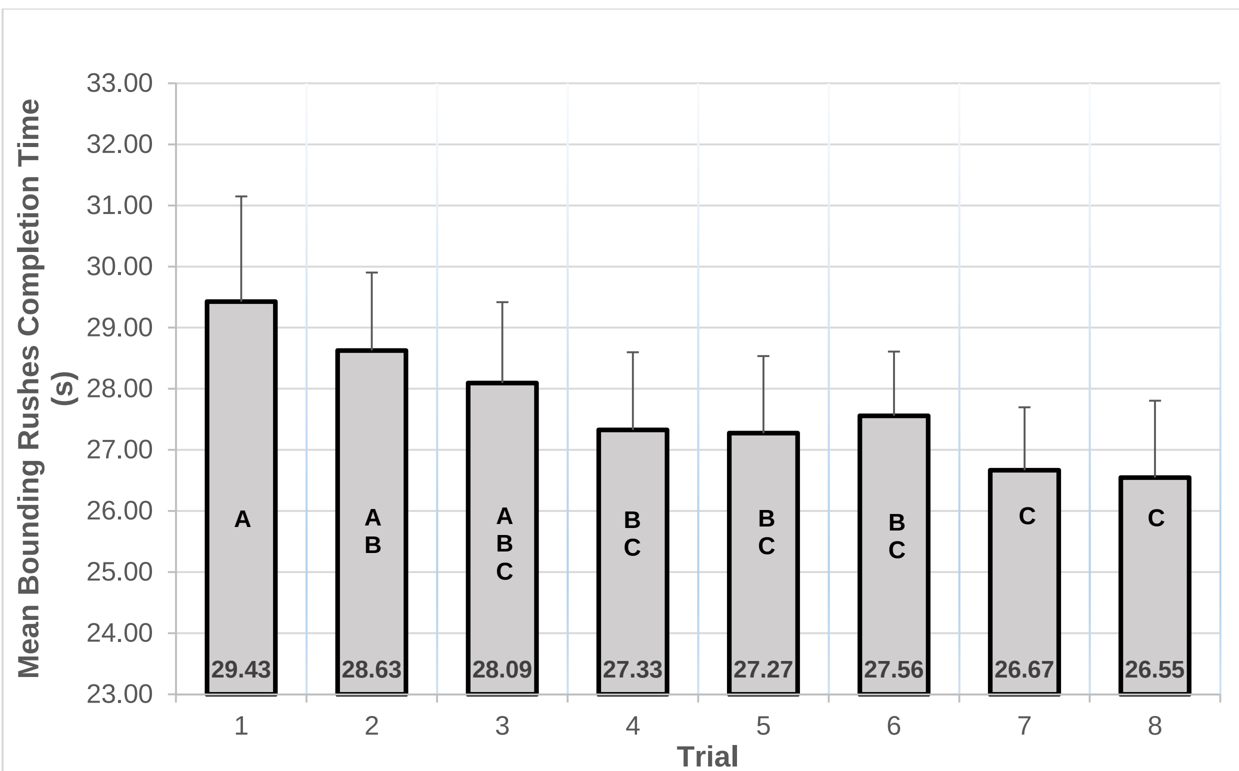


Figure G-7. Trial means (+1*SE*) and results of post hoc tests for Bounding Rushes obstacle (Segment 17). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

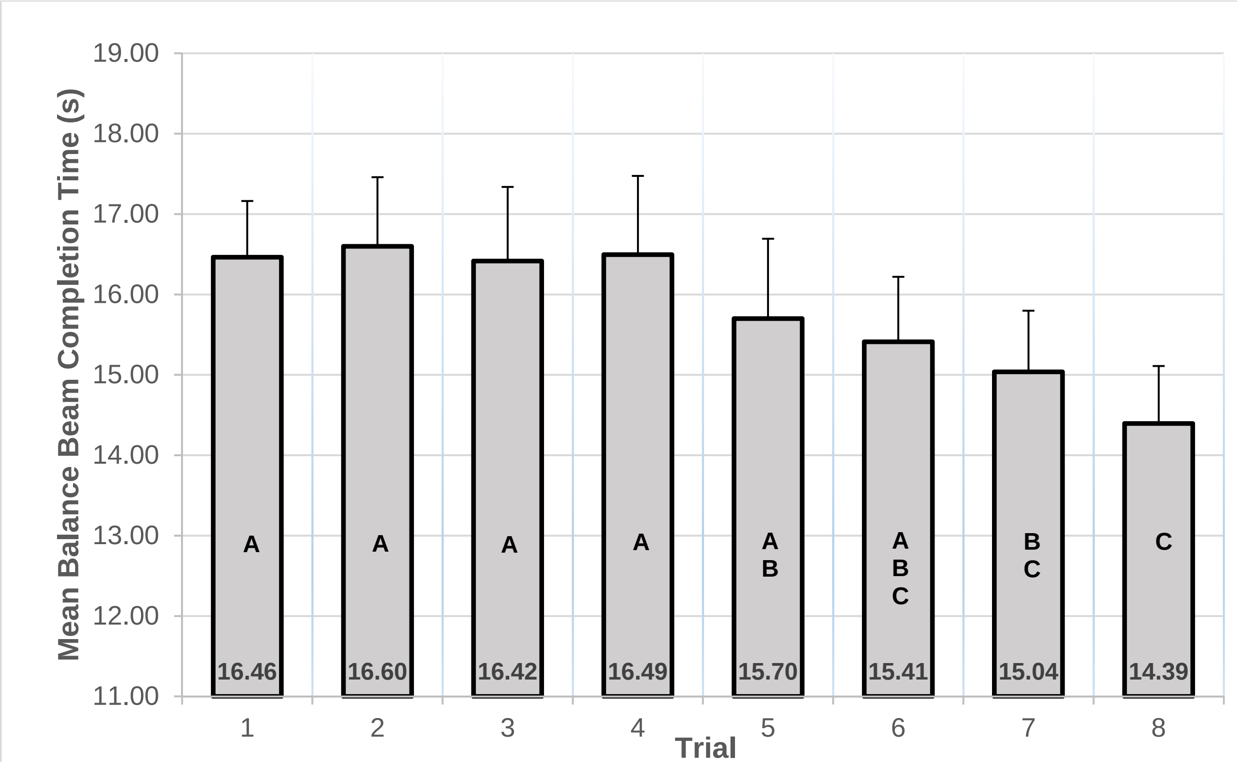


Figure G-8. Trial means (+1*SE*) and results of post hoc tests for Balance Beam obstacle (Segment 18). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

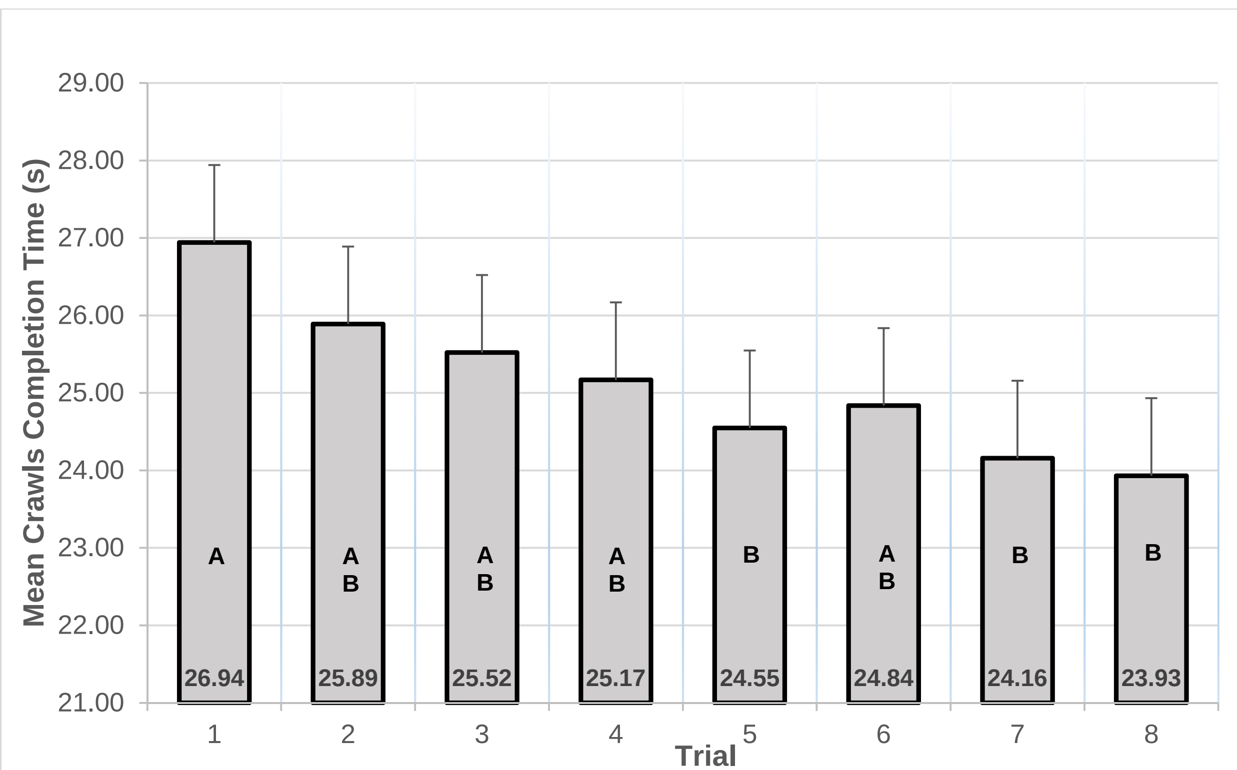


Figure G-9. Trial means (+1*SE*) and results of post hoc tests for Crawls obstacle (Segments 20-22). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

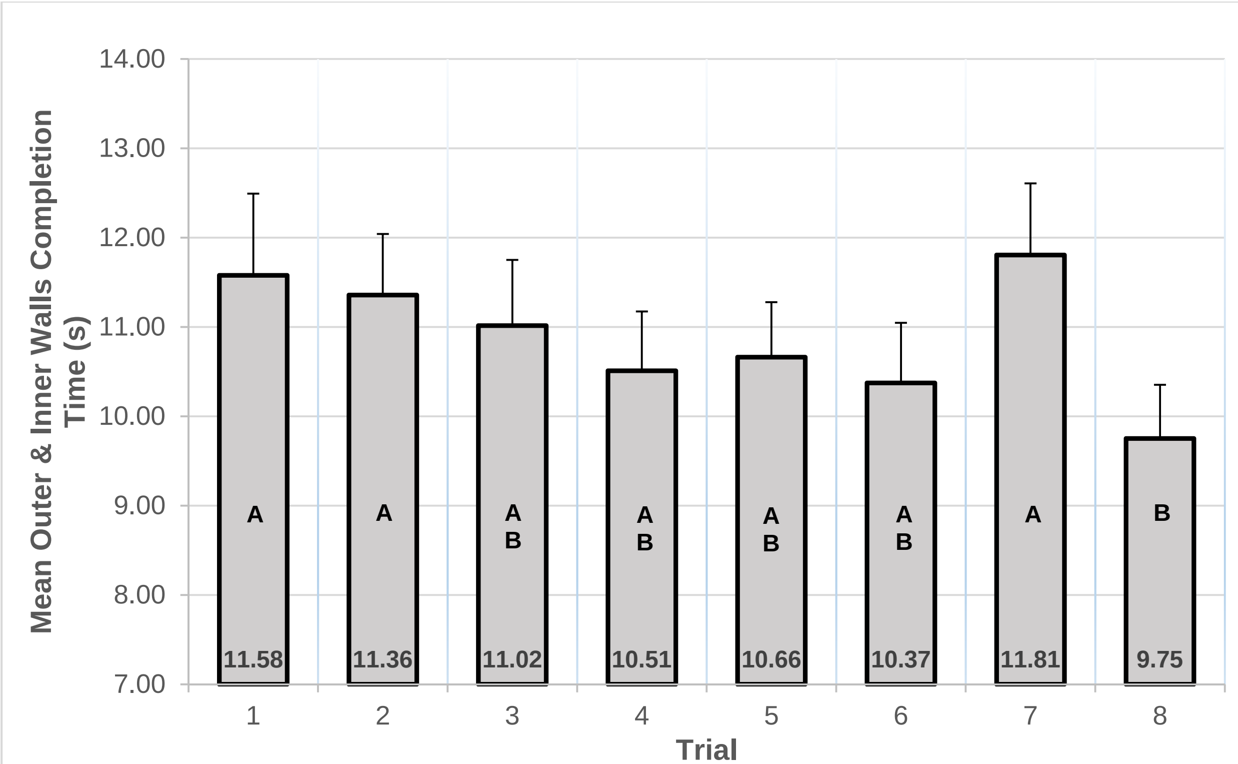


Figure G-10. Trial means (+1*SE*) and results of post hoc tests for Outer & Inner Courtyard Walls obstacle (Segments 24-25). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).

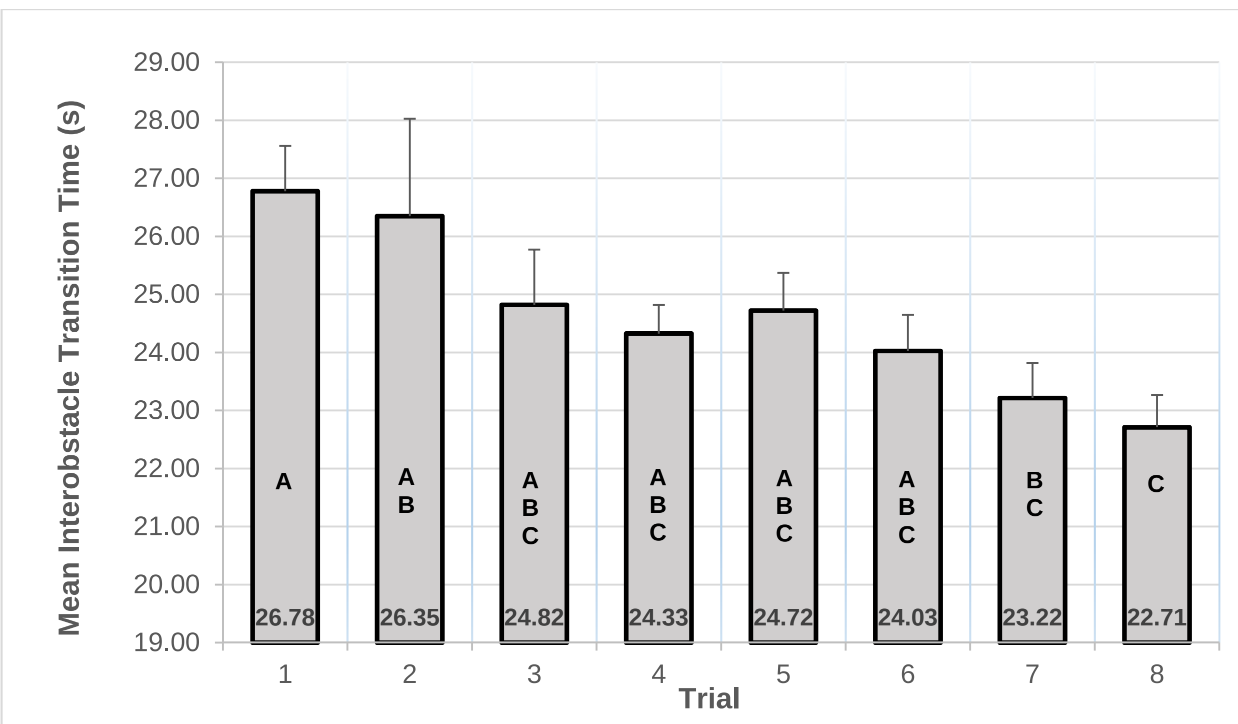


Figure G-11. Trial means (+1*SE*) and results of post hoc tests for Interobstacle Transitions (Segments 2, 4, 10, 12, 14, 19, 23). Trials that do not share the same letter differed significantly in post hoc tests ($p < .05$; $N = 19$).