U.S. Army Center for Health Promotion and Preventive Medicine

USCHPPM

THE CASE FOR PRE-ENLISTMENT PHYSICAL FITNESS TESTING: RESEARCH AND RECOMMENDATIONS

USACHPPM REPORT NO. 12-HF-01Q9D-04

US Army Center for Health Promotion and Preventive Medicine
Aberdeen Proving Ground, MD

US Army Research Institute of Environmental Medicine
Natick, MA

Center for Accessions Research
Ft Knox, KY

Readiness Thru Health

U.S. Army Center for Health Promotion and Preventive Medicine

The lineage of the U.S. Army Center for Health Promotion and Preventive Medicine (USACHPPM) can be traced back over 50 years. This organization began as the U.S. Army Industrial Hygiene Laboratory, established during the industrial buildup for World War II, under the direct supervision of the Army Surgeon General. Its original location was at the Johns Hopkins School of Hygiene and Public Health. Its mission was to conduct occupational health surveys and investigations within the Department of Defense's (DOD's) industrial production base. It was staffed with three personnel and had a limited annual operating budget of three thousand dollars.

Most recently, it became internationally known as the U.S. Army Environmental Hygiene Agency (AEHA). Its mission expanded to support worldwide preventive medicine programs of the Army, DOD, and other Federal agencies as directed by the Army Medical Command or the Office of The Surgeon General, through consultations, support services, investigations, on-site visits, and training.

On 1 August 1994, AEHA was redesignated the U.S. Army Center for Health Promotion and Preventive Medicine with a provisional status and a commanding general officer. On 1 October 1995, the nonprovisional status was approved with a mission of providing preventive medicine and health promotion leadership, direction, and services for America's Army.

The organization's quest has always been one of excellence and the provision of quality service. Today, its goal is to be an established world-class center of excellence for achieving and maintaining a fit, healthy, and ready force. To achieve that end, the CHPPM holds firmly to its values which are steeped in rich military heritage:

- ★ *Integrity is the foundation*
- ★ *Excellence is the standard*
- ★ *Customer satisfaction is the focus*
- ★ *Its people are the most valued resource*
- ★ *Continuous quality improvement is the pathway*

This organization stands on the threshold of even greater challenges and responsibilities. It has been reorganized and reengineered to support the Army of the future. The CHPPM now has three direct support activities located in Fort Meade, Maryland; Fort McPherson, Georgia; and Fitzsimons Army Medical Center, Aurora, Colorado; to provide responsive regional health promotion and preventive medicine support across the U.S. There are also two CHPPM overseas commands in Landstuhl, Germany and Camp Zama, Japan who contribute to the success of CHPPM's increasing global mission. As CHPPM moves into the 21st Century, new programs relating to fitness, health promotion, wellness, and disease surveillance are being added. As always, CHPPM stands firm in its commitment to Army readiness. It is an organization proud of its fine history, yet equally excited about its challenging future.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE August 2004	3. REPORT TYPE AND DATES COVERED Final

4. TITLE AND SUBTITLE
The Case for Pre-Enlistment Physical Fitness Testing: Research and Recommendations

5. FUNDING NUMBERS

6. AUTHOR(S)
Joseph J Knapik, Bruce H Jones, Marilyn A Sharp, Salima Darakjy, Sarah Jones, Keith G Hauret, Gene Piskator

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

US Army Center for Health Promotion and Preventive Medicine
Aberdeen Proving Ground, MD

US Army Research Institute of Environmental Medicine
Natick, MA

8. PERFORMING ORGANIZATION REPORT NUMBER
12-HF-01Q9D-04

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

Center for Accessions Research
Ft Knox, KY

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

20041008 223

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
This paper reviewed the concept of physical fitness and outlined suggestions for a pre-enlistment physical fitness test. The components of physical fitness were identified by reference to factor analysis studies and to physiological concepts related to the components of physical fitness. The physiological validity and reliability of relatively simple tests of physical fitness were considered. Criteria against which physical fitness tests have been validated include a) performance on military tasks, b) injuries and c) attrition. The review resulted in three suggested courses of action. Course of Action 1 (COA1) was to keep the current Reception Station Physical Fitness Test consisting of push-ups (PU), sit-ups and a 1-mile run. Course of Action 2 (COA2) involved a physical fitness test battery based on 1) findings in the literature, 2) the assumptions that the major components of physical fitness should be measured and 3) the assumption that fitness tests should be related to some criterion measure. The COA2 test consisted of an incremental dynamic lift (IDL), PUs and a 1-mile run. Course of Action 3 (COA3) recommended a research project that involves 6 major steps: 1) determining a set of critical military criteria, 2) determining a battery of physical fitness tests that are assumed to measure the fitness components associated with these criteria, 3) obtaining performance data on a representative sample of soldiers 4) validating and cross-validating the fitness measures against the military criteria, 5) selection of fitness test score that represent acceptable performance on the criterion tasks, 6) periodic re-evaluation of the fitness tests. The current entry-level physical fitness test (COA1) possesses some validity since individuals who do not pass the test are more likely to be injured or to attrite from service but the relationship with job performance is weaker and muscle strength is not measured by the test battery. The test battery could be immediately improved by eliminating the SU and incorporating an IDL. In the long term, an entry-level physical test should be developed through a comprehensive research program that involves well established methods of relating physical fitness tests to criterion measures important to the military like job performance, injuries, and attrition.

14. SUBJECT TERMS
Physical training, muscle contraction, aerobic endurance, VO2max, muscle strength, muscular endurance, coordination, flexibility, balance, body composition, gender, task performance, injury, attrition

15. NUMBER OF PAGES
101

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

Table of Contents

DEPARTMENT OF THE ARMY
U.S. ARMY CENTER FOR HEALTH PROMOTION AND PREVENTIVE MEDICINE
5158 BLACKHAWK ROAD
ABERDEEN PROVING GROUND, MARYLAND 21010-5403

REPLY TO
ATTENTION OF

MCHB-TS-DI

Executive Summary

THE CASE FOR PRE-ENLISTMENT PHYSICAL FITNESS TESTING: RESEARCH AND RECOMMENDATIONS

USACHPPM REPORT NO. 12-HF-01Q9D-04

1. INTRODUCTION. Many studies and reports over the years have recommended that new recruits should possess some minimum levels of physical fitness prior to entry to basic training. A baseline fitness requirement was mandated in 1999 when all new recruits were required to pass a Reception Station Physical Fitness Test on arrival for the Basic Combat Training (BCT). However, the rationale for this test and the passing criteria was unclear. The Center for Accessions Research (CAR) requested that the Army Center for Health Promotion and Preventive Medicine (CHPPM) make recommendations for a physical fitness test that could be given to Army applicants in the pre-enlistment phase. The CAR desired to move the fitness test from the reception station into the recruiting process, in order to save time and resources. The major purpose of this paper were to 1) review the concept of physical fitness, 2) review tests available to measure the components of physical fitness, 3) review previous work on pre-employment and pre-enlistment physical fitness testing, 4) recommend options for a pre-enlistment physical fitness test.

2. DEFINING PHYSICAL FITNESS. To determine an appropriate physical fitness test it was first necessary to define physical fitness. In general, physical fitness is a set of attributes that allows individuals to perform purposeful, coordinated physical activity in a satisfactory manner. The attributes or capabilities that make up physical fitness are called the "components" and these can be used to quantify physical fitness in individuals. The literature indicated that factor analysis was the major statistical technique used to identify the components of physical fitness. Factor analysis assembled physical tests into groupings that had a hypothetical common performance requirement. Complementing factor analytic studies were physiological investigations that linked specific fitness components to the physical principles involved, the energy systems recruited to fuel the activity, muscle fiber types associated with the activity, and the neuromuscular control necessary to accomplish the movement. By combining the factor analysis approach and physiological studies, the major components of physical fitness were identified as strength, muscular endurance, cardiorespiratory endurance, flexibility, coordination, and balance. Strength is the ability of a muscle group to exert a maximal force in a single voluntary contraction. Muscular endurance is the ability of a muscle group to perform short-term, high-intensity physical activity. Cardiorespiratory endurance is the ability to sustain long-term, low-power physical activity. Flexibility is the ability to voluntarily stretch, flex or otherwise lengthen various parts of the body as far as possible. Coordination is the ability to synchronize the simultaneous movement of a number of body parts. Balance is the ability to maintain the entire body in a fixed position when static, or maintain equilibrium when moving.

3. PHYSICAL FITNESS TESTS. We considered relatively simple tests of physical fitness that could be administered quickly and easily in the MEPS station or by recruiters with minimal training and equipment. We also considered the reliability and physiological validity of the test. A test is reliable if an individual produces similar scores over two or more trials. A test has physiological validity if it has a high correlation with a physiological test related to that fitness component. The analysis was limited to tests of muscle strength, muscular endurance, cardiorespiratory endurance, and body composition.

Strength can be tested either statically or dynamically as the maximum force or power that an individual exerts. There is no accepted single physiological criterion for muscular strength so it is not possible to examine physiological validity. In a sample of strength tests, reliabilities ranged from 0.62 to 0.99.

Muscular endurance tests can involve static or dynamic contractions and absolute or relative (% of an individual's maximum) loads. There is no single accepted physiological test for muscular endurance so physiological validity cannot be established. In a sample of muscular endurance tests reliabilities ranged from 0.57 to 0.90.

Tests of cardiorespiratory fitness include a) maximal effort runs for time over fixed distances, b) maximal effort runs for fixed times completing as much distance as possible, and c) aerobic shuttle run tests. Physiological validity of cardiorespiratory endurance tests can be determined using VO_2max which is proportional to the maximal rate at which energy can be supplied to fuel longer-term physical activity. Physiological validity coefficients ranged from 0.28 to 0.95 with 41 of 60 sampled coefficients being greater than 0.70. Reliability coefficients range from 0.78 to 0.98.

Estimates of body composition can be obtained from anthropometric measures such as circumferences, skinfolds, girths, and diameters. Physiological validity has been established by relating the anthropometric measures to body composition determined from densitometry (underwater weighing) and other methods. Physiological validity coefficients in a sample of studies ranged from 0.68 to 0.92. Military specific equations using circumferences have been developed for Army, Navy, Air Force and Marine samples with physiological validities ranging from 0.80 to 0.82 and standard errors of estimate ranging from 3.1% to 3.3% body fat.

4. CRITERIA FOR SELECTION OF PHYSICAL FITNESS TESTS. Our approach to developing a pre-enlistment physical fitness test was to determine criteria that are important from a military standpoint and examine the relationship of these criteria to various measures of physical fitness (criterion-related validity). Military criteria that have been described as important in the literature include job performance, injuries, and attrition from service. With regard to job performance, the Equal Employment Opportunity Commission (EEOC) has published Uniform Guidelines on Employee Selection Procedures which define acceptable criteria for a pre-employment selection test. A large civilian literature has developed on the association between physical fitness tests and occupational task performance apparently motivated by efforts to comply with these guidelines. Sampled studies show correlations between job tasks and physical fitness measures ranging from 0.57 to 0.95. Military studies conducted in the British, Canadian, Dutch, and United States Armies generally show that a wide variety of measures of muscle strength, muscular endurance, cardiorespiratory endurance, and body composition are related to performance of specific military tasks involving lifting, lifting and carrying, repetitive lifting, road marching, digging, and casualty evacuation. Other studies show that military personnel have a higher likelihood of injury if they have: a) low performance on 1-mile runs, 1.5-mile runs, 2-mile runs, aerobic shuttle runs, or 3000 m runs, b) low performance on sit-ups or push-ups, c) both high and low extremes of flexibility as measured by the sit-and-reach. Attrition from service is related to lower performance on push-ups, sit-ups, 2-mile runs, aerobic shuttle runs, pull-ups and the incremental dynamic lift, and injury.

5. RECOMMENDATIONS FOR AN ENTRY-LEVEL PHYSICAL FITNESS TEST. Three courses of action for a pre-accession physical fitness test were identified. Course of Action 1 (COA1) is to keep the current Reception Station Physical Fitness Test consisting of push-ups, sit-ups and a 1-mile run. We examined individuals who did and did not pass the test based on the current criteria and entered BCT without further physical training. Compared to individuals who passed the test, those who did not pass the test were 1.6 to 3.9 times more likely to get injured and 1.9 to 3.2 times more likely to attrite from training. Thus, the current test has some validity if the validity criterion involves injury or attrition. The relationship of the test with military job performance is weaker and the test does not measure muscle strength (it does measure muscular endurance and cardiorespiratory endurance).

Course of Action 2 (COA2) suggests a physical fitness test battery based on findings in the literature. Two assumptions are made: a) that the major components of physical fitness (muscle strength, muscular endurance, cardiorespiratory endurance) should be measured, and b) that the fitness tests should be related to some criterion measure. COA2 involves a test incorporating the incremental dynamic lift (IDL), PUs and a 1-mile run. The passing criteria for PU and the 1-mile run remain the same as in COA1. The criteria for passing the IDL are based on MOS. For MOS that have light, medium or moderate lifting requirements as defined in Army Regulation 611-201, the requirement is to lift 40 lbs. For MOS having heavy or very heavy lifting requirements as defined in Army Regulation 611-201, the requirement is to lift 70 lbs. The IDL has been shown to be related to a variety of military tasks while PUs and the 1-mile run have been shown to be related to injuries and attrition.

Course of Action 3 (COA3) complies with the EEOC guidelines on employee selection procedures and takes advantage of information and techniques garnered from past military and civilian studies on pre-employment testing. COA1 recommends a research project that involves 6 major steps: 1) determining a set of critical military criteria, 2) selecting a battery of physical fitness tests that measure the fitness components associated with these criteria, 3) obtaining performance data on a representative sample of soldiers 4) validating and cross-validating the fitness measures against the military criteria, 5) selecting fitness test scores that represent acceptable performance on the criterion tasks, 6) periodic re-evaluation of the fitness tests to account for technological changes in equipment and materials and for changes in the level of fitness of potential military recruits.

6. CONCLUSIONS. Several studies show that the current entry-level physical fitness test possesses some validity since individuals who do not pass the test are more likely to be injured or to attrite from service. However, the current physical fitness entrance test could be immediately improved by eliminating the SU and replacing it with the IDL. In the long term, an entry-level physical test should be developed through a comprehensive research program that involves well established methods of relating physical fitness tests to criterion measures important to the military like job performance, injuries, and attrition. A physical fitness test battery established from these research procedures would have a strong rational basis, be legally defensible, and would place testing of the physical capability of potential recruits on a footing similar to cognitive ability testing which has been performed since WWI.

DEPARTMENT OF THE ARMY
U.S. ARMY CENTER FOR HEALTH PROMOTION AND PREVENTIVE MEDICINE
5158 BLACKHAWK ROAD
ABERDEEN PROVING GROUND, MARYLAND 21010-5403

REPLY TO
ATTENTION OF

MCHB-TS-DI

THE CASE FOR PRE-ENLISTMENT PHYSICAL FITNESS TESTING: RESEARCH AND RECOMMENDATIONS

USACHPPM REPORT NO. 12-HF-01Q9D-04

1. REFERENCES. References used in this report are in Appendix A.

2. INTRODUCTION.

Many studies and reports over the years have recommended that new recruits should possess some minimum level of physical fitness prior to entry to basic training (76,77,109,238,250,261). A 1984 US Army Training and Doctrine Command (TRADOC) study group examining the Army Trainee Discharge Program (now called Entry Level Separation) noted that many recruits arrived in poor physical condition and that this lack of physical conditioning was a major reason for discharges. They recommended a physical fitness prescreening in the Military Entrance Processing Station (MEPS) (250). A 1999 report on basic training discharges noted that 26% of recruits given an entry level separation failed their first APFT and over 70% of these failed multiple events. APFT failure was among the 3 most common items found on counseling statements. Analogous to the requirement for educational and intelligence credentials required for service entry, the report recommended a fitness screening prior to service (261). A 1998 General Accounting Office (GAO) report (77) indicated that service officials acknowledge that poor physical condition of recruits contributes to attrition. The GAO recommended that the Secretary of Defense implement a policy of administering fitness tests to recruits before basic training and the Acting Assistant Secretary of Defense concurred with this recommendation.

Based on a program first conducted at Ft Jackson South Carolina in 1998 (146), a physical fitness requirement for entry into service was mandated for all 5 Army Basic Combat Training (BCT) posts in 1999 (249). The requirement called for a specific physical fitness test that was to be given to all trainees on arrival at the BCT reception station. Trainees who failed the test were given a special physical training program in the reception station and once the trainee could pass test he or she could begin BCT. The 3 test events and passing standards are shown in Table 1. The tests were administered in the order shown and a recruit who could not meet the standard on any one event was considered a test failure.

At Ft Jackson, the Army's largest BCT post, trainees had only one try to meet the sit-up (SU) and the 1-mile run standard. On the push-up (PU) test, if the trainee failed on the first attempt, they were given specific, individualized instruction on how to perform a correct PU and a second attempt was allowed. For the 1-mile run, recruits were provided a "pacer" who ran at the exact pace required to pass the test. In addition, "chasers" attempted to motivate recruits who fell behind the pacer and reminded recruits where the pacer was located. While some research has been conducted on the validity of this test (146,148), the rationale for the test events and the passing standards remains unclear.

Table 1. Fitness Criteria to Enter BCT

Event	Men	Women
Push-ups (repetitions)	13	3
Sit-ups (repetitions)	17	17
One-Mile Run (minutes)	8.5	10.5

The Center for Accessions Research (CAR) requested that the Army Center for Health Promotion and Preventive Medicine (CHPPM) determine courses of action for a physical fitness test that could be given to Army applicants in the pre-enlistment phase. The original concept was to have a test in the MEPS to save the time and expense of shipping the recruit to the reception station and maintaining an infrastructure to train low fit recruits. However, since the original CAR request, Recruiting Command took independent action to have all recruiters administer a fitness test as a condition of enlistment. The exact test has not yet been determined as of this writing.

The major purpose of this paper is to outline suggestions for a pre-accession physical fitness test. Three courses of action were determined and the rationale for each is provided. The paper is organized to first define and analyze the concept of physical fitness to achieve a common understanding of the concept for the purposes of this paper. Tests of physical fitness will be outlined so the variety of available fitness tests can be appreciated. The civilian and military literature involving pre-employment/pre-accession testing will be reviewed but emphasis will be placed on previous studies of military pre-enlistment testing conducted in the US and foreign countries. Finally, courses of action for selective pre-enlistment physical fitness tests will be suggested along with the rationale for each course of action.

3. DEFINITION OF PHYSICAL FITNESS

Before making recommendations on physical fitness tests for pre-accession screening we need to define the concept of physical fitness. In the literature there appears to be general agreement on what constitutes physical fitness but different authors have defined the term in somewhat different ways (28,37,53,92,117,204,257). A commonly cited definition is "the ability to carry out daily tasks with vigor and alertness, without fatigue, and with ample energy to enjoy leisure-time pursuits and to meet unforeseen emergencies" (1). Another definition is "a set of attributes that relate to the ability of people to perform physical activity" (185). The World Health Organization defined fitness as "the ability to perform muscular work satisfactorily" (28). Fleishman (69) calls fitness "the functional capacity of individuals to perform certain kinds of tasks requiring muscular activity". Daniels et al. (53) defined fitness for the purposes of the US Army as "those factors which determine one's ability to perform heavy physical work and contribute toward maintaining good health and appearance".

The term "physical fitness" implies the ability to move in an energetic, optimal, or at least satisfactory manner (i.e., "fitness") in the corporeal (i.e., "physical") environment. For human movement to occur, muscular contraction is needed and to accomplish a task, muscular contractions must be coordinated and goal directed. Physical fitness is not a single characteristic but has a number of attributes or components that can be identified and quantified. Based on these physical and physiological considerations a more appropriate definition of physical fitness might be "a set of attributes that allows individuals to performance of purposeful, coordinated physical activity in a satisfactory manner".

These definitions provide a very broad description of physical fitness but most are very general and do not afford a way of measuring fitness. Since the 1930s a large number of studies have contributed to refining the concept of physical fitness by describing the specific types of behaviors, attributes or capabilities involved in the concept. These behaviors, attributes or capabilities are termed the "components" of physical fitness and these components provide a way to quantify physical fitness. In reviewing the literature we found that there were two broad approaches that had been used to determine the components of physical fitness. These might be termed the Factor Analytic Approach and the Physiological Approach. These approaches are complementary and provide different types of validity for the concept of physical fitness.

The Factor Analytic Approach was named after the statistical technique used by investigators in this field who were primarily physical educators. The Factor Analytic Approach involved presenting individuals with a broad array of physical tasks for which quantitative performance measures could be obtained. Correlational and factor analysis techniques were used to assemble the physical tasks into groupings that were assumed to have a hypothetical common performance requirement. Over time, and using many types of physical performance tasks, a number of constructs or fitness components were identified (21,45,47,50,67,68,69,70,115,166,186,197,204,207,276). In individual studies, the factors or components that were identified depend to a large extent on the tests that were administered as part of the test batteries. Early studies concentrated on various measures of strength and few studies included what we would now consider cardiorespiratory endurance measures. As particular factors emerged, later studies included additional tests that might be related to a particular factor and the components of physical fitness was further refined. The

Factor Analytic Approach provided construct validity because tests that were supposedly related to the construct of physical fitness were found to have a specific "structure" identified by the tests.

The second approach to identifying fitness components can be termed the Physiological Approach. The Physiological Approach characterized the components of fitness by describing the physical principles involved, the energy systems recruited to fuel specific fitness components, muscle fiber types associated with the activity, and the neuromuscular control necessary to accomplish the movement (85,116,153,185,257). The Physiological Approach ties fitness components to the underlying physiological and metabolic factors.

a. Components of Physical Fitness Identified by Factor Analysis

To determine fitness components identified in factor analytic studies we modified an approach used by Nicks and Fleishman (201) and Fleishman (69). An Excel® file was created that contained factors identified in each study along with the physical fitness tests and the rotated factor loadings. By sorting the file by tests, factors, and factor loadings, an effort was made to identify common factors in each study. In many cases factor names were relatively consistent across studies. However, some studies might give a particular factor an unusual name but it was apparent from the tests and the factor loading what the factor had been named in other studies. The names we gave to particular factors were those most commonly used in the literature and those most descriptive of the general fitness component. Although we reviewed and analyzed individual articles, we depended heavily on the work of Fleishman and colleagues (69,70) to help identify specific fitness factors since their work on categorization of physical abilities was the most comprehensive.

(1) Strength

There is strong support for a fitness component that is generally termed strength (18,29,36,43,47,50,51,67,69,99,106,110,119,122,124,127,129,165,166, 180,186,187,188,193,197,207,214,233,263). This factor is characterized by tests that involve exerting as much force as possible in a single voluntary effort lasting for a very short period of time (less than about 5 seconds). In studies that included an adequate number and variety of tests in the test battery, some additional subcomponents emerge that might be termed static strength and power. In addition, some research suggested that the strength component of fitness should be further broken down into upper body, lower body and trunk strength.

(a) Static Strength. Many factor analytic studies have defined a separate static strength factor (36,67,69,99,107,119,124,127,129,165,166,197,207,214, 233). This factor is characterized by tests that involve the ability to voluntarily exert maximal force against a fairly immovable object for a brief period of time (less than about 5 sec). Tests that have typically demonstrated high factor loading on this fitness component include isometric tests involving hand grip, knee extension, elbow flexion, or shoulder extension (36,69,119,127,166,197).

(b) Power. A factor that can be termed power is identified in a wide variety of investigations (36,43,47,50,67,69,99,106,107,110,119,124,186,189, 193,197,207,214). This fitness component is characterized by tests that involve rapidly projecting objects or rapidly projecting the body in a single maximal effort. Tests that have high factor loadings on this fitness component include the standing broad jump, vertical jump, softball throw, shot put, medicine ball throws, and short sprints. Many names have been ascribed to this fitness component including velocity (36,47,99,119), speed (43,107,110,193,207,214), sprinting (124), energy mobilization (106) and explosive strength (67,69,197). Since the central characteristic is the ability to develop force rapidly, and power is force/time, power seems like an appropriate term for this fitness component.

(c) Upper Body, Lower Body and Trunk Strength. Some factor analytic or cluster analytic studies provide support for separate strength factors for the upper body, and lower body (26,47,50,110,123,124,214), although there is conflicting evidence (69,197). There was weak support for the existence of a separate trunk strength factor in one study (123).

(2) Muscular Endurance

This factor has been consistently identified in a large number of studies (21,47,50,67,69,106,165,166,186,189,197,228,276). The muscular endurance component of fitness is characterized by tests that involve repeated high intensity muscular contractions for relatively short periods of time (less than about 2 minutes) while supporting the body or supporting an external weight. The number of muscular contractions that can be performed progressively decreases over time. The performance measure is usually how many contractions can be performed in a set period of time or until fatigue, or how long an isometric contraction can be held. Tests that have high factor loadings on this component of physical fitness include push-ups, pull-ups, and dips.

Many studies have called this factor dynamic strength (29,67,69,95,165,166,197). However, this term is likely to confuse the factor with a more general strength concept mentioned above. The term muscular endurance avoids this confusion and has more general acceptance in the physical education and epidemiological communities (37,204). Other names ascribed to this factor include dynamic gross motor ability (228), limb strength (106), and strength/endurance (21,276).

(3) Trunk Muscular Endurance

There is strong support for a separate trunk muscular endurance factor (21,29,47,67,69,106,197,207,276). This factor is characterized by tests that involve repeated high intensity contraction of trunk muscles for relatively short periods of time (less than about 2 minutes). Tests with higher factor loading on this component include leg lifts, V-sits and SUs.

(4) Cardiorespiratory Endurance

Surprisingly few factor analytic studies (21,67,69,106,180,186,197,276) have identified this factor despite the strong support for it in the physiological literature (185). This is apparently because few of the early factor analytic studies included tests of sufficient duration to tax the cardiovascular system. In fact, it was not until 1971 that a factor analytic study included a running test that involved distances longer than 300 yards (21). Cardiorespiratory endurance is characterized by tests that involve low intensity muscle contractions that are sustained for long periods of time. Tests that demonstrate high factor loadings on this fitness component include time to run specific distances, distances completed in specific times, heart rate counts on step tests or cycle ergometers, or maximal oxygen uptake (VO_2max) tests. This factor has been called stamina in some studies (67,69,197).

(5) Flexibility

A few studies have defined a separate fitness component that is termed flexibility (29,67,106,180,197). Few early factor analytic studies contained tests that could isolate this factor. This fitness component is characterized by tests that involve stretching, flexing or otherwise lengthening various parts of the body as far as possible. It involves the suppleness of the muscles, tendons, ligaments and other structures of a single joint while the rest of the body is held static. Tests that demonstrate high factor loading on this fitness component include the sit-and-reach, toe touching, and twist and reach. There are studies indicating that flexibility is specific to the joint being measured (58,100).

Fleishman (67) isolated factors for 2 types of flexibility that he termed extent flexibility and dynamic flexibility. Extent flexibility is defined in the paragraph above. Dynamic flexibility was proposed to have a speed component requiring rapid movement of the trunk or limbs reaching long distances. Examination of the tests that load on this component suggests this factor may relate more to speed of movement rather than extending body parts to maximal distances (50,51,67).

(6) Coordination

A coordination factor has been identified in a number of studies (49,50,106,165,187,193,263). Various names have been used to describe this factor including gross body coordination (106,165), agility/coordination (49,50), motor educability (193,263), sensorimotor control (263) and large muscle coordination (187). This factor is characterized by complex tests that require synchronizing the simultaneous movement of a number of body parts. Tests that demonstrate high loadings on this factor include squat thrusts, cable jumps, and sports skills like basketball shooting and catching a ball. Fleishman (69) included three tests that he hypothesized would involve coordination but was unable to isolate a separate coordination factor.

(7) Balance

A few studies have isolated a balance factor (19,49,51,67,106,122) but this factor has not been well characterized because few factor analytic studies have included tests that might involve this fitness component. Balance is characterized by tests that involve either maintaining the entire body in a fixed position when either static or maintaining equilibrium when moving. Tests that demonstrate high factor loadings on the balance component include standing on one foot, rail walking and rail balancing.

There is some suggestion that separate balance factors may exist dependent on whether the eyes are open or closed (19,67,122). Fleishman (69) distinguished between gross body equilibrium and balance with visual cues. Gross body equilibrium appears to involve the ability to maintain balance when forces are attempting to disrupt that balance and the main cues are vestibular and kinesthetic; however, some tests involving visual cues also had relatively high loadings on this factor. The tests that best characterized this factor were balancing on a beam with eyes closed and rail walking with eyes open. Balance with visual cues more clearly involves vestibular, kinesthetic and visual sensory input to maintain balance. The test that best characterized this component was balancing on a beam with eyes open.

Cumbee (49) suggested that there was a separate balancing objects factor but this was only partly supported in a follow-up study (51) and has not been supported in an independent study designed to measure this potential factor (67). Considerably more work needs to be done to determine the structure of the balance component of physical fitness.

(8) Body Weight, Body Fat, Muscle Mass

By including body weight in the factor analyses several studies have identified relationships between body weight and other fitness components. Body weight is generally negatively associated with whole body power tests like the broad jump, vertical jump and short sprints (36,43,99,214) and positively associated with upper body power tasks (99,188). Excessive weight would be a disadvantage on tests requiring powerful whole body movements because of the additional mass that would have to be moved. The positive relationship with upper body power tasks may reflect the muscle component of the body weight.

Three studies included measures of body fat from skinfolds (18,180,199). Generally, body fat was found to be negatively associated with tests of cardiorespiratory endurance and tests that require leg power.

b. Ability Requirements Approach

Fleishman and Quaintance (70) developed the Ability Requirement Approach. The general objective of the Ability Requirement Approach was to describe the fewest independent ability categories that were useful and meaningful in describing human performance on the widest possible variety of tasks. The physical proficiency factors described in Table 2 were identified using the factor analytic techniques described above (68,69,70,115,197). Note that the Ability Requirement Approach does not attempt to identify physical fitness components *per se* but rather it attempts to characterize human physical capabilities in both physical and cognitive domains. The capabilities shown in Table 2 are those that require primarily physical rather than cognitive performance.

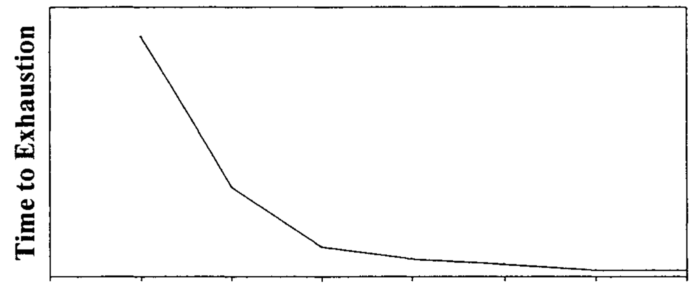
Table 2. Human Physical Capabilities Defined from the Ability Requirement Approach (from Reference Number 70)

Physical Capability	Definition
Static Strength	Ability to exert maximal strength against a fairly immovable object
Explosive Strength	Ability to expend a maximum of energy in one burst or a series of bursts
Dynamic Strength	Ability to exert muscular force repeatedly or continuously over time
Trunk Strength	Ability to exert muscular force of the trunk muscles repeatedly or continuously over time
Stamina	Ability to sustain physical effort involving the cardiovascular system
Gross Body Coordination	Ability to perform movements that simultaneously involve the entire body
Gross Body Equilibrium	Ability to maintain or regain body balance, especially where equilibrium is threatened
Extent Flexibility	Ability to extend or stretch the body
Dynamic Flexibility	Ability to move trunk and limbs quickly and through a wide range of motion

c. Relationships Between Fitness Components and Physiological Factors

The Physiological Approach refines the Factor Analytic Approach by linking the components of physical fitness to particular physical and physiological characteristics (85,116,153,185,257). It thus provides another type of validity for the components of physical fitness identified by the Factor Analysis Approach. The Physiological Approach shows body composition to be an important fitness component because the quantity and distribution of muscle, fat and other tissue will largely determine the capacity for different types of physical activity. Figure 1 shows the relationship between time to exhaustion and various physical and physiological measures. Figure 1 is a useful reference for the discussion that follows.

Figure 1. Relationship Between Exhaustion Time and Various Physiological and Physical Measures



Power: Lo------------------------Moderate--------------------------Hi
Contraction Force: Lo-------------------------Moderate--------------------------Hi
Energy System: Glucose/Fat---------------Glycogen/Glucose-------ATP-CP--ATP
Oxygen Req.: Oxidative----------------------------------Non-Oxidative
Fiber Type: ST--FT-----------FT/ST
Contraction No.: High------------------------------Moderate--------------------Low
Fitness Component: CR Endur--------------------------M.Endur---Power--M.Strngh

Abbreviations: Req.=Requirement, No.=Number, ST=Slow Twitch, FT=Fast Twitch, CR Endur= Cardiorespiratory Endurance, M.Endur=Muscular Endurance, M.Strngh=Muscular Strength

(1) Energy Production for Physical Activity

Physical activity requires muscular contraction. Energy for muscular contraction is derived primarily from splitting phosphagen molecules from adenosine triphosphate (ATP) located in the active muscles. The supply of ATP can last only a few seconds but ATP can be rapidly replenished by creatine phosphate (CP) in the active muscle. This ATP/CP system can only supply energy for a few more seconds. As the length of the activity increases further ATP can be replenished by the enzymatic breakdown of glycogen (in the muscle), or glucose (primarily from the liver) in the glycolytic pathway. As the length of activity increases further glucose, glycogen, and fats can be used enzymatically to produce APT in the presence of oxygen in the Krebs cycle. Thus there are four energy systems that can be identified: endogenous ATP, the ATP/CP system, the glucose/glycogen system and the oxidative glucose/fat system (78,85,116). Figure 2 shows these energy systems and provides some examples of activities associated with each (257).

Figure 2. Sources of Energy for Muscular Contraction

Energy Source	Metabolic Pathway	Activity Example
ATP	Phosphagen Splitting →	Lift Heavy Box
CP	Phosphagen Splitting →	Lift Several Heavy Boxes
Glycogen/glucose	Glycolytic Pathway →	Sprint
Glucose/Fat	Krebs Cycle →	Run/Walk

Modified from Reference Number [Vogel, 1985 #638]

It should be noted that these energy systems overlap and energy is seldom, if ever, supplied from only one system. However, because of the length of time that energy can be provided, each energy system is predominately associated with a particular type of muscle contraction. Endogenous ATP provides energy for very high-intensity, short-term muscle contractions like a maximal hand grip squeeze lasting about 3 seconds. CP rapidly replenishes ATP. The APT/CP system provides energy for high intensity muscle contractions lasting about 10 seconds like a short sprint. The glucose/glycogen system significantly overlaps the oxidative glucose/fat system since energy can be produced from glucose/glycogen in both systems. However, tasks lasting less than 1.5 minutes obtain energy predominately from the glucose/glycogen system. Activities lasting over 1.5 minutes derive energy from the glucose/fat system (78,85,116). Figure 1 links energy systems to fitness components identified in factor analytic studies. Note that there is actually a continuum: each energy system is used in approximate proportion to the force of the contraction and the length of time the contractions are carried out.

(2) Muscle Fiber Types

There are two basic muscle fiber types called fast twitch (FT) and slow twitch (ST). The FT fibers break down into at least two subtypes, and possibly more (240,241,246), but for the purposes of this paper only the two subgroups will be considered. The names of the FT and ST fibers come from the time it takes these fibers to reach peak tension when electrically stimulated. ST fibers reach peak tension in about 110 ms while FT fibers reach peak tension in about 40 ms. FT fibers contain an isoform (version) of enzyme called ATPase that can split ATP quickly. This faster enzyme allows for the faster contraction. ST fibers contain a slower isoform of ATPase resulting in the slower contraction.

A nerve and the muscle fibers it innervates are called a motor unit. Motor units are generally composed of either FT or ST fibers. All the muscle fibers attached to a nerve contact together when stimulated (all or none principle). A motor nerve innervating a FT motor unit typically contains 300-800 FT muscle fibers while a nerve innervating a ST motor unit typically has 10-180 ST muscle fibers. Because of the faster speed of contraction and greater number of muscle fibers FT motor units reach peak tension and generate more force than ST motor units (66,89,270).

FT and ST muscles are structurally and enzmyatically different. FT fibers contain a more highly developed structure called the sarcoplasmatic reticulum that allows for the faster contraction velocity by more rapidly releasing calcium to activate muscle contraction. FT fibers also contain large amounts of gylcolytic enzymes that make them well suited to producing energy non-oxidatitively from the glucose/glycogen system. ST fibers have many capillaries that provide for the more efficient transport of oxygen, fats, and glucose into the muscle from outside sources. ST fibers have a larger number of mitochondria that contain the Krebs cycle enzymes and more myoglobin for the storage of oxygen. ST fibers are thus well suited to producing energy oxidatively from the glycogen/fat system. As with energy systems, there is a continuum of characteristics (structure or enzyme profile) among different muscle fibers (89,93,206,211).

When muscles contract there is a selective recruitment of muscle fiber types that depends on the force or power required for the activity. During maximal contractions lasting a few seconds, both FT and ST fibers are recruited. With lower muscle forces that last a considerable period of time like long-distance running, ST fibers provide most of the muscle force. For events requiring short, high power output like short sprints, FT fibers are particularly recruited. It should be noted that both types of muscle fibers are used in most types of muscle contractions but one type is used predominately more than another type (65,86,211,270). Figure 1 shows the association between muscle fiber types and fitness components.

(3) Body Composition Factors

Body composition refers to the amount of various tissues in the body. Body composition can be quantified by a number of methods and the human body can be partitioned into compartments that include fat mass and fat-free mass (226). The fat-free mass compartment includes everything that is not fat and is composed primarily of muscle, bone, and mineral. Some techniques allow bone tissue to be partitioned out of fat-free mass so that 3 compartments (fat, bone, and lean tissue) can be distinguished. In this 3 compartment model, the lean tissue compartment has a larger proportion of muscle tissue mass since the bone is not included (176).

Muscle mass is highly correlated with absolute strength (120,121,184), power production (96), cardiorespiratory endurance (258), and the performance of many physical tasks (97,258). Individuals with more fat tend to have more difficultly performing certain tasks, especially those requiring weight bearing activity and cardiorespiratory endurance (52,258).

d. Consolidation of Factor Analytic and Physiological Studies

Physiological factors like energy systems, muscle fiber types, and body composition can be linked to fitness components identified by factor analysis. The fitness components that can be linked include strength, power, muscular endurance, cardiorespiratory endurance, body composition.

Strength can be defined as the ability of a muscle group to exert a maximal force in a single voluntary contraction. Maximal muscle contractions (i.e., 100% of maximum voluntary force) derive energy primarily from ATP. Both FT and ST muscle fibers are involved in maximal contractions but the FT muscle fibers provide most of the contractile force (270). The major determinate of strength appears to be the total cross-sectional area of muscle mass in the muscle group exerting the force. Individuals with more cross-sectional muscle mass are able to exert more force (120,121,182,184), and whole body fat-free mass is associated with greater strength in lifting tasks that involve a large proportion of the body muscle mass (258). The absolute amount of force generated is also dependent on muscle-bone architecture (74,174).

Muscular power is related to muscle strength but also involves a time component. Power is defined as force/time and thus muscular power is the ability of a muscle group to develop high force quickly. Power may be a subcomponent of strength in factor analytic studies because rapid, powerful movements depend on FT muscle fibers to a greater extent than other types of muscle contractions (211). There is a strong relationship between a high proportion of FT fibers and power production (17,27). Power can involve a single short contraction (peak power) or it can be sustained for a short period of time (sustained power). Peak power is well correlated with muscle strength in military populations (196). For peak power, energy will be derived primarily from ATP in the active muscles. For sustained power (less than about 10 sec), not only ATP but also CP in the active muscle will be used as an energy source (185). Examples of peak power events are quickly lifting a heavy weight or jumping up to reach the top of a wall. An example of a sustained power event is a short sprint. Like strength, power production depends on the total amount of muscle mass and muscle architecture.

Muscular endurance is the ability of a muscle group to perform short-term, high-intensity physical activity. Early in the muscular endurance activity (first few seconds) energy will be derived from ATP and CP but as the activity lengthens beyond about 10 seconds, energy will be derived from glycogen in the active muscles. Muscle glycogen can be mobilized rapidly to provide energy for the resynthesis of ATP through the glycolytic pathway. However, the byproducts of this rapid energy mobilization are associated with rapid fatigue (30,200,264). The muscle is working at a high percentage of its maximal capacity (50% to 90%), and probably recruits both FT and ST fibers depending on the length of the contraction. Individuals with more muscle mass are able to continue these high intensity muscle contractions (at an absolute exercise intensity) for a longer period of time, presumably because they have more muscle tissue over which to spread the load of the repeated contractions so that when some motor units become fatigued other motor units can continue to contract.

Cardiorespiratory endurance is the ability to sustain long-term, low-power physical activity. There is strong physiological evidence for this component of physical fitness. Energy for the low intensity, long term muscle contractions of this sort is primarily derived from the glucose/fat system, and the predominate muscle fiber used in these types of contractions are the ST (185,211). Oxygen is used to produce this energy and the amount of oxygen can be directly linked to the amount of energy produced (185). Morphologially, cardiorespiratory endurance depends on the functioning of the circulatory and respiratory systems. The ability of lungs to deliver oxygen to the blood, the ability of blood to deliver oxygen to the active muscles, and the ability of the muscles to take up and use this oxygen to produce energy from glucose and fat are all linked to the ability to perform long-term physical activity (211).

Coordination and balance involve muscular contraction and will recruit energy systems and muscle fiber types in proportion to the intensity of the contraction and the duration of the activity. However, neuromuscular control is a primary characteristic of tasks requiring coordination or balance. For example, consider an obstacle course. The ability to quickly move over, under and around obstacles requires the coordinated action (neuromuscular control) of a number of muscle groups. The movement is "agile" in proportion to the speed of completion and to the extent that unnecessary movements are avoided (economical movement). An activity requiring coordination on an obstacle course may recruit different energy systems at different times for different types of activities. An individual may be required to run between obstacles (cardiorespiratory endurance), jump up and pull himself/herself over a wall (power and muscular strength), and rapidly traverse a series of logs (muscular endurance). In the case of balance, neuromuscular control is used to inhibit unwanted muscular contractions to obtain a required state of static (little or no movement) or dynamic (movement in a specific direction) equilibrium. For example, consider standing on a narrow board. The individual is "balanced" in proportion to his or her ability to sustain static muscular contractions that result in little or no movement on the board.

As noted earlier, many fitness components (muscular strength, muscle power, muscular endurance, cardiorespiratory endurance) actually exist on a physiological continuum. This continuum is characterized by the intensity of the muscular contraction, the relative proportions of FT and ST fibers used, the predominate energy source, and how soon fatigue ensues. The continuum is shown in Figure 1. The Factor Analytic Approach suggests discrete fitness components while the physiology variables suggest a continuum. However, the results are complementary because at widely separated points on the continuum major differences in physiological factors do exist.

e. Components of Physical Fitness: Consolidated Definition

Table 3 shows the relationship among terms used to describe the components of physical fitness derived from the review of the Factor Analytic Approach, Ability Requirements Approach, and the Physiological Approach. The generic terms for the fitness components are based on common concepts and associations in each approach. Generic terms are more closely linked to the energy systems involved and the terms are more easily understood and generally accepted (37,45,204).

While it is possible to link physiological energy systems with some components of fitness, this linkage for other components (coordination, balance flexibility) depends on the activity and the length of time the activity is performed. For example, the ATP-CP-glycogen system would be primarily involved in a coordination (agility) task that requires an individual to move quickly around a series of obstacles and takes 30 seconds to complete at a maximal effort. Longer coordination events taking several minutes to complete at a submaximal effort might recruit glycogen/glucose/free fatty acids. Flexibility movements through a range of joint motion that takes 1 to 2 seconds to complete would require energy only from ATP. A flexibility movement like a static stretch that is held for several minutes would recruit other energy systems.

Table 3. Consolidated Definition of Physical Fitness Components

Generic Term	Factor Analytic Approach	Human Ability Approach	Physical Measure	Energy System
Muscular Strength	Static Strength Power	Static Strength Explosive Strength	Maximal Force Maximal Power	ATP[a]
Muscular Endurance	Muscular Endurance Trunk Endurance	Dynamic Strength Trunk Strength	Short-term sustained force or average power	ATP-CP[b] Glycogen/Glucose
Cardiorespiratory Endurance	Cardiorespiratory Endurance	Stamina	Speed/distance or long-term sustained force/power	Glycogen/Glucose/ FFA[c]
Coordination	Coordination	Gross Body Coordination	Speed/distance (deviation from desired movement)	[d]
Balance	Balance	Gross Body Equilibrium	Distance (deviation from desired posture)	[d]
Flexibility	Flexibility	Extent Flexibility Dynamic Flexibility	Distance (range of motion)	[d]
Body Composition	Body Weight, Body Fat, Muscle Mass		Mass (body tissue amount)	(related to tissue type)

[a]ATP= adenosine triphosphate
[b]CP=creatine phosphate
[c]FFA=free fatty acids; minor amounts of protein also used
[d]Varies depending on power output and length of time of movement

4. PHYSICAL FITNESS TESTS

For the purposes of this review we considered relatively simple tests of physical fitness. Simple tests were those quickly and easily understood by the individual being tested and that could be administered in the MEPS station or by recruiters with minimal training and equipment. Although we tried to minimize equipment because of the initial expenses and maintenance costs, a test requiring equipment was considered if it substantially improved the reliability or validity of a test.

In addition to procedural and equipment considerations we considered the reliability and physiological validity of the test. A test is reliable if it produces a similar score over a series of tests. For example, if an individual performs a test and achieves a particular score on one occasion they should achieve a similar score on a second occasion. The correlation coefficient between scores on the 2 tests administrations demonstrates the magnitude of the reliability. A test has physiological validity if it has a high correlation with a physiological test related to that fitness component. For example, any simple measure of cardiorespiratory endurance should have a high correlation with VO_2max because VO_2max is the physiological test that measures cardiorespiratory endurance (20,194). Many fitness components do not have accepted criteria and so physiological validity cannot be established.

We did not consider tests of flexibility, balance, or coordination. These components of physical fitness have not been well characterized in factor analytic studies, there are few standard tests available for some of these fitness components (balance or coordination), and there is little data on the reliability of balance or coordination tests. Further, these components have not been identified as limiting military task performance nor have they been related to injuries or attrition from military service.

a. Tests of Muscular Strength

Strength can be tested either statically or dynamically as the maximum force or power that an individual exerts. On some tests, surrogate measures of force or power are measured (e.g., maximal distance of projecting an object). In isometric testing, the individual exerts as much force as he or she can against a fairly immovable object. Spring loaded tensiometers were used to measure the force early in objective strength testing (42), but load cells later replaced this technology (160). Dynamic strength tests can be separated into three broad categories: a) tests involving isoinertial maximal tests, b) tests involved in projecting objects, and c) tests involving projecting the body weight. Isoinertial maximum tests involve determining a one-repetition max (1RM) in which the weight the individual lifts is progressively increased in a systematic manner until the maximum weight that the individual can lift is determined. Tests involving projection of objects generally entail throwing or "putting" objects as far as possible (e.g., softball throws for distance, shot put). Tests of body projection involve propelling the body forward or upward as far as possible in a single maximal effort (e.g., vertical jump, broad jump).

Isometric testing requires the use of some equipment but this equipment can be easily acquired, is durable, and is relatively easy to maintain. It takes little time to train test administrators on these simple tests. Individuals can easily understand the test requirements and can be tested quickly. Isoinertial tests are also relatively inexpensive and equipment maintenance is low. The time required to train administrators is short but it can be somewhat more time consuming to find an individual's 1RM because individuals must lift a series of heavier weights. Tests of projecting objects can be especially inexpensive but additional administrators are required (one to monitor the individual and one to watch where the object falls), and some space is required over which the object can be thrown.

Conceptually, tests of body projection (e.g., vertical jump, standing broad jump) would seem like the cheapest and easiest to administer. However, they also tend to be less "standardized" than other methods of measuring strength because of variations in body weight. Body weights can vary considerably among individuals. Projecting a larger body mass takes more muscular power and thus the larger body mass tends to reduce the distance the body can be projected. Body weight is moderately correlated with strength (26,36,43,99,119,188), indicating that heavier individuals tend to have more strength. However, this relationship does not standardize strength to body weight because strength is not exactly proportional to body weight. One vertical jump method considers the body weight and height in an equation that provides a measure of absolute peak and average power (98).

There is no accepted single physiological measure for muscular strength so it is not possible to examine physiological validity. However, the construct validity of muscle strength has been more adequately established than any other component of physical fitness as discussed above. A number of studies have reported on the reliability of various measures of muscular strength and Table 4 shows some of these studies. This is not a comprehensive list but merely a sampling of the literature. Reliability values are relatively high. For isometric strength tests, reliability coefficients range from 0.75 to 0.98; for isoinertial (dynamic) tests, coefficients range from 0.88 to 0.99; for tests involving object projection (baseball throw, softball throw, medicine ball put, shot put), coefficients range from 0.70 to 0.97; for tests involving body projection (vertical jump, broad jump, bar snap, high jump, rope climb) reliability ranges from 0.62 to 0.98.

Table 4. Reliability of Tests of Muscular Strength

Test	Study (Reference Number)	Subjects	Reliability Coefficient
Isometric Hand Grip	69 271 225 225 212 214 119	201 Naval recruits 116 California Highway Patrolmen 350 male Naval recruits 269 female Naval recruits 12 laboratory personnel 51 athletes 406 boys in physical education classes	0.91 0.75 0.93 0.90 0.98 0.81-0.82 0.95
Isometric Plantar Flexion	212	12 laboratory personnel	0.83
Isometric 38-cm Upright Pull	160	270 Soldiers	0.97
Isometric Wrist Flexion	164 42	50 male college students 64 college students	0.80[a] 0.93
Isometric Elbow Flexion	35 42 155	36 male college students 64 college students 352 male infantry soldiers	0.94[a] 0.96 0.98
Isometric Knee Extension	155 42	352 male infantry soldiers 64 college students	0.98 0.94
Isometric Squat	25	14 athletic men	0.97
Dynamic Bench Press	271 232 118	116 California Highway Patrolmen 14 young men 24 male university students and staff	0.88 0.99 0.94
Dynamic Squat	118	24 male university students and staff	0.94
Softball Throw	69	201 Naval recruits	0.93
Baseball Throw	43	100 college men	0.91
Medicine Ball Put (9 lbs) standing	69	201 Naval recruits	0.70
Medicine Ball Put (9 lbs) sitting	69	201 Naval recruits	0.73
Shot Put, 4 lbs	214	51 athletes	0.90
Shot Put, 6 lbs	119	406 boys in physical education classes	0.97
Shot Put, 12 lbs	119	406 boys in physical education classes	0.97
Vertical Jump/Sargent Jump	69 214 187	201 Naval recruits 51 athletes Fourth to twelfth grade boys	0.90 0.80-0.82 0.98
Standing Broad Jump	69 119 18	201 Naval Recruits 406 boys in physical education classes 95 boys (7-11 yrs), summer sports program	0.90 0.96 0.76
Bar Snap	228	103 male college freshmen	0.92
Running High Jump	119	406 boys in physical education classes	0.96
Rope Climb (6 sec)	69	201 Naval recruits	0.80

[a]Three trial reliability

b. Tests of Muscular Endurance

Tests of muscular endurance involve repeated high intensity muscular contractions that are continued for relatively short periods of time (less than about 1.5 minutes). Muscular endurance tests can involve static or dynamic contractions and absolute or relative loads. There are at least four possible types of tests described in the literature. One type involves repeatedly moving a fixed load or a fixed proportion of one's strength as many times as possible in a set time or until fatigue. An example is performing as many contractions as possible in 30 sec on a bench press with a load of 37 lbs or a load of 30% of one's maximal strength. Another type of muscular endurance test involves statically holding a fixed load or a fixed proportion of one's maximal strength. An example is holding 30 lbs of force on a hand grip or 50% of one's maximal strength until fatigue ensues. A third type of muscular endurance test involves repeatedly moving the body or a portion of the body in a specific period of time or until fatigue ensues. Examples are PUs or SUs. A fourth and final type of muscular endurance test involves statically holding the body or a portion of the body in a fixed position until fatigue ensues. Examples include the flexed arm hang or holding a half SU.

Absolute muscular endurance requires individuals to hold (static) or move (dynamic) a specific force (weight or resistance). An example of an absolute endurance test is asking an individual to flex and extend his or her elbow with a 20 lb weight to a cadence. The measure is the amount of time the individual is able to maintain the cadence. Relative muscular endurance requires individuals to hold or move a certain proportion of their maximal strength. An example of a relative muscular endurance test would be asking the individual to flex and extend his or her elbow to a cadence with a weight that is 30% of his or her maximal strength. The measure would remain the same as in the absolute endurance test. Absolute muscular endurance tests more closely approximate situations experienced in the real world. This is because objects of fixed mass are typically those that have to be handled, held, lifted, carried, or otherwise moved. Loads are not set dependent on a mass relative an individual's maximal capacity.

For tests involving repeatedly lifting loads or statically holding fixed loads administrators can be quickly taught the well-standardized tests. Individuals can be tested rapidly since the time is set or fatigue rapidly ensues. Tests of this type generally require some minimal equipment which, in the simplest case, is only a set of free weights. Muscular endurance tests dependent on the body weight or portions of the body weight (e.g., PUs, SUs) share the potential shortcomings discussed earlier in the section on muscle strength with regard to differences in body weights. However, tests of this type require no equipment, little time to train administrators, and can be administered very quickly.

There is no single accepted physiological measure for muscular endurance so no physiological validity can be established. Construct validity has been well established and has been discussed above. Table 5 shows some tests of muscular endurance and the reported reliability. This is not a comprehensive list and only shows the variety of tests available to measure muscular endurance. The two tests shown that involve repeatedly moving an external weight (bench press, rowing) have reliabilities of 0.90 to 0.80. The two tests involving static hand grip or leg press have reliability coefficients of 0.68 to 0.60. Tests that involve moving the body or a portion of it (pull-ups, PUs, dips, leg lifts, deep knee bends, squat thrusts, anaerobic shuttle run, sprints) have reliabilities ranging from 0.57 to 0.97. Tests that involve holding the body in one position (flexed arm hang, hold half-sit up, hold half PU) have reliabilities of 0.74 to 0.85.

Table 5. Reliability of Tests of Muscular Endurance

Test	Study (Reference Number)	Subjects	Reliability Coefficient
Bench Press Repetitions (37 lbs, max reps in 20 sec)	69	201 Naval recruits	0.90
Rowing Repetitions (37 lbs, max reps in 20 sec)	69	201 Naval recruits	0.80
Hand Grip Endurance (hold ½ maximal strength to fatigue)	34	56 male college students	0.60
Leg Press Endurance (hold 300 lbs to fatigue)	63	34 aviation students	0.68
Pull-up (to fatigue)	46 141 69 187 228 18	14 college physical education majors 150 tenth grade males 201 Naval recruits Adults 103 male college freshmen 95 boys (7-11 yrs), summer sports program	0.89 0.89 0.88 0.91 0.95 0.86
Pull-ups (20 sec)	69	201 Naval recruits	0.95
Modified Pull-ups (legs on floor)	71	147 high school girls	0.82
PUs (15 sec)	69	201 Naval recruits	0.76
PUs (to fatigue)	69	201 Naval recruits	0.88
Sit-up	141 69 71	150 Tenth grade males 201 Naval recruits 139 high school girls	0.57 0.72 0.61
Dips (to fatigue)	228 69 18	103 male college freshmen 201 Naval recruits 95 boys (7-11 yrs), summer sports program	0.92 0.91 0.77
Dips (10 sec)	69	201 Naval recruits	0.92
Leg Lifts	214 119	51 Athletes 406 boys in physical education classes	0.67 0.95
Leg Lifts (20 sec)	69	201 Naval recruits	0.84
Deep Knee Bends	69	201 Naval recruits	0.85
Squat Thrust	69 71 228 187	201 Naval recruits 142 high school girls 103 male college freshmen Adults	0.70 0.74 0.87 0.72
Anaerobic shuttle Run	69	201 Naval recruits	0.85
30 yd dash	214	51 Athletes	0.88
50 yd dash	69	201 Naval recruits	0.86
60 yd dash	119	406 boys in physical education classes	0.97
Flexed Arm Hang (chin touch bar)	46	14 college physical education majors	0.74
Flexed Arm Hang (elbows to 90°)	46	14 college physical education majors	0.83
Flexed Arm Hang (eyebrows at bar)	69	201 Naval recruits	0.77
Hold Half Sit	69	201 Naval Recruits	0.88
Hold Half PU	69	201 Naval recruits	0.85

c. Tests of Cardiorespiratory Endurance

In the literature there are three types of tests of cardiorespiratory endurance that meet the general criteria described above for an acceptable test of this fitness component. These include a) running tests for time over fixed distances, b) running tests at fixed times completing as much distance as possible, and c) aerobic shuttle run tests. An innovative step test was also considered for the present purposes and is described later. Tests of cardiorespiratory endurance that used heart rate to predict VO_2max (12,132,142) were not considered here because of the relative complexity of the procedures, amount of equipment required, and the training needed for accurate measurement. Further, it became apparent early in the review that these heart rate methods did not appear to be more valid, and in some cases they were less valid (6,277) than more simple measures that used time, distance, or speed. There are a number of assumptions in the use of heart rate to predict VO_2max and violation of one or more of these assumptions may account for the lower validity. These assumptions include 1) the assumed linearity of the heart rate-VO_2max relationship, 2) assumed relationship between age and maximal heart rate, 3) assumed constant mechanical efficiency, and 4) day-to-day variations in heart rate (55,185).

(1) Physiological Validity of Cardiorespiratory Endurance Tests

For the cardiorespiratory endurance component of fitness a well established physiological measure exists. This measure is the maximum rate at which oxygen is used by the body (VO_2max) during physical activity. VO_2max is the highest rate at which oxygen can be taken up and used by the body during physical activity (20). The faster the rate of oxygen usage, the faster the rate of energy production to fuel longer-term physical activity. Oxygen used by the body is directly linked to oxidative energy production. One liter of oxygen taken up by the body is the energy equivalent of 4.85 kilocalories produced from fats, carbohydrates, and protein. Thus, VO_2max is a measure of cardiorespiratory endurance because it is a direct measure of the maximal rate at which energy can be supplied to fuel longer-term physical activity (156).

Table 6 shows studies that have examined the relationship between VO_2max and times achieved on maximal-effort runs of varying distances. The studies are arranged in the table by the distance of the run test with the exception of 5 studies on the bottom, which involved multiple distances. Distances range from 0.1 miles to 26.2 miles. Where adequate descriptions of subject samples were provided (64,80,94,101,168,183,191,192,203,215,224, 237) participants tended to be physically active, although in a few cases untrained individuals served as subjects (231,266). Run tests using untrained individuals (231,266) had lower correlations than studies using physically active subjects. The average ages of individuals in these studies were within those that might be expected among basic trainees (17-35 years of age) with the exception of 4 studies (94,168,203,224) that examined middle-aged subjects. Body weights were similar to those of recruits (148,150,157). All studies in Table 6 validated the run against a VO_2max test on a treadmill. Most studies used a graded uphill running protocol (32,64,94,101,183,191,192,203,213,224,277), but some used a graded uphill walking protocol (168,215). Two studies used a single stage test (231,266) in which subjects ran at 7 miles/hr and 8.6% grade and this could have underestimated VO_2max in the most fit subjects. In one case (237) the VO_2max protocol was not specified. The negative correlations indicate that as VO_2max increases, run times decrease.

Longer running distances or longer running times result in more use of aerobic energy sources (4,116) and because of this higher correlations between VO_2max and running performance were expected at longer distances. Such a trend cannot be seen across different studies in Table 6 but the trend can be seen across studies examining single distances. This is likely due to methodological differences between separate studies. Single studies examining multiple distances use the same methods for all distances making it easier to see the relationship between VO_2max and distance. Examining studies involving multiple distances in Table 6 suggests that run distances as short as 1 mile provide acceptable physiological validity but distances of 2 miles or more appear optimal. Distances below ½ mile generally have lower validity.

An alternative to a distance run is a timed run. In this type of test individuals complete as much distance as possible in a set time. The relationship between VO_2max and distances achieved on 12-min runs are shown in Table 7. In studies that provide adequate descriptions of tested subjects, individuals tended to be physically active (44,88,177,191), although one study used a mixed group of physically active and sedentary subjects (139). Ages and weights of individuals tested tend to be very similar to those of recruits (148,150,157). Aerobic capacity (VO_2max) of the individuals tested tends to be higher than those of recruits (205,230) in all but 2 studies (139,274). Most studies used a graded uphill running test to determine VO_2max (32,44,88,190,191), but 2 used an uphill walking protocol (139,177), and 1 study used a protocol increasing exercise intensity by speed alone (274). These fixed-time studies generally show higher correlations between running distances and VO_2max than fixed-distance studies. In fixed-time tests, the time left to complete the run can be called out and this may provide more motivation. However, fixed-time tests are more difficult to administer because of the necessity to calculate individual distance.

The aerobic shuttle run involves running back and forth between 2 markers placed 20 m apart. Exercise intensity (pace) is determined by a metronome that provides an auditory signal. When the metronome sounds the participant must be at one of the 2, 20 meter markers. The goal of the test is to complete as many 20-m circuits as possible. The test is terminated when the participant can no longer maintain the metronome pace. Participants start running at either 8.0 km/h (5 miles/h) or 8.5 km/h (5.3 miles/h). Speed is increased 0.5 km/h (0.3 miles/h) every minute. The original test (171) had 2 minute stages but participants often became bored with the test and stopped before reaching their maximal capacity (170). Most studies have used 1 minute stages which results in less test time and equivalent physiological validity.

Table 6. Studies Examining Relationships between VO_2max and Running Tests at Various Distances

Study (Reference Number)	Distance (miles)	Subjects	Age (yr)[a]	Weight (kg)[a]	VO_2max (ml/kg*min)[a]	Physiological Validity[ab]
101	1.2	9 men in the British Royal Air Force	31±2	70±4	64±3	-0.83
80	1.5	21 female college joggers	20±2	57±8	46±6	-0.92
271	1.5	106 California Highway Patrolmen	~31[c]	~83[c]	39.9[c]	-0.68
277	1.5	38 women	33±3	64±8	41±7	-0.79
191	1.5	32 male college physical education majors	20±0	74±3	60±6	0.87[d]
224	2.0	24 moderately well trained men	40±6	80±11	49±6	-0.86
192	2.0	44 men, 17 women, active duty Army	31±7M 28±4W	78±9M 61±8W	50±8M 42±6W	-0.91M -0.90W
168	2.0	70 male US Army War College students	43±2	80±8	43±5	-0.78
215	3.0	14 male Marines	[e]	[e]	[e]	-0.65
213	3.1	36 men, 38 women	19-36	71±8M 57±9W	59±7M 47±6W	-0.76M -0.83W
203	6.2	9 endurance trained Men	35±6	74±6	59±10	-0.95
173	18.6	11 marathoners	32±6	68±5	66±2	-0.71
94	26.2	50 marathoners	36±8	70±6	65±6	-0.63
183	26.2	18 male and 10 female marathoners	34±7M 30±7W	68±9M 59±8W	61±10M 52±6W	-0.88M -0.63W
237[h]	26.2	35 marathon runners	30	67	66	0.78[f]
224	 0.1 0.3 0.5 1.0 2.0	11 college students, moderately well trained	20±1	72±9	57±4	 -0.05 -0.31 -0.67 -0.79 -0.85
266	 0.25 1.0 2.0 3.0	30 untrained college men	21±2	74±12	53±6	 -0.22 -0.29 -0.47 -0.43
32	 0.1 0.3 1.0	44 college men	22±3	78±11	53±6	 -0.52 -0.78 -0.74
231	 0.1 0.3 0.5 1.0 2.0 3.0	30 untrained college men	23±3	76±13	54±6	 -0.08 -0.29 -0.35 -0.43 -0.76 -0.82
64	 2.0 6.0 9.3 12.0 26.2[g]	18 experienced male distance runners	28±9	70±8	62±8	 0.83[f] 0.86[f] 0.89[f] 0.91[f] 0.91[f]

[a]M=Men; W=Women
[b]Correlation between VO_2max and run time
[c]Values are approximate since not all subjects completed both tests
[d]Not a correlation between VO_2max and run performance but rather between directly measured VO_2max and VO_2max estimated from a simple linear regression
[e]Data not reported in study
[f]Correlation is between VO_2max and running speed rather than run time
[g]Only 13 individuals ran the 26.2 mile distance
[h]Age, weight and VO_2max values were calculated as the weighted average of 3 groups in the article

Table 7. Studies Examining Relationships between VO_2max and 12-min Timed Running Tests

Study (Reference Number)	Subjects	Age (yr)[a]	Weight (kg)[a]	VO_2max (ml/kg*min)[a]	Physiological Validity[a]
44	115 male Air Force officers	22	76	[c]	0.90
274	25 male laboratory workers	30±8	78±14	44±9	0.94
139	36 college women: 12 athletes, 10 physical education majors, 14 sedentary	20±1	59±7	39±5	0.67
32	44 college men	22±3	78±11	53±6	0.90
177	26 women, varsity athletes	20±2	62±9	41±4	0.70
190	15 male and 15 female college students	26±5M 26±5W	74±7M 60±7W	62±7M 51±6W	0.84M[b] 0.82W[b]
191	32 male college physical education majors	20±0	74±3	60±6	0.87[b]
88	22 men involved in endurance sports	22±2	72±9	60±8	0.86

[a]M=Men; W=Women
[b]Not correlation between VO_2max and run performance but rather between directly measured VO_2max and VO_2max estimated from a simple linear regression
[c]Data not reported in article

Table 8 shows studies that have examined the physiological validity of the aerobic shuttle run. The studies are arranged by the year in which the studies were conducted with the earlier studies listed first. The average age in most of these investigations (81,88,191,203,244) was similar to that of basic trainees (17-35 years of age). An exception was one study (203) that had older participants (26 to 47 years of age). Body weights were similar to those of recruits (148,150,157) but the cardiorespiratory endurance levels tended to be higher than those of recruits (205,230). Where adequate descriptions of subject samples were provided (81,88,191,203,244), participants tended to be physically active. Most studies validated the run against a VO_2max test on a treadmill using a graded uphill running protocol (81,88,170,191,203,213,244). An exception was the original aerobic shuttle run study (171) that used a retroextrapolation procedure. Retroextrapolation involved collecting a series of timed expired gas samples in Douglas bags immediately after the run. VO_2max was determined by extrapolating the VO_2-time curve back to the end time of the run exercise (172).

Table 8. Studies Examining Relationships between VO_2max and the Aerobic Shuttle Run

Study (Reference Number)	Subjects	Age (yr)[a]	Weight (kg)[a]	VO_2max (ml/kg*min)[a]	Physiological Validity[a]
171	59 men and 32 women	25±6M 27±9W	71±10M 57±9W	52±8M 39±8W	0.84[c]
203	9 endurance trained men	35±6	74±6	59±10	0.95
213	36 men, 38 women	19-36	71±8M 57±9W	59±7M 47±6W	0.83M 0.93W
170	53 men, 24 women	31±8W 31±7M	72±10M 53±7W	49±10	0.90[c]
88	22 men involved in endurance sports	22±2	72±9	60±8	0.92
191	32 male college active college physical education majors	20±0	74±3	60±6	0.82[b]
81	10 runners and 10 squash players	22±3	71±8	61±3	0.67
244	60 male and 60 female athletes	25±5M 25±5W	77±11M 64±9W	55±8M 47±6W	0.77M 0.66W

[a]M=Men; W=Women
[b]Not correlation between VO_2max and run performance but rather between directly measured VO_2max and VO_2max estimated from a simple linear regression
[c]Did not provide a separate correlation for men and women

(2) Reliability of Cardiorespiratory Fitness Tests

We found surprisingly few studies that had examined the reliability of tests of cardiorespiratory fitness. This may be because of the difficulty of having volunteers perform such a physically demanding test on multiple occasions. Studies on the reliability of cardiorespiratory endurance tests are shown in Table 9. Reliability coefficients for times to complete distances of 0.3 to 2 miles range from 0.82 to 0.92. Reliability coefficients for timed runs of 5 to 12 minutes range from 0.78 to 0.94. Reliability coefficients for the 20-m shuttle run are 0.87 and 0.98.

Table 9. Reliability of Tests of Cardiorespiratory Endurance Tests

Test	Study	Subjects	Reliability Coefficient
600-yd (0.3 mile) run	60	60 ninth and tenth grade girls	0.87
1-mile run	163	Trained first grade girls Trained third grade boys	0.82 0.92
2-mile run	224	10 well trained middle-aged men	0.91
5-min run	60	100 ninth and tenth grade girls	0.79
8-min run	60	45 ninth grade girls	0.87
9-min run	60	43 ninth grade girls	0.84
10-min run	60	123 ninth and tenth grade girls	0.90
11-min run	60	45 ninth grade girls	0.88
12-min run	139	36 college women: 12 athletes, 10 physical education majors, 14 sedentary	0.78
	177	26 women, varsity athletes	0.87
	178	80 high school boys	0.92
	59	154 ninth grade boys	0.94
	60	145 ninth and tenth grade girls	0.92
20-m Shuttle Run	81	10 runners and 10 squash players	0.87
	171	59 men and 32 women	0.98

(3) Innovative Step Test

In considering tests that could be easily administered, we conceptualized a step test that could be conducted in a small space. This test involves a repetitive bench stepping task that uses a standard stair height of 7 inches. The test could be conducted either in stages (like the aerobic shuttle run) or for a fixed time. The major advantage of a step test is that it can be administered in a relatively confined space and would thus be more appropriate for the MEPS or recruiter station.

The bench stepping in stages would start at a slow cadence (e.g., 30 steps/min) set to a metronome. Every minute the metronome cadence would be increase by a set rate. The individual taking the test would continue until he/she was unable to keep up with the cadence.

The bench step for a fixed time would require individuals to complete as many steps as possible in a fixed time (e.g., 10 minutes). A plate counter at the top of the step would count the number of steps completed in the set time. It may also be possible to use a pedometer attached to the individual's belt or waist band to count the number of ascents and descents, although this might introduce some error if the pedometer did not catch every step.

It would be necessary for pilot studies to be conducted to determine adequate times and/or paces. Validity could be examined by directly measuring a VO_2max using both versions of the graded step test. For the stair step in stages, oxygen uptake at each stage of the test could be determined by direct measurement. A table would be developed showing the VO_2 value at each stage of the test. For the maximum steps in a fixed time a sampling of steps could be obtained and a curve developed by extrapolating steps/10 min to oxygen uptake. Test-retest reliability can be established by having individuals repeat the step test at least twice.

d. Body Composition

Measures that meet the standards for an acceptable assessment of body composition by recruiters or MEPS personnel are simple anthropometric measures such as body circumferences, girths, diameters, and/or skinfolds. All of these measures require minimal and easily maintained equipment, are quickly learned by administrators, are easy to administer, and are passive measures for the individuals being tested since they require no action on the testee's part.

Physiological validity of various anthropometric measurements have been established by relating the anthropometric measures to body composition determined from densitometry (underwater weighing, air plethysmography), dual energy X-ray absorptiometry (DEXA), and other direct measures (226). The largest amount of literature is on underwater weighing. Underwater weighing uses whole body density and assumptions about the density of fat and fat-free mass to determine body composition. The density of any object is calculated as mass/volume. A body that is submerged in water is buoyed up by a force equal to the weight of the water that is displaced (Archimedes's principle). Thus, the body volume is equal to the individual's weight in water. An individual's density can be calculated as the weight in air minus the weight in water corrected for the density of water (84). Corrections must also be made for the air in the lungs and air in the gastrointestinal track. Air in the lungs can vary considerably among individuals and must be measured directly on land. Gases in the intestinal track are small and can be estimated (33). Based on data from animal carcasses and human cadavers the density of fat can be assumed to be 0.90 g/mL and that of fat-free mass, 1.10 g/mL (31). These numbers have been shown to vary somewhat based on age, race, and degree of obesity (84). The classic formula for conversion of body density (D_b) into fat is that of Siri (235,236) which is:

$$\%Fat = (4.95/D_b - 4.50) * 100\%$$

Another densitometry measure (whole body plethysmography) uses the relationship between pressure and volume (Boyle's Law) to determine whole body volume and hence calculate body density (57). Another method for measuring body composition is dual X-ray absorptiometry (DEXA). This method determines tissue amounts and densities from the attenuation of two low energy beams of X-ray radiation projected into the body (5).

Table 10 shows the association of body fat determined from underwater weighing with various anthropometric measures. This is by no means a comprehensive list of studies in this area but Table 10 shows the wide variety of anthropometric measures that can be used to predict body fat.

Table 10. Studies Examining Relationship Between Body Fat (Determined from Underwater Weighing) and Various Anthropometric Measures

Study (Reference Number)	Subjects	Age (yr)[a]	Anthropometric Measures	Body Fat (%) from Underwater Weighing	Physiological Validity (Correlation Coefficient)
275	Women from college community	20±2	Subomphalion skinfold Lower rib skinfold Triceps skinfold Suprailiac skinfold	28.7±4.1	0.68
140	64 college women	19 to 23	Triceps skinfold Buttock girth Upper arm girth Scapula skinfold	21.5±5.7	0.70
268	133 college men	22±3	Abdominal skinfold Bi-iliac diameter Neck circumference Chest circumference Abdominal circumference	14.6±5.5	0.83
269	128 college women	21±4	Scapula skinfold Knee diameter Neck circumference Minimal abdominal circumference Maximal abdominal circumference	25.7±4.5	0.76
138	53 college Men	19±2	Tricep skinfold Scapula skinfold Abdominal circumference Forearm circumference	15.3±5.7	0.89
	69 college women	20±2	Scapula skinfold Iliac skinfold Elbow diameter Thigh diameter	25.6±6.4	0.84
209	60 middle aged women	45±6	Axilla skinfold Superiliac skinfold' Thigh skinfold Chest girth Waist girth Cup size	29.8±6.7	0.91
	83 healthy female college students	20±1	Superiliac skinfold Thigh skinfold Chest girth Waist girth Chest diameter Knee diameter	24.8±6.4	0.84
208	84 healthy middle aged men	45±5	Chest skinfold Axilla skinfold Abdominal girth Gluteal girth Arm girth	24.7±5.9	0.84
	95 health male college students	25±6	Tricep skinfold Abdominal skinfold Waist girth Calf girth Ankle girth Wrist girth	13.4±6.0	0.88

			Biacromial diameter Bitrochanteric diameter		
128	308 men	33±11	ΣChest, abdominal, thigh skinfolds Age Forearm circumference	17.7±8.0	0.92
129	249 women	31±11	ΣTricep, thigh, superiliac skinfolds Age Gluteal circumference	24.1±7.2	0.85

Table 11 show the anthropometric measures used by the military services to predict body fat. The physiological validity is similar to other studies in Table 10. Friedl and Vogel (72) performed a cross-validation of the male equations for the Army, Navy and Marine Corps using DEXA. They found that all three equations had similar validity (r=0.80 to 0.82) and standard errors of estimate (3.1 to 3.3% body fat).

Table 11. Military Services Estimates of Body Composition

Study (Reference Number)	Subjects	Age (yr)[a]	Anthropometric Measures	Body Fat (%) from Underwater Weighing	Physiological Validity (Correlation Coefficient)	Standard Error of Estimate (% body fat)
272 273	297 male Marines	29±8	Abdominal circumference Neck circumference	16.5±6.2	0.81	3.7
	181 female Marines	23±6	Biceps circumference Forearm circumference Neck circumference Abdominal circumference Thigh circumference	23.1±5.9	0.73	4.1
259	1126 male Soldiers	30±9	Height Abdominal circumference Neck circumference	20.6±7.0	0.82	4.0
	266 female Soldiers	24±5	Hip circumference Forearm circumference Neck circumference Wrist circumference Weight	28.0±6.1	0.82	3.6
112 113	602 male Sailors	32±7	Abdominal circumference Neck circumference Height	21.6±8.1	0.90	3.5
	214 female Sailors	27±5	Abdominal circumference Hip circumference Neck circumference Height	27.0±6.9	0.85	3.7
73 111	197 Air Force men	37	Flexed biceps circumference Height[b]	20.3 (range 5.9 to 35.6)	0.84	3.0[b]
		[c]	Forearm circumference Height	[c]	0.84	3.0

[a]Equation predicts fat-free mass rather than body fat
[b]Standard error is for fat-free mass in kg
[c]Body fat unknown; obtained from secondary source (111)

Circumference measures are somewhat more reliable than skinfold measures and it takes less time to train individuals on the proper circumference techniques (111,195). Thus, if body fat is to be estimated it is recommended that a circumferential method be selected. The most appropriate estimate of body composition would involve the Army equations since they were developed on an Army sample.

5. CRITERIA FOR SELECTION OF PHYSICAL FITNESS TESTS

Our approach to developing a physical fitness test was to determine criteria that are important from a military standpoint and examine the relationship between these criteria and various measures of physical fitness. In this way criterion-related validity could be established. Criteria that have been defined as important in the literature include job performance, injuries, and attrition from service. Each of these is reviewed below.

a. Physical Fitness and Job Performance

The Equal Employment Opportunity Commission (EEOC) has published Uniform Guidelines on Employee Selection Procedures which is abstracted in Appendix B. The guidelines define acceptable criteria for a pre-employment selection test. A large body of literature has developed regarding the association between physical fitness tests and job performance since employers have attempted to comply with the EEOC guidelines. Many occupational tasks are physically demanding and studies have shown that specific physical fitness tests are related to performance on these tasks. The practical goal of much of the research has been to develop a test battery that identifies whether or not individuals have the physical capability to perform a particular job and thus be hired for that job. This section on job performance will review civilian and military studies that have examined the relationship between physical fitness tests and job task performance. Because studies differ substantially in methods and tests, each one is reviewed individually below.

(1) Civilian Studies

(a) Steelworkers. Arnold et al. (10) examined the relationship between anthropometric/physical fitness tests and the ability to perform simulated work tasks at three steel working sites. The authors performed a job analysis to determine the tasks involved in the job then selected 12 criterion tasks to serve as a sample of the work performed. These criterion tasks included lifting 50 and 75 lb bags (max in 5 min), shoveling earth (inches in work bin), shoveling slag (inches in work bin), working with a jackhammer, wheelbarrowing, hooking a chain, and other tasks. Generally, the tasks and criteria were not well defined in the article. A criterion work score composite was developed by standardizing and summing performance on all 12 criterion tests. Anthropometric and physical performance tasks included height, weight, isometric leg, arm, and back strength (exact methods not described), leg lifts, PUs, pull-ups, squat thrusts, the Harvard step test, balancing on a 1 inch board and a flexibility test. Multiple correlation analysis showed that the isometric arm test had the highest correlation with the criterion work score composite (r^2=0.67 to 0.72, depending on work site) and adding additional tests only marginally increased the correlation. An analysis of gender-specific regression lines generally showed that women were slightly over predicted resulting in a small bias against the men. A utility analysis was conducted assuming, based on conservative estimates from collected data, that wages were $18,000 per year and the performance of the strongest workers was 6 standard deviations above that of the weakest. This resulted in an estimated savings of about $5,000 per worker hired per year or about $9 million per year.

(b) Policemen. Wilmore and Davis (271) gave a physical fitness test battery to 217 male and 13 female California Highway Patrol officers then had them perform two simulated work tasks. The fitness tests involved a 1.5 mile run, isometric hand grip, bench press (1RM), vertical jump, sit-and-reach, and body composition (estimated from skinfolds). The criterion simulated work tasks were a barrier-surmount task and dummy drag. The barrier-surmount task involved running and scaling 2 walls (4'10" and 6'), simulating a handcuffing, then scaling the 2 walls again back to the starting point as fast as possible. The dummy drag involved pulling a 165 lb dummy from a car and dragging it 50 feet as fast as possible. The multiple correlations (r-values) between the fitness tests and the barrier surmount were 0.62 and between the fitness tests and the dummy drag, 0.57.

Arvey et al. (11) developed a physical fitness test for police officers in Minneapolis, Minnesota. Important and critical physical activities involved in the job were determined from examination of internal police reports on the use of force, officer generated reports on physical effort on the job, and surveys sent to officers. From examination of these data and theoretical considerations the underlying constructs of the job were defined as strength (ability to exert force against a load) and endurance (ability to sustain or recover from exertion of effort over time). A series of physical fitness tests (performance tests and physiological tests) and job performance ratings were obtained on 115 incumbent officers. The authors considered the fitness tests to include a 100-yd dash, dummy drag (120-lb dummy 50 ft, timed), obstacle course (jump hurdle, ditch, zigzag, crawl, climb 6-ft fence in 60 yards, timed), isometric grip strength, dummy wrestle (80-lb dummy, rotate, roll, place on spot, timed), SUs (1 min), dips,1-mile run, VO_2max (estimated from bicycle ergometry), body composition (skinfold estimate), height, and weight. For physical performance, supervisors rated officers on a 5 point scale (poor to superior) for running, wrestling, lifting and carrying, climbing, crawling, balancing, pushing/pulling, endurance, general physical fitness, and overall job performance. Confirmatory factor analysis produced 2 latent variables that were termed strength and endurance. The highest factor loadings on the strength factor were grip strength, lift and carry rating, push and pull rating, wrestling rating, and dummy wrestling. The highest factor loadings on the endurance factor were the obstacle course, 1-mile run, dips, 100-yd dash and SUs. The factorial structure of the physical components of the police jobs was confirmed. Portions of the model were cross-validated on 161 police applicants and the fit of the model was high.

(c) Firefighters. Davis et al. (56) examined the relationship between criterion simulated firefighting tasks and a variety of what the authors termed fitness and physiological measures. The fitness measures included anthropometry and body composition (height, weight, body fat estimates), strength (hand grip, SUs, chin-ups, long jump, PUs), and flexibility (sit-and-reach). The physiological measures included blood pressure, resting pulse pressure, resting heart rate, a 5-minute step test, and a Balke treadmill test. The simulated firefighting tasks included extending and retracting a long ladder, carrying a 33-kg hose up 5 flights of stairs, pulling a 24-kg rolled hose from the ground through a 5th story window, dragging a 53-kg dummy down 5 flights of stairs, and striking a rail 30 times with a sledge hammer to simulate forcible entry. One-hundred professional firefighters were selected from the District of Columbia area and tested. Canonical correlations identified two dimensions that defined the relationship between the simulated tasks and the fitness measures. These dimensions were a physical activity factor that involved muscle strength and aerobic endurance and a resistance to fatigue factor. Multiple regression analysis resulted in equations for predicting the two physical activity factors. The fitness/physiological tests involved in prediction the physical activity factors and their multiple correlations with the factor are shown in Table 12.

Table 12. Tests Predicting Physical Activity Factor and Resistant to Fatigue Factor in Study of Davis et al (56)

	Physical Activity Capacity Factor		Resistance to Fatigue Factor	
	Tests	Multiple R^2	Tests	Multiple R^2
Physiological and Fitness (Field) Tests	Hand Grip (kg)	0.90	Body Fat (%)	0.80
	Sit-ups (reps)		Fat-Free Mass ((kg)	
	Long Jump (cm)		HR_{max} (beats/min)	
	O_2 Pulse (mLO_2/beat)		Treadmill Grade (%)	
	HR_{max} (beats/min)			
Fitness (Field) Tests Alone	PUs (reps)	0.54	Body Fat (%)	0.60
	Sit-ups (reps)		Fat-free Mass (kg)	
	Hand Grip (kg)		Step Test (ml/kg/min)	

Gledhill and Jamnik (83) performed an analysis of jobs required by firefighters and developed tests that simulated the physically demanding tasks performed in the occupation. The main considerations in selecting the job tasks were that they were commonly encountered and essential, usually performed during a fire, and normally performed by a single firefighter. In developing task simulations the tests had to simulate as closely as possible the actual task, be measured in a standard and reliable manner, and be conducted in protective gear (48 lbs total weight) or simulated protective gear. There were 7 tasks developed. The Ladder Climb (untimed) involved going 40 feet up a ladder, uncouple and recouple a wall-mounted hose connection, and returning to the ground. The Claustrophobia Test (untimed) required wearing a blacked facemask, searching in an unlighted, narrow (14-in) passageway (30 feet) and recovering an 18-in doll. The Latter Lift (untimed) required removing a 24 ft, 56 lbs ladder from a bracket on a wall, placing it on the ground, and then returning it to the bracket. The Rope Pull (timed) involved lifting a hose roll weighing 50 lbs up to a third floor window (16 feet), then lowering it (repeated 4 times). The Hose Advance/Drag (timed) involved pulling a weighted sled 50 feet (requiring 154 lbs of force). The Hose Carry/Stair Climb (timed) involved lifting an 85 lb hose bundle and carrying that bundle up 5 floors (50 vertical feet). The Victim Drag (timed) involved grasping and dragging a 200 lb dummy for 50 feet weaving in and out of cones. Fifty-three firefighters (5.4 years of experience) completed the tasks. Acceptable times for the timed tasks were established as the mean plus one standard deviation and maximal times were the mean plus two standard deviations. Firefighters rated each task with regard to whether or not the tasks were 1) similar to the on-the-job task and 2) required the similar physical demands as the related on-the-job task. A Likert scale was used with 1 indicating strongly agree and 7 indicating strongly disagree. Average ratings of 1.4 to 2.5 indicated that the firefighters thought the task were similar to the job. Average ratings of 1.4 to 2.4 indicated that the firefighters thought the task had similar physical demands to the actual job. Although these tasks were not related to physical fitness this study is important because of the way the job analysis was developed.

Williford and coworkers (267) examined the relationship between a battery of health and fitness tests and simulated firefighting tasks. There were 91 male firefighters who were assessed. The health and fitness battery consisted of age, height, weight, resting heart rate, blood pressure, body composition (3-site skinfolds), pull-ups, SUs, grip strength, sit-and-reach flexibility, and a 1.5 mile run. The criterion firefighting tasks were done in sequence for time and involved 1) a stair climb (5 stories carrying a 22 kg hose section), 2) hoist (hoist a hose section weighing 16 kg from the ground to the fifth floor), 3) forcible entry (with a 4-kg sledge hammer drive a 75 kg I-beam 1.5 meters using an overhead stroke), 4) hose advance (carry a charged hose over the shoulder and move 30 meters), and 5) a victim rescue (drag an 80 kg dummy 31 meters). Fat-free mass and 1.5 mile run time produced a multiple correlation of 0.71 with the total time on the performance assessment. The addition of pull-ups increased the multiple correlation to 0.73.

Schonfeld and coworkers (227) had 25 men (not firefighters) from the Kennedy Space Center perform 3 simulated firefighting tasks and various physical fitness tests. The criterion firefighting tasks were performed in full firefighting gear (24 kg). The tasks included a Stairclimb (7 flights, 21 vertical meters), Chopping Simulation (3.6 kg sledgehammer, 30 strokes), and a Victim Drag (81-kg dummy, 26 m). All criterion tasks were timed. Physical fitness tests included isometric hand grip, PUs (1 min), SUs (2 min), sit-and-reach, Wingate upper body and lower body tests, isokinetic knee extension and flexion (60^{o}/sec, peak torque and average power), and a treadmill VO_2max test (Bruce protocol). Body composition was determined by skinfolds. Stepwise multiple linear regression showed that the total time on all 3 tasks could be predicted by treadmill time and knee flexion peak torque with an r=0.89.

(d) Gas Company Workers. Jamnik and Gledhill (130) developed an applicant screening test for a large multifaceted natural gas company. To develop the test, several steps were taken. A detailed job analysis was conducted that included time-motion studies, examination of tools and working environments, and measurements with experienced workers (posture, heart rate, force application). The jobs at the gas company were placed into 5 physical demand categories (high to low). Task simulations were developed that consisted of 7 items. Each item had different lifting and/or lifting and carrying requirements depending on the physical demand category (related to particular jobs). A Two-Handed Lift involving lifting a box from the floor to waist height. The Two-Handed Lift and Carry Task involved lifting a box from the floor to an upright position, ascending and descending 20 steps, then placing the box on a table 3 feet above the ground. The One-Handed Lift and Carry Task involved picking up a simulated tool box, ascending and descending 20 steps, changing hands, ascending and descending 20 more steps, then returning 165 feet to the start. The Two-Handed Lift to Chest Task involved picking up a steel box from a ledge 4 feet from the ground and returning it to the ground a number of times (depending on physical demand category). The Upright Appliance Push and Pull Task involved tilting a simulated appliance so the front was 6 inches above the ground then pulling the appliance. The Simulated Shoveling Task involved lifting 20, 15-lb shovel loads from the ground and placing them 4 feet above the ground. Sledgehammering involved 30 controlled but forceful 2-handed overhead stokes with a 10 lb sledgehammer. The first 4 tests were performed by all applicants but the last 3 tests were performed only by applicants for specific jobs. The tasks were validated by having incumbent workers rate each task with regard to whether or not they were 1) similar to the on-the-job tasks, 2) required similar physical demands compared to on-the-job task. A Likert scale was used with 1 indicating strongly agree and 7 indicated strongly disagree. Average ratings of 1.9 to 2.6 indicated the incumbents thought the task were similar to the job. Average ratings of 2.0 to 2.7 indicated the incumbents thought the task had similar physical demands to the actual job. Again, this study did not validate fitness tests against the simulated tasks but the study is important for the job analysis in a very complex industrial environment.

(e) Divers. Marcinik et al. (179) examined the relationship between the U.S. Navy Fleet Diver Physical Screening Test (FDPFT) and tasks involved in Navy diving. The FDPFT (with passing criteria in parentheses) involved a 500-yd swim (14 min), PUs (42 in 2 min), SUs (50 in 2 min), pull-ups (6 with no time limit) and a 1.5 mile run. The Navy diving tasks involved a Tool Bag Swim (swim 200 ft in scuba gear carrying a 10 kg tool bag without touching pool bottom), Fin Kick (wearing fins, remain on water surface for 5 min with arms and hands out of the water), Ladder Climb (time to ascend and descend a 14-ft ladder with scuba gear), SCUBA-Bottle Carry (carry SCUBA bottles 450 feet), and the Umbilical Pull (pull a 100 lb umbilical line 50 ft upward). There were 146 diver candidates in the study. The authors showed scatterplots that suggested little relationship between the job performance tasks and items in the FDPFT. The authors developed a "shipboard task performance score" that was said to be the time to complete the tasks. However, since two Navy diving tasks were pass/fail (tool-bag carry, fin-kick) it is unclear how or if these tasks were included. The authors stated that "results of the regression analysis between the physical screening test and shipboard tasks showed screening test scores were not predictive of the 3 representative shipboard tasks". However, the regression results were not presented in the article. The authors suggest that tests of muscular strength involving moving external objects should be included in the FDPFT since fleet diver jobs involve high strength demands on the arms, legs and back.

(f) Multi-Occupational Study. Hogan (115) performed a secondary analysis of 7 studies that examined a variety of tests administered to adults in a variety of occupations. These occupations included grocery warehouse workers, outdoor telephone workers (pole climbers), oil refinery workers, steelworkers, metal/chemical processing maintenance operators and workers, chemical/plastic/synthetics/paint maintenance and production workers, and chemical/refining/drilling maintenance and production technicians. She showed that 3 major fitness components could account for the structure of physical performance in these occupations: muscle strength, cardiorespiratory endurance and movement quality. Movement quality was specific to the job and included factors like balance, flexibility and coordination.

(2) Military Studies

(a) British Army. Rayson et al. (217,218,221,222) examined the relationship between a series of criterion military tasks and physical fitness tests in a very comprehensive set of studies. They first performed a job analysis (221) which consisted of obtaining information from various specialists on the most physically demanding tasks and observing, filming, and measuring a sample of these tasks. It was found that activities most frequently performed were lifting (88%), carrying (48%), pulling (6%), pushing (3%), climbing (3%), marching (2%), running (2%), and crawling (2%). About 55% of tasks involved a combination of activities with lifting and carrying comprising 89% of these. Vertical lifting distance ranged from ground to overhead with 70% of the lifts beginning at ground level. Fifty-seven percent of the lifts were to waist height, 28% to shoulder height, and 15% overhead. Distance of carries were from 2 to 32 m with 62% of carries <10 m, 18% of carries from 11 to 50 m, 6% of carries from 51 to 100 m, and 15% of carries >100 m. Where external loads were involved, forces ranged from 10 kg to 111 kg. Heart rates ranged from 55 to 88% of the maximum and oxygen uptake ranged from 1.2 to 2.9 L/min. This job analysis resulted in the development of 4 criterion tasks with 3 levels each. The tasks and levels are shown in Table 13. From these tasks measurable criterion tasks were developed and are shown in Table 14.

Table 13. Selected Common Military Tasks and Task Levels in the Job Analysis by Rayson et al. (217)

Task Level	Single Lift of Ammunition Box	Jerry Can Carry – 2 Cans, 20 kg, One in Each Hand	Repetitive Lift and Carry of Ammunition Box	Road March of 12.8 km in 120 min
1	44 kg to 1.70 m	210 m	44 kg, ground to 1.45 m, 1/min, 20 min	25 kg load
2	35 kg to 1.45 m	90 m	22 kg, ground to 1.45 m, 3/min, 15 min	20 kg load
3	20 kg to 1.45 m	30 m	10 kg, ground to 1.45 m, 6/min, 10 min	15 kg load

Table 14. Criterion Tasks and Measures Developed in the Rayson et al. Study (217)

Task Level	Single Lift of Ammunition Box (Measure: Max load up to 75 kg)	Carry (Measure: Time to exhaustion)	Repetitive Lift and 10 m Carry of Ammunition Box (Measure: Time to exhaustion up to 60 min)	Road March of 12.8 km (Measure; Time to complete)
1	To 1.70 m	Jerry cans, 20 kg each, one carried in each hand, 1.5 m/sec	44 kg, ground to 1.45 m, 1/min	25 kg load
2	To 1.45 m		22 kg, ground to 1.45 m, 3/min	20 kg load
3			10 kg, ground to 1.45 m, 6/min	15 kg load

In a subsequent study, tasks were related to a series of anthropometric and physical fitness measures (217). Subjects were 340 men and 75 women from various specialties in the British Army. Anthropometric measures include height, weight, arm span, biacromial diameter, elbow diameter, neck girth, chest girth, waist girth, and gluteal girth. Fitness measures covered strength, muscular endurance, cardiorespiratory endurance and body composition components. Isometric strength was measured with the upright pull, arm flexion, hand grip, back extension, and plantar flexion. Dynamic lifting strength was measured with a hydro-dynamometer and an incremental dynamic lift (IDL). Muscular endurance was measured with SUs, PUs, pull-ups, isometric arm flexion (time holding 14 kg at 90° of elbow flexion), dynamic arm flexion (repeatedly lifting 15 kg to cadence), and dynamic shoulder flexion (repeatedly pulling 15 kg to cadence). Cardiorespiratory endurance was measured with the aerobic 20-m aerobic shuttle run. Body composition was determined from skinfolds. Separate regression equations were developed for each of the criterion tasks (Table 15). Single lifting tasks demonstrated high relationships with fat-free weight and muscle strength measures. The carrying models incorporated strength variables and anthropometrics but errors were large. The repetitive lifting models included muscular strength, muscular endurance, and anthropometric measures but errors of prediction were large. Road march tasks were predicted from the aerobic shuttle run, body weight, body fat, and arm flexion endurance.

Table 15. Criterion Tasks and Models for Prediction from Rayson et al. Study (217)

Criterion Task	**Model**[a]	R^2
Single Lift 1.7 m (men)	-22.5+0.011*UP+0.829*FFM+0.014*BES	0.59
Single Lift 1.7 m (women)	-19.1+0.930*FFM+5.817*IDL145/WEIGHT	0.40
Single Lift 1.45 m (men&women)	-13.2+0.017*BES+0.999*FFM+6.706* IDL145/WEIGHT -6.013*GENDER	0.88
Carry (men&women)	Exp(0.35+0.022*PU+0.022*ARM+0.019* LogDAFE-0.174*GENDER)	0.70
Repetitive Lift, 44kg (men)	406.0+1.527*LIFT POWER-606.689* IDL170/WEIGHT+0.027*(SU*WEIGHT)	0.55
Repetitive Lift, 22 kg (women)	-1440.1+16.51*DAFE+3.284*HG	0.55
Loaded March 10 kg (women)	-801.1+2.608*UP	0.38
Loaded March, 25 kg (men)	142.7-19.765*VO2+0.530*WEIGHT-0.052*SAFE	0.40
Loaded March, 20 kg (men&women)	132.7-0.072*VO2+14.134*GENDER	0.55
Loaded March, 15 kg (men&women)	233.4-0.108VO2-11.661-11.66+LogSAFE-0.534*BF	0.75

[a]FFM=fat-free mass; BES=isometric back extensor strength; UP=isometric upright pull ; IDL145=incremental dynamic lift to 1.45 m; IDL170=incremental dynamic lift to 1.7 m; WEIGHT =body weight; PU=pull-ups; VO2=VO_2max predicted from aerobic shuttle run; FAT=body fat; Exp=Exponential; log=logarithm; DAFE=dynamic arm flexion endurance; SU=sit-up; HG=hand grip; BF=body fat; LIFT POWER=power on hydrodynamometer; ARM=arm span; SAFE=static arm flexion endurance

The predictive models were cross-validated in a separate study (218). Cross-validation involved testing regression equations developed on one sample on a second sample to see how well the original equations fit the second sample. There were 214 men and 112 women that served as subjects for the cross validation. The soldier military occupational specialties (MOS) included infantry, engineering, administration (adjutant general), intelligence, and logistics specialties. Essentially the same fitness measures taken in the previous study were administered to recruits at Weeks 1, 5, and 9 of basic training. The road march tests were only given at Week 9 for fear of injuring the trainees. The only modifications to the criterion tasks in Table 14 included a 90 minute maximum time for the repetitive lift. Successful predictive models were defined as those with a) consistent prediction from the validation and cross-validation studies, b) similar standard deviation (SD) of residuals between measured and predicted task values in validation and cross-validation samples, and c) similar mean change scores for measured and predicted task values at different stages of training. Models that best met the criteria are shown in Table 16. In some cases, single gender models had to be developed to account for different regression intercepts and/or slopes and to improve the accuracy of the model. Results showed that the three single lifting models had accuracy across the validation and cross-validation samples with a small number of misclassifications. The authors recommended these single-lift prediction models for the evaluation of recruits at weeks 1, 5 and 9 with no further validation. The carry model had several anomalies and errors that caused the authors to recommend further validation. The models from the validation sample of the repetitive lifting tasks were not tested against the cross-validation sample because of differences in the test procedures (60 min vs 90 min). The repetitive lifting cross-validation samples also produced large SDs and a low r^2 for the 10 kg task leading the authors to recommend further validation trials. The loaded march models involving 15 and 20 kg had accuracy across the validation and cross-validation samples with small numbers of misclassifications. The authors recommended use of the 15 and 20 kg road march models for the evaluation of recruits at Week 9 with no further validation. The loaded march model involving 25 kg required a larger sample of women before the authors would recommend it for use. It was recommended that 9 new physical fitness tests be adopted including body weight, % body fat, static and dynamic lift strength, back extension strength, static arm endurance, pull-ups, and the 20-m aerobic shuttle run.

Table 16. Criterion Tasks and Models for Prediction from Rayson et al. Study (218)

Criterion Task	Model[a]	R^2	Status[b]
Single Lift 1.7 m (men)	-33.386+0.75*FFM+0.011*BES+0.012*UP+6.15* TRAINEE+1.88*STEP	0.55	RFU (Weeks 1,5,9)
Single Lift 1.7 m (women)	-28.624+0.827*FFM+ILM145/WEIGHT+1.333*VISIT	0.50	RFU (Weeks 1,5,9)
Single Lift 1.45 m (men&women)	-20.92+0.935*FFM+11.04*ILM145/WEIGHT-5.67*GENDER+3.657*STEP	0.83	RFU (Weeks 1,5,9)
Carry (men)	Exp(2.68+0.16*FFM+0.015*PU+0.00029*UP+0.136* logSAFE+0.334*TRAINEE)	0.49	RFV
Carry (women)	Exp(1.815+0.0015*UP+0.0022*SAFE+0.702* TRAINEE)	0.58	RFV
Repetitive Lift, 44kg (men&women)	-4641.0+3.59*SAFE+117.84*PUL+84.6*FFM	0.47	RFV
Repetitive Lift, 22 kg (men&women)	$(-28.24+6.53*VO2+0.67*FFM+4.5*STEP)^2$	0.55	RFV
Repetitive Lift 10 kg (women)	Exp(5.44+0.0029*UP-0.049*FAT+0.57*STEP)	0.30	RFV
Loaded March, 25 kg (men&women)	161.37-16.543*VO2+0.353*WEIGHT-0.044*SEFA-9.175*TRAINEE	0.43	RFV
Loaded March, 20 kg (men&women)	120.45-0.052*VO2-0.013*BES+12.31*GENDER+ 6.663*TRAINEE	0.45	RFU (Week 9)
Loaded March, 10 kg (men&women)	192.95-0.088*VO2-6.04*logSAFE-0.016*BES	0.71	RFU (Week 9)

[a]FFM=fat-free mass; BES=isometric back extensor strength; UP=isometric upright pull ; TRAINEE=recruit or soldier; STEP=week of training; ILM145=incremental lift machine to 1.45 m; WEIGHT =body weight; VISIT=week of visit; PU=pull-ups; SAFE=static arm flexor endurance;VO2=VO_2max predicted from aerobic shuttle run; FAT=body fat; Exp=Exponential; log=logarithm
[b]RFU=ready for use; RFV=requires further validation

In a subsequent analysis (222), the 9 measures (now called the Physical Selection Standards for Recruits (PSSR)) were validated against specific measures of recruit success in basic training. The measures of recruit success (criterion tasks) were performance on 4 specific representative military tasks (not specified in the article), number of duty days lost for medical conditions, attrition from training, and self, peer, and supervisor performance ratings. The PSSR measures were fat-free mass, isometric back extensor strength, 38-cm upright pull, IDL to 1.45 m, body weight, pull-ups, static arm endurance, VO_2max predicted from aerobic shuttle run, and body fat. There were 315 recruits (271 men, 44 women) who completed all testing. The PSSR correctly predicted outcomes on all 4 recruit success criteria in 75% of recruits. Compared to those that failed their PSSR, those that passed had fewer medically restricted days, were more likely to complete training, and had higher job performance ratings.

We spoke to Mark Rayson (14APR04) to get an update on the current status of the PSSR. The PSSR was implemented by the British Army in 1998 for recruit selection. An individual's predicted criterion task scores were compared to criterion scores for specific jobs to see if the individual qualified. From 1998 to 2002 several changes were made to the basic training program that called into question the criterion tasks selected. In addition, a 2.4-km run was added to the test battery so that 2 highly intercorrelated cardiorespiratory endurance tests (2.4-km run and the aerobic shuttle run) were included in the test battery. In 2001-2002, Dr. Rayson and colleagues conducted additional studies to confirm the validity of the PSSR. In the new studies, criterion tasks included a single lift to 1.45 m, a carry of jerry cans (20kg, one in each hand, 1.5 m/sec pace to exhaustion), marches of 6 miles with loads of 15 or 20 kg, and a march of 8 miles with 25 kg. The lift and carry tasks were not included in the criterion test battery because of problems in trying to predict them. The criterion tasks were administered at the end of recruit training. Obtained from records of recruits on entry to service were weight, height, body composition, static arm endurance, back extension strength, 38-cm upright pull, IDL, pull-ups, the aerobic shuttle run, 2.4 mile run and other data. Based on an analysis of these data, factors in the revised models are shown in Table 17. There were large errors in trying to predict the single lifting tasks so it was suggested that these criterion tasks actually be performed in place of predictive fitness tests. Because of the importance of lifting and carrying it was also suggested that a lift and carry task actually be performed in place of fitness tests. Road-march performance could be predicted. Passing standards for each test are set by the Arms and Services on each criterion task and there is a table listing every MOS in a pamphlet entitled "Fit to Fight". A risk management approach was taken such that probabilities of successful performance (90%, 80%, 70% etc.) could be assigned to different scores. This allows the British Army to accept recruits with less likelihood of passing when it is necessary to fill recruiting quotas.

Table 17. Factors in Revised Predictive Models for British Army

Criterion Task	Predictor Tests[a]	R^2
Single Lift, 1.45 m (recruit)	IDL, BES	0.58
Single Lift,1.45 m (infantry)	IDL, PU, WEIGHT, BES	Not specified
Carry	Not Recommended	<0.21
Loaded March, 15 kg, 6 mile	UP, 2.4-km time	0.50
Loaded March, 20 kg, 6 mile	pVO_2max	0.39
Loaded March, 25 kg, 8 mile	pVO_2max , UP	0.53

[a]BES=isometric back extensor strength; UP=isometric upright pull ; IDL=incremental dynamic lift to 1.45 m; PU=pull-ups; pVO_2max=predicted VO_2max from run test

(b) Canadian Military Services

Lee (169) related criterion military tasks to laboratory measures of aerobic capacity and anaerobic endurance. Based on literature reviews, interviews, and field observations, a committee of Canadian Army personnel selected the following as criterion military tasks: dig a slit trench using a standard issue shovel, perform a loaded road march (25 kg load), evacuate a casualty (over the shoulder fireman's carry for 100 m), carry/empty a jerry can (carry 35 kg can 35 m, empty can; repeat 3 times), and lift ammunition boxes (1.3 m lift, 48 boxes). The fitness test measures included 1) a direct, graded, uphill running treadmill VO_2max, 2) Wingate arm test and 3) Wingate leg tests. A total of 99 infantry soldiers completed all the criterion tasks and the physical fitness tests. Separate equations were developed for each criterion variable using stepwise multiple linear regression. The predictor variables and r^2 are shown in Table 17. In general, the correlation coefficients were low.

Table 17. Criterion Tasks and Predictor Variables in Lee Study (169)

Criterion Test	Predictor Variables	Multiple R^2
Dig Slit Trench	Leg Maximal Power Output	0.39
	VO_2max	
	Arm Power Decline	
	Arm Peak Power	
	Leg Power Decline	
Loaded Road March	Leg Maximal Power Output	0.03
	Arm Power Decline	
	Arm Peak Power	
	Leg Power Decline	
	VO_2max	
Evacuate a Casualty	Leg Maximal Power Output	0.24
	Leg Power Decline	
	Arm Power Decline	
	VO_2max	
	Arm Peak Power	
Carry/Empty Jerry Can	VO_2max	0.09
	Arm Power Decline	
	Leg Maximal Power Output	
	Leg Power Decline	
	Arm Peak Power	
Ammunition Box Lift	Leg Maximal Power Output	0.23
	VO_2max	
	Leg Power Decline	
	Arm Power Decline	
	Arm Peak Power	

Chahal (38) used the same criterion tasks as Lee (169) but examined body composition, muscular strength and muscular endurance measures. Test/retest reliabilities for the tasks were: trench dig, 0.86; casualty evacuation 0.85, jerry can carry, 0.83 and ammunition box lift, 0.90. The fitness tests consisted of isometric hand grip, isometric arm flexion, isometric trunk flexion and extension, isokinetic knee extension and flexion (180°/sec), concentric and isokinetic arm flexion (30°/sec), concentric and isokinetric trunk flexion and extension (15°/sec), concentric and isokinetic bench press (30°/sec), concentric and isokinetic shoulder extension (30°/sec), concentric and isokinetic leg extension, isometric hand grip endurance (hold 21 kg), isometric elbow flexion endurance (hold 20 kg at 105° elbow angle), and dynamic shoulder extension endurance (10 contraction/min, 21 kg, to exhaustion). Body composition was determined by densitometry. Subjects were 116 infantry soldiers from the Canadian Forces. Table 18 shows the results of the stepwise multiple linear regression analysis. Cutoff scores for successful performance on each task were suggested by a panel of 5 expert military judges and are shown in Table 19. Based on these judgments and discriminate function analysis, minimal scores on the performance tests were set as shown in Table 20.

Table 18. Stepwise Multiple Regression Results from Chahal Study (38)

Criterion Task	Predictor Variables	r^2	SEE[a]
Casualty Evacuation	Static trunk flexion, body fat	0.19	8 sec
Ammunition Box Lift	Static trunk extension, body fat	0.15	48 sec
Jerry Can Task	Static trunk flexion	0.08	29 sec
Digging Task	Leg extension strength, dynamic shoulder extension endurance	0.28	38 sec
Road March	[b]	[b]	[b]

[a]SEE=Standard error of estimate

[b]None of the performance tests met the $p<0.05$ criteria set by the investigator

Table 19. Cutoff Scores for Successful Criterion Task Performance from Chahal Study (38)

Criterion Task	Suggested Times (sec)
Casualty Evacuation	60
Ammunition Box Lift	300
Jerry Can Task	300
Digging Task	360
Road March	No Suggested Time

Table 20. Fitness Measures Suggested Field Task Performance Standards From Chahal Study (38)

Fitness Measures	Suggested Performance Level
Static trunk flexion	58 kg
Static trunk extension	145 kg
Leg extension strength	203 kg
Dynamic shoulder extension endurance	74 reps
Body fat	23.4%

Singh et al. (234) used the same data as Chahal (38) and Lee (169) and reported slightly higher relationships with the physical performance variables when muscle strength, muscular endurance and cardiorespiratory endurance variables were included in the regression models. The results are shown in Table 21. Using the same professional judgment and discriminate function analysis as Chahal (38), Table 22 shows the suggested criteria for the task performance and body composition. Criterion tasks and performance/body composition measures were also obtained on 45 female soldiers but no multiple regressions were performed and no attempt was made to combine the data with that of the men to develop gender-free models.

Table 21. Stepwise Multiple Regression Results from Singh Study (234)

Criterion Task	Predictor Variables	r^2	SEE
Casualty Evacuation	Static trunk flexion, dynamic shoulder endurance	0.24	7 sec
Ammunition Box Lift	VO_2max, body fat	0.25	46 sec
Jerry Can Task	Static trunk flexion	0.08	29 sec
Digging Task	Static trunk flexion, VO_2max, leg peak power	0.36	37 sec
Road March	VO_2max	0.05	49 min

Table 22. Strength and Body Composition Characteristics of Soldiers for Suggested Field Task Performance Standards From Singh Study (234)

Performance or Body Composition Characteristic	Suggested Performance Level
Static trunk flexion	58 kg
Static trunk extension	145 kg
Leg extension strength	203 kg
Dynamic shoulder endurance	74 reps
Body fat	23.4%
VO2max	3.1 L/min
Leg Peak Power	630 W

Stevenson et al. (242,243) developed a set of criterion military tasks and attempted to validate a modified Canadian Standardized Test of Fitness (also called the Exercise Prescription Test or EXPRES test) against these criterion tasks. The fitness tests included isometric hand grip (both hands), SUs (1 min), PUs (1 min), and a step test (to estimate VO_2max). The criterion military tasks included land evacuation (one person test with wheels on back of litter, 80 kg person on litter, 0.75 km), sea evacuation (dressed in fire fighting gear, move an 80 kg person 12.5 m on a stoker litter, then push up and down staircase), entrenchment dig (move 0.5 cubic m of crushed rock from one box to another), sandbag carry (move 20 kg sandbags 50 m and do as many as possible in 10 min), and low/high crawl (low crawl 30 m, high crawl 45 m with fatigues, helmet and rifle). Test-retest reliabilities ranged from 0.93 to 0.99 (no reliability reported for the sandbag carry). In separate studies, younger (<35 years) and older (≥35 years) individuals were tested. Older individuals were restricted to working at no higher than 90% maximum heart rate on the criterion tasks based on American College of Sports Medicine guidelines (2). For the study of younger soldiers, 66 men and 144 women were selected such that the group was evenly distributed across EXPRES quartiles; older subjects were similarly distributed and consisted of 100 men and 66 women (not clear if this latter group involved soldiers). In multiple stepwise regression for younger individuals, r^2 between criterion tasks and various EXPRES scores ranged from 0.14 to 0.48 for men and 0.14 to 0.41 for women. The r^2 for the older individuals are not presented except to say that the highest values were 0.49 for men (sandbag carry) and 0.55 for women (low/high crawl). The authors state that the EXPRES test was related to task performance but could not well predict it. Individuals who were above the 75th percentile for all criterion task performances were identified. Their EXPRES scores were converted to Z scores and the 95th percentile identified for each fitness test. The 5th percentile on each fitness test for those who achieved the 75th percentile on all criterion tasks became the passing score. The proposed passing scores are shown in Table 23. In examining the number of individuals falsely classified it was found that for younger individuals, there were only 8% false negatives (failing a person who could actually perform the criterion tasks) and 28% false positives (passing a person who could not do all the criterion tasks); for older individuals there were 7% false negatives and 28% false positives. The majority of false positives were women. The restriction of 90% maximum heart rate probably influenced the older individuals' scores. This hypothesis was tested by having older individuals perform an unrestricted entrenchment digging task (no maximum heart rate). The unrestricted task resulted in a 38% improvement in time.

Table 23. Proposed Minimum Passing Standards for EXPRES Test (From Reference 242)

Test Item	Men		Women	
	<30 years	≥35 years	<30 years	≥35 years
Predicted VO_2max (ml/kg/min)	39	35	32	30
Hand Grip (both hands, kg)	75	73	50	48
Sit-ups (n)	19	17	15	12
PUs (n)	19	14	13	7

A problem with the studies by Stevenson et al (242,243) is that the authors took an existing test (the EXPRES) and attempted to show a relationship to task performance. A better approach would have been to test a wide variety of fitness tasks and use these to predict task performance as in the studies by Singh et al. (38,169,234). Further, the 90% maximum heart rate restriction on older individuals was shown to affect criterion task performance probably resulting in an underestimate of the fitness level required (dig entrenchment). Stevenson et al. did go to great lengths to standardize the tests and establish reliability.

(c) Royal Netherlands Army

As described by Bertina (23), the Royal Netherlands Army previously had an Assessment of Physical Capabilities (APC) Test that includes the measurement of isometric muscle strength (5 muscle groups), vertical jump, cycle ergometer predicted VO_2max, and body fat assessment. Entry standards were based on functional groupings of MOS and on percentile rankings. The functional groups (FG) were: a) FG1 which includes combat units like infantry and engineer, b) FG2 which includes combat support units like artillery, and 3) FG3 comprised of "logistical units". Entry standards were based on percentiles of the young male population with performance required at or above the 50th percentile for FG1, 25th percentile required for FG2, and 10th percentile for FG3. Women have a lower standard for FG3.

These entry standards were based on percentiles and not actually job demands so a study was conducted. The criterion military tasks were established by NATO working groups. The criterion military tasks included road marching, repetitive lifting, digging, and carrying. A Digging Task required soldiers to empty a container containing one cubic meter of sand as rapidly as possible using a standard entrenching tool. The Loaded Road Marching Task involved a progressive, interrupted test in which the intensity was increased by manipulation of the load and speed. Loads of 25 kg, 38 kg and 50 kg were carried in sequence at a speed of 6 km/h; a 63kg load was carried at 6, 6.5 and 7 km/h. The performance measure was distance covered until the soldier was unable to maintain the pace. The Repetitive Lifting Task involved a progressive, interrupted lifting of a box from the floor to 145 cm. The initial weight in the box was 12 kg and soldiers were required to lift the box 1 time/10 sec for 9 repetitions. Thirty sec of rest was given then the weight was increased in 4 kg increments. This sequence was repeated until the soldier could not keep up with the pace. The performance measure was the number of repetitions. The Carry Task involved a progressive, interrupted jerry can carry of 90 m at a pace of 5.4 km/h. The initial load was 15 kg was increased by 4 kg each trip with 1 min rest between trips. The task ended when the soldier could not maintain the pace and the performance measure was the distance covered (252,253,255,256).

Physical performance tasks included the APC battery (described above), as well as laboratory measures of fitness. A "functiongram" and "somagram" were developed. The functiongram was a 5-digit code for a particular MOS that describes the general fitness requirement and requirement on the 4 criterion tasks. The somagram described the physical profile of the individual (23).

Dr. Jos van Dijk provided us with English summaries of efforts by the Royal Netherlands Army to validate physical fitness test measures against the criterion military tasks (252,253,255,256). There were about 137 men and 61 women (soldier numbers differ slightly on each task) who were administered the criterion tasks and a number of physical fitness tests. The entire list of the administered fitness items not known because only the English abstracts of the studies could be reviewed. Separate equations for predicting the criterion tasks were developed for men and women. Table 24 provide the test items included in the in the equations used to predict the criterion tasks and the squared correlation coefficients.

Table 24. Royal Netherlands Army Criterion Tasks and Physical Fitness Test Predictors of Criterion Tasks (From References Numbers 252,253,255,256)

Criterion Task	Men		Women	
	Fitness Tests	r^2	Fitness Tests	r^2
Digging	Cycle ergometer VO_2max Static leg extension Fat-free mass 12-min run distance	0.30	Fat-free mass Arm ergometer VO_2max Elbow flexion isometric strength Cycle ergometer VO_2max	0.45
Loaded Road Marching	Height Static trunk extension 12-min run distance Squat strength (isokinetic)	0.56	Static lifting force at 40 cm Push-ups (2 min) Sit-ups (2 min) Bench Press	0.66
Repetitive Lifting	Elbow flexor isometric strength Isokinetic lifting force Shoulder press (isokinetic) Fat free mass	0.62	Elbow flexor strength Static trunk extension Static lifting force at 140 cm	0.72
Carry	Arm ergometer VO_2max Leg length Grip strength (weak hand) Static leg extension Push-ups (2 min)	0.39	Arm ergometer VO_2max Body length Static lifting force	0.49

(d) US Air Force and Navy Studies

In a 4-year investigation (1978-1982) Ayoub et al. (14,15) developed a fitness test for assigning Air Force personnel to physically demanding MOS. A job analysis was conduced using surveys, interviews of supervisors, reviews of technical manuals, and measurements of weights and forces. Manual material handling was found to be the most common physical activity with lifting, carrying, holding, and pushing/pulling the most common types of tasks. A series of simulated lifting, carrying, holding and pushing/pulling tasks were developed. From an original group of 28 tasks, 13 were selected that accounted for 90% of the tasks identified for all the MOS. For the lifting, carrying, and holding tasks, subjects adjusted their weight to the maximum they thought they could lift and/or carry (modified psychophysiological approach). For the pushing/pulling tasks subjects exerted maximal isometric force. Fitness tests included several IDL tasks (maximum weight lifted to 6 feet, to elbow height, and to knuckle height), an IDL task to elbow height that was held to exhaustion, isometric hand grip, isometric 38-cm upright pull, isometric one-handed pull, and an isometric two-handed elbow height pull. Separate multiple regression equations were developed for each of the 13 criterion tasks. Stepwise multiple regression showed that the IDL to 6 feet (IDL6) was the first variable to enter the equations (accounting for most of the variance) in 11 of the 13 task models and it was the second variable to enter the equation in the other two. Linear regression coefficients (r^2) with only the IDL6 as the predictor variable ranged from 0.34 to 0.80 with 10 equations above 0.64 (weighted models). IDL6 standards for each MOS were determined based on 1) IDL6 equivalent for each activity in the MOS,

2) IDL6 scores for the top 25 activities for all MOS, 3) assignment of "weights" to the activities based on proportion of airmen performing the task, frequency of task performance, and task criticality, 4) calculation of MOS weighted demand score in terms of IDL6, and 5) adjustment of weighted IDL6 based on the number of airmen in the MOS (assumes if MOS has more airmen some can assist in demanding tasks).

Robinson (225) examined the association of a battery of fitness tests to a cranking task for the Navy. The fitness tests were isometric hand grip, isometric arm pull, isometric arm lift, PUs, SUs, pull-ups, bent arm hang, body weight, height and a skinfold estimate of body composition. The criterion performance task was to turn as rapidly as possible the handles of an ergometer set to 600 kgm/min to simulate turning or pumping activity. The test sample consisted of 350 men and 493 women beginning Navy recruit training. In the article, individual correlations were presented between the fitness tests and the cranking task for men and women separately but no multiple correlation analysis was presented. Measures of hand grip, total body weight, fat-free mass, and weight/height were found to have the highest relationship with the arm cranking task. Robinson's physical performance tests were not selected as a result of job analysis but rather based on physical fitness constructs from the literature assumed to be involved in Naval tasks (225). The criterion task was actually an upper body Wingate task (using an absolute exercise load) designed originally to test upper body peak power and average power (16).

(e) US Army Studies

In 1976, a GAO report recommended that the military services develop fitness standards for more effective operational performance. The GAO report stated that the standards should be job specific and there should be no differentiation in standards between men and women (75). In July 1977, the Army Vice Chief of Staff directed that research be conducted to develop a gender-free occupationally-related fitness test that could be used for both 1) assignment to an Army MOS and 2) for Army physical training standards (260).

Vogel et al. (260) began the development of a system for establishing gender-free fitness standards that were occupationally related. There were 5 assumptions in the development of these standards. The first was that standards should be developed for two components of fitness, strength and cardiorespiratory endurance. Despite the fact that Vogel et al. identified 3 components, one (muscular endurance) was thought to overlap the first two (see Figure 1) and was not considered for the sake of simplicity. The second assumption was that the standards should be based on objectively determined demands of the MOSs. The third assumption was that the standards should be developed for clusters of MOS because many appeared to have similar demands. The fourth assumption was that the standards should be based on the task with the highest physical demand in each MOS. The fifth assumption was that the application of the standards in the field should be as simple as possible with relative gross resolution as long as the tests were meaningful in terms of job performance.

A task list was obtained from each Army service school that provided a detailed description of the physical demands of the MOSs. The MOSs were grouped by inspection into clusters having similar physical demands using the empirically derived criteria in Table 25. Four to 6 of the most physically demanding tasks in each cluster were selected for measurement and the weights soldiers lifted and energy cost of the tasks were measured. It was assumed, based on the literature, that an individual was capable of working at 45% VO_2max for an 8 hour day, and thus the VO_2max of each aerobic demand category could be set (e.g., a task of 8 kcal/min requires a maximal energy production rate of 18 kcals/min or a VO_2max of 3.6 L/min). Two sets of fitness tests were developed, one more technically involved for the MEPS and one for application in the field as shown in Table 26. The capacities determined from testing of soldiers would be related to the fitness measures using regression analysis. Several examples are shown (Table 27) but the full analysis is not presented nor are the cut-off values for the fitness tests. The analysis resulted in 5 clusters shown in Table 28.

Table 25. Criteria Used to Cluster Various MOS in the Study by Vogel et al. (260)

Intensity Rating	Strength (kg lifted floor to waist height)	Aerobic Demand (kcal/min)
Low	<30	<7.50
Medium	30-40	7.50-11.25
High	>40	>11.25

Table 26. Proposed Fitness Tests for Entrance to Service (MEPS) and On-The-Job (Field Test) from the Vogel et al. Study (260)

Component	Entrance (MEPS)	On-The-Job (Field Tests)
Aerobic	Heart rate from step test, body fat	2-mile run
Strength	Upright Pull	PUs, sit-ups, squat thrusts

Table 27. Examples of Representative Tasks in Different MOS Clusters (From Reference 260). For Echo Cluster Entire List is Shown.

Cluster	Representative Task
Alpha	Carry 45 kg bag 1000m in 20 min
Bravo	Lift and carry 41 kg ammunition box 6.7 m 32 times per hour
Charlie	Lift 25 kg projectile to 132 cm and carry 15 m, 50 times per hour
Delta	Lift and carry 27 kg container 15 m 40 times per hour
Echo[a]	Complete 8 km march in 120 min Dig 1-man emplacement in 45 min Lift and carry 23 kg, 50 m 8 times in 10 min Rush 75 m in 25 sec Low and high crawl 75 m in 90 sec

[a]The Echo cluster "representative tasks" are actually all tasks required for the Echo cluster

Table 28. Clusters of MOS by Strength and Aerobic Demands

Cluster	Physical Demand		MOS (n)	Total Personnel (%)
	Strength	Aerobic Demand		
Alpha	High	High	10	19
Bravo	High	Medium	39	13
Charlie	High	Low	63	21
Delta	Medium	Low	53	21
Echo	Low	Low	184	26

In a subsequent report, Sharp et al. (229) outlined 2 models to predict aerobic capacity and strength capacities. The model for aerobic capacity used VO_2max as the criterion, while the model for strength used maximal lifting capacity (MLC) as the criterion. The criterion VO_2max measure was obtained directly using a progressive uphill running protocol. The criterion MLC task involved lifting as much weight as possible from the floor to a 132 cm height (the height of the bed of a 2.5-ton truck). Two separate groups were tested, a group of recruits at Ft Jackson SC and a group of active duty soldiers at Ft Stewart GA. Although the initial group sizes were large, drop-outs and incomplete testing reduced the sample sizes to 86 for Ft Jackson (42 men, 44 women) and 222 for the Ft Stewart sample (181 men and 41 women). Fitness measures on the Ft Jackson sample included VO_2max estimated from heart rate on a step test, weight, age and body composition estimated from skinfolds. Fitness measures on the Ft Stewart group included isometric strength of the leg extensors, upper torso and back extensors, hand grip, upright pull (38 and 132 cm from ground) and body composition. Cross-validation was accomplished by splitting the samples into 2 approximately equal groups and analyzing them separately. The Ft Jackson sample was used to develop a model to predict directly measured VO_2max. The final model involved gender, step test predicted VO_2max and percent body fat (%BF) producing r^2 of 0.80, 0.78 and 0.84 in the validation, cross-validation (ridge regression was used to compensate for multicollinearity in the second sample) and combined samples, respectively. Because of the resources required for the step test heart rate monitor, a two factor model involving gender and %BF was developed and found to have r^2 of 0.78, 0.76 and 0.82 in the validation, cross-validation and total samples, respectively. The standard error of estimate (SEE) in predicting VO_2max was 3.5 mL/kg/min. The Ft Stewart data was used to develop a model to predict MLC. A model involving fat-free mass, upright pull (38 cm), and gender had an r^2 of 0.75, 0.74 and 0.79 in validation, cross-validation and total samples, respectively. The SEE in predicting MLC was 6.6 kg.

Based on this study, it was recommended in September 1980 that 2 tests be implemented in the MEPS. These tests were a skinfold estimate of body composition and the upright pull at 38 cm. Due to concerns on how this might affect manpower, the decision to implement these tests was deferred. In 1981 the Office of the Deputy Chief of Staff for Personnel showed renewed interest in an Army screening test for physical ability. A Women in the Army Policy Review Group conducted another task analysis of Army MOS and grouped the MOS into modified Department of Labor standards based only on lifting requirements (9). This system is shown in Table 29 (248). It should be noted that this job analysis emphasized lifting requirements and may have neglected other aspects of physical fitness such as cardiorespiratory endurance and muscular endurance.

Table 29. Modified Department of Labor Physical Demand Classification Standards (From Reference Number 248)

MOS Lifting Category	Occasional Lifting Requirement (kg)[a]	Frequent Lifting Requirement (kg)[b]
Light	9.0	4.5
Medium	22.7	11.3
Moderately Heavy	36.3	18.1
Heavy	45.3	22.7
Very Heavy	>45.3	>22.7

[a]Occasional is <20% of the time
[b]Frequent is >20% and <80% of the time

From 1982 to 1983, Teves et al. (248) performed a three phase study in which a group of new recruits was tested on entry to BCT (Phase 1), during the last week of BCT (Phase 2), and near the end of AIT (Phase 3). Sample sizes were 1) Phase 1: 1,984 (980 men and 1004 women); b) Phase 2: 202 (89 men and 113 women); and c) Phase 3: 970 (473 men, 497 women). AIT posts included Ft Jackson (281 men, 140 women), Ft Gordon (151 men, 234 women), Ft Sam Houston (19 men, 99 women) and Ft Lee (22 men, 24 women). Included in the physical ability test battery (called the Military Entrance Physical Strength Capacity Test or MEPSCAT) were isometric hand grip, isometric 38-cm upright pull, an IDL to 2 heights (152 cm and 183 cm), a bicycle test of predicted VO_2max (Astrand-Rhyming test), a step test of predicted VO_2max, and a skinfold estimate of body composition. The criterion task performance was a MLC from the floor to a height of 132 cm (MLC132). Results indicated that both men and women made substantial gains in strength (9% to 24%) and predicted aerobic capacity (16% for men and 20% for women) from Phase 1 to Phase 3 with the largest gains occurring from Phase 1 to Phase 2. Men and women were grouped by MOS Lifting Category (Table 29) based on their ability to lift the weight in the MLC test to 132 cm in Phase 3. The proportion of individuals who could lift the weight in each category is shown in Table 30. Using fat-free mass and IDL to 183 cm to predict MLC to 132 cm produced multiple regression correlation coefficients (r^2) of 0.33, 0.11 and 0.47 for men, women, and combined genders, respectively. Since the SSE was 18 kg for the gender combined equation it was not recommended for further use. The correlation (r^2) between the MLC132 and the IDL to 152 and 183 cm was 0.42 and 0.44, respectively, in a gender combined sample.

Table 30. Proportion of Individuals Who Could Lift to 132 cm the Weights Required By Their MOS Lifting Category (From Reference 248)

		MOS Physical Demand Category			
		Light/Medium	Moderately Heavy	Heavy	Very Heavy
Men	N	113	12	70	268
	Pre-BCT	100%	100%	96%	86%
	Post-AIT	100%	100%	100%	95%
Women	N	149	2	124	202
	Pre-BCT	97%	50%	1%	0%
	Post-AIT	100%	100%	12%	1%

Myers et al. (198) reported a separate analysis of the 3 phase study by Teves et al. (248) described above. They also collected and analyzed additional data. As a first step, Myers et al. performed a job analysis using data gathered from the Women in the Army Policy Review Group. They determined that the majority of physically demanding tasks in the Army involved lifting, carrying, pushing, and applying torque (turning a wrench) and devised criterion tasks based on these activities. The physical fitness test battery for Phase 1 included hand grip, 38-cm upright pull, predicted VO_2max (bicycle and step test), PUs, SUs, and a 1-mile run. Phase 2 included a 2-mile run. In their statistical analysis the actual criterion task used was a single number that served as a composite of all the individual tasks. How this single number was calculated was not described in the paper. Multiple correlation analysis was performed using the combination task criteria as the dependent measure and MLC to 152 inches, fat-free mass and 38-cm upright pull as independent measures. This produced an r^2=0.67 (men and women combined). Male and female equations were examined separately and found to have significantly different intercepts but similar slopes. When mean values were put into the general equation and into a female-specific equation there was little difference in the predictive values. However, the general equation (non-gender specific) slightly over-predicted women's performance (gave them a 4% higher score).

IDL machines were placed in the MEPS station in 1983. During the time the IDL was in place it was used only to suggest to enlistees that they might not meet the strength requirements of a particular MOS but it was not used to prohibit them from a particular MOS. Table 31 shows the weights recommended to be lifted to 132 cm (56 inches) by MOS category. Observation indicated that almost all recruits could meet the Light/Medium/Moderately Heavy category.

Table 31. Suggested Weights To Lift on IDL For MOS Categories

MOS Category (Modified Department of Labor Categories)	Recommended Weight Lifted (lbs)
Light / Medium / Moderately Heavy	40
Heavy / Very Heavy	70

VanNostrand et al. (254) analyzed data from the first 1.25 years of IDL use (January 1984 to March 1985). The proportion of individuals who could lift the "occasional" weight required in their MOS (Table 29) is shown in Table 32. These data are similar to those of Teves et al. (248) in Table 30. It was determined that if the IDL had been used in the MEPS for screening this would have resulted in a shortfall of 3,358 soldiers which represents 4% of the total recruit population but 33% of the female recruit population. It was not known how many individuals might have selected or did select another MOS as a result of the IDL.

Table 32. Proportion (%) of Individuals Who Could Lift in the MEPS Station the Occasional Lifting Requirement in their MOS (From Reference 254)

	MOS Physical Demand Category				
	Light	Medium	Moderately Heavy	Heavy	Very Heavy
Men	100	100	98	90	90
Women	100	92	17	6	6

Table 33 summarizes the US Army studies. Predictors of criterion task performance include gender, predicted VO_2max, body fat, fat-free mass, 38-cm upright pull, IDL183, and MLC152. In several equations fat-free mass and the 38-cm upright pull are included.

Table 33. Summary of Criterion Tasks and Predictor Tasks in US Army Studies

Study	Criterion Tasks	Predictor Tasks	R^2	SEE
Sharp et al (229)	Directly measured VO_2max	Gender, VO_2max from step test, %body fat	0.80	3.5mL/kg/min
	Directly measured VO_2max	Gender, % body fat	0.78	3.5 mL/kg/min
	Maximal lift to 132 cm	Fat-free mass, 38-cm upright pull, gender	0.75	7 kg
Teves et al. (248)	Maximal lift to 132 cm	Fat-free mass, IDL183	0.47	18 kg
Myers et al (198)	Combination of 4 criterion measures (lifting, carrying, pushing/pulling, applying torque)	MLC 152 inches, fat-free mass, 38-cm upright pull	0.67	Not specified

b. Physical Fitness and Injury Risk

Physical fitness tests could also be validated by examining their relationship with injury. Both military (105,133,134,144,147,157,158,210,220, 223,265) and civilian (108,239) studies have suggested that individuals who have low levels of physical fitness are more likely to become injured during occupational job activities. We have limited this section to examining associations between fitness and injuries in the military. With a few exceptions which will be discussed, the data is relatively consistent and can be summarized briefly.

Military personnel have a higher likelihood of injury if they have: a) low levels of cardiorespiratory endurance measured with 1-mile runs, 1.5 mile runs, 2-mile runs, aerobic shuttle runs, or 3000 meter runs (105,133,134,147,157,158, 210,220,265), b) low VO_2max measured with an uphill running protocol (158), c) low levels of muscular endurance measured with SUs or PUs (135,144,158,223), d) both high and low extremes of flexibility as measured by the sit-and-reach (135,158). Performance on the IDL was not shown to be associated with injury (48). Table 34 shows the associations between running performance and injury risk from several Army and Marine studies.

Table 34. Injury Risk during U.S. Military Basic Training, by Level of Aerobic Endurance (From Reference Number 82)

Gender/n	Branch of US Military	Location and Year of Study	Measure of Injury	Quartile 1 (fastest)	Quartile 2	Quartile 3	Quartile 4 (slowest)	p-value for trend
Women/ 79	Army	Ft. Jackson 1984	Proportion Injured Risk Ratio	36% 1.0	33% 0.9	57% 1.6	61% 1.7	0.03
Men/ 140	Army	Ft. Jackson 1984	Proportion Injured Risk Ratio	14% 1.0	10% 1.4	26% 1.9	42% 3.0	0.02
Women/ 680	Army	Ft. Jackson 1998	Proportion Injured Risk Ratio	39% 1.0	55% 1.4	59% 1.5	60% 1.5	0.02
Men/ 488	Army	Ft. Jackson 1998	Proportion Injured Risk Ratio	21% 1.0	23% 1.1	32% 1.5	30% 1.4	0.01
Women/ 265	Marine Corps	Parris Island 1993	Odds Ratio (95% CI)[a]	1.0[b]	2.2 (1.1-4.4)	2.2 (1.1-4.5)	2.4 (1.2-5.1)	NA
Men/ 369	Marine Corps	Parris Island 1993	Odds Ratio (95% CI)	1.0[b]	2.1 (1.1-4.2)	1.3 (0.6-2.6)	2.1 (1.1-4.3)	NA

[a]95% Confidence Interval
[b]No confidence interval since this is the reference category

The association of BMI with injuries is less certain but there is a suggestion of a J-shaped curve. This means that risk is somewhat elevated at very low BMI levels, there is reduced risk at moderate BMI levels, and risk is again elevated at higher levels (54,102,104,133,135,158,223,262). It should be noted that there are specific height weight standards for entry into service and stricter standards for retention in service (159). Thus, individuals with higher BMIs are not likely to enter or be retained in service so the BMI distribution is somewhat skewed.

Myers et al. (198) examined the relationship between physical fitness on entry to service and sick call and restricted duty in BCT. Fitness measures included hand grip, 38 cm upright pull, predicted VO_2max (bicycle and step test), PUs, SUs, and a 1- or 2-mile run. They collected medical data including sick call visits and days of restricted duty but the source of the data is not clear and they noted that the medical data was very incomplete. In correlational analysis, they found little relationship between restricted duty days and any of the physical performance tests ($r=-0.29$ to 0.22). Correlational analysis was an inappropriate statistical technique in this case. Many individuals would have zero scores (no sick call days) resulting in a highly skewed distribution. A more appropriate analysis would have involved separating fitness scores into risk groups and analyzing sick call visits and/or profile days in these groups. No analysis was done on injury incidence (who was and was not injured) which could have yielded additional data.

In summary, low aerobic fitness, low muscular endurance and both high and low levels of flexibility are strongly associated with higher injury incidence. There is a suggestion that both high and low levels of BMI are also associated with higher injury incidence. The use of inappropriate statistical techniques, as in the study by Myers et al. (198), can lead to incorrect or misleading conclusions.

c. Physical Fitness and Attrition Risk

For the present purposes, attrition can be defined as the failure of a service member to complete his or her contractual enlistment obligation (167). Because of the importance of attrition to the military (143), this can serve as another criteria against which fitness measures can be validated.

Bernauer and Bonanno (22) administered a 40-item test to 241 job applicants at a pole climbing school. The tests included measures of physical characteristics and body composition (age, gender, height, weight, body fat and fat-free mass from skinfolds), static strength (e.g., hand grip, shoulder abduction, elbow flexion and extension), muscular endurance (chin-ups, SUs, PUs), cardiorespiratory endurance (e.g., treadmill walk time, treadmill recovery heart rate, bicycle estimated VO_2max), balance (beam walking), flexibility, and response time. Factor analysis reduced the 40 item battery down to 7 factors that accounted for 89% of the test variance. Items with the highest loadings on the 7 factors were percent body fat, grip strength, SUs, recovery heart rate, body weight, reaction time and beam walking. The reduced test battery (with a step test substituted for the recovery heart rate) was given to 300 pole climbing school applicants. Difference in test items between those who successfully completed the school and those who did not were compared. The step test and balance tests were significantly different for successful and unsuccessful students. For men only, body fat was also different between successful and non-successful students. A major problem with this study was that the authors compared only mean values of successful and unsuccessful pole climbing school students. The appropriate statistical test would have been chi-squares or logistic regression to identify factors that determined the odds of success or failure.

Gunderson et al. (91) examined the relationship between graduation from underwater demolition training and measures of physical fitness and health status. A sample of 293 enlisted men in underwater demolition school were examined. The physical fitness tests are not fully described in the article but the health status measures included the Health Opinion Survey (HOS) and the Cornell Medical Index (CMI). The HOS was a 20-item symptom list while the CMI included 195 items used to aid in medical diagnosis and symptom identification. Two subsamples were examined (n=147 and 146) and separate regression equations developed. For Subsample 1, SUs, pull-ups, body weight, a CMI subscale, and squat jumps produced a multiple correlation of 0.54 with attrition. For Subsample 2, SUs, pull-ups, body weight, a CMI subscale and age produced a multiple correlation of 0.47 with attrition. Cross-validation using the equation of the opposite subgroups produced multiple correlations of 0.37 and 0.40 for Subgroups 1 and 2, respectively.

Several studies have examined associations between fitness and attrition in Army BCT. Findings from these studies are not totally consistent. Some studies suggest that low physical fitness is associated with attrition (77,145,210,220,238,254) but other investigations have found mixed results (40,162,247). It may be possible to resolve these differences.

There are 7 studies that show low fitness is associated with attrition in Army BCT. One study (145) found that men and women who scored at or below the 25th percentile on any of the APFT events at entry were 1.9 to 3.3 times more likely to be discharged than those scoring at or above the 75th percentile. There was a dose-response showing that lower fitness was systematically associated with higher discharge rates. Fitness was independently associated with discharge when race, educational level, martial status and injuries in basic training were considered in a multivariate analysis. An Australian study (210) demonstrated that the least aerobically fit basic trainees (based on a progressive 20-meter aerobic shuttle run) were about 6 times less likely to complete training than trainees of average fitness. A study on Army infantry basic trainees (238) showed that men with lower performance on any one of the three APFT events in infantry basic training were at 4.1 to 8.0 times higher risk of discharge. A study cited in a 1998 GAO report (77) indicated that those who failed the initial Marine physical fitness test were 1.8 times more likely to be discharged than those who passed the test (24% vs. 13%). Army basic training data from Ft Jackson South Carolina in 2003 shows that those who fail their first APFT were 2.2 times more likely to be discharged (7.3 vs. 3.3%) (Knapik, 2003, Unpublished data from reference number 148). In the British Army, attrition was found to be strongly associated with VO_2max predicted from a 2.4 km run (220). A secondary analysis from the VanNostrand et al. study (254) showed that individuals who had low scores on the IDL at the MEPS were more likely to drop out of BCT. Table 35 shows attrition among men and women who could and could not lift the weight required by their MOS category. The only exception is the female very heavy category.

Table 35. Proportion of Men and Women Discharged in BCT Based on MOS Lifting Category and Whether or Not Lifted Occasional Load for MOS Category (From Reference 254)

		Could Lifted (% Attrition)	Could Not Lift (% Attrition)	Risk Ratio (Could Not Lift/Could Lifted)
Men	Light	[a]	[a]	---
	Medium	[a]	[a]	---
	Moderately Heavy	9.0	18.2	2.0
	Heavy	10.5	13.1	1.2
	Very Heavy	10.0	13.8	1.4
Women	Light	[a]	[a]	---
	Medium	15.2	22.0	1.4
	Moderately Heavy	13.0	15.7	1.2
	Heavy	14.3	18.2	1.3
	Very Heavy	15.9	15.9	1.0

[a]No comparison possible because all could lift required load

Not only is lower fitness on entry associated with higher attrition in basic training but Army trainees who have great difficulty achieving the basic training Army Physical Fitness Test (APFT) graduation standards have higher attrition later in service. A study (152) examined individuals who did not pass the APFT by the end of basic training and were sent to a special program where they worked exclusively on their fitness (the APFT Enhancement Program). The final BCT graduation rates of these individuals were lower than recruits who graduated with their peers. One-year attrition for men in the special program was 26% compared to 8% for men who did not have to enter the special program (i.e., passed their final APFT); 1-year attrition for women in the special program was 37% compared to 16% for women who did not have to enter the program.

Four studies (40,162,198,247) on associations between attrition and fitness have found mixed results. The first study by Tate (247) examined the association of 6-month attrition and PU performance for men; for women, attrition and an index composed of PUs and flexed arm hang was studied but it is not clear how this index was developed. Unfortunately, other APFT measures were not investigated. For men, PUs were associated with all separations from the Army but the strength of the relationship was considerably diminished when examining non-medial separations suggesting the association was stronger for medically-related separations. For women the PU/flexed arm hang index was not associated with 6-month attrition.

The second study showing mixed results was that of Kowal et al. (162) who examined the influence of a number of physical fitness factors (cardiorespiratory endurance, body composition, muscle strength) on discharges from basic training. The analysis of each fitness variable alone was not shown but discriminate function analysis indicated that for men, self perception of fitness (fitness rated on 5 point scale) discriminated between men who were and were not discharged. Self-perception of fitness, isometric trunk strength and isometric leg strength discriminated between women who were and were not discharged. VO_2max, predicted from a progressive step test, did not independently discriminate between those discharged and those not discharged.

A third study by Chin et al. (40) examined associations between attrition from Air Force basic training and passing the Air Force cycle ergometry test and/or two-mile run times. The Air Force cycle test estimated VO_2max on the basis of changes in heart rate to specific power outputs. The basic training attrition rate (which included discharges, medical holds and recycles) was 13%. Failure on the cycle test based on Air Force standards (indicative of very low fitness) was not related to attrition ($p=0.72$). However, secondary analysis indicated the small sample size ($n=50$ men and 50 women) and low attrition rate resulted in a statistical power of only 14%. Chin et al. (40) also examined associations between attrition and 2-mile run times. Those completing basic training had faster 2-mile run times than those not completing basic training (22.5 ± 0.4 vs. 21.3 ± 1.4 minutes) but the 1.2 minute difference was reported as not statistically significant. Calculation of average run times was not the appropriate statistical technique to determine attrition risk; chi-square statistics should have been used. Unfortunately, Chin et al. did not provide sufficient information for a secondary calculation of risk of discharge based on lower fitness levels.

The fourth and final study by Myers et al. (198) examined the relationship between attrition and physical fitness. The fitness tests included hand grip, 38-cm upright pull, predicted VO_2max (bicycle and step test) in addition to PUs, SUs and a 1- or 2-mile run. They collected medical data including sick call visits and days of restricted duty but it is not clear how this was done and they noted that the medical and attrition data provided to them was very incomplete. In correlational analysis, they found little relationship between discharge data and the physical performance tests ($r=-0.14$ to 0.12). Correlational analysis was an inappropriate statistical technique in this case. Attrition scores would have been very restrictive considering there are only 2 values (attrited or not). A more appropriate statistical technique would have been logistic regression so the odds of attrition could be related to the fitness measures.

Most studies that have used performance tests of cardiorespiratory endurance or muscular strength/endurance (77,145,210,238, Knapik, unpublished data) have demonstrated that those scoring higher have less attrition. The exceptions (40,162,198,247) have methodological problems cited above. Besides these methodological problems another possible explanation for the discrepancy in two cases (40,162) may involve a distinction between performance tests and predicted physiological tests. A performance test can be defined as an evaluation that requires a particular fitness component or a number of fitness components and is related to the accomplishment of a specific task under the volitional control of the individual (e.g., 2-mile run, PU). A physiological test can be defined as a task that measures a specific physiological capability or condition (VO_2max measuring cardiorespiratory endurance, or densitometry to measure body composition). Individuals who perform well on performance tests of fitness may indicate both a higher level of physical capacity and a higher level of motivation. The higher fitness level eases their effort in performing physical tasks while their higher motivation helps them complete what they start. This may be a partial explanation of the association between fitness and attrition.

Two studies that did not show a relationship between cardiorespiratory endurance and attrition used heart rate to predict aerobic capacity just prior to basic training (40,162). Heart rate can be elevated by stress, especially in new situations (90,175,245) and new trainees are under considerable initial stress in basic training (202). The bicycle test used by Chin et al. can generate higher heart rates among non-cyclists than field or treadmill tests at the similar power outputs (13). An elevated heart rate at a set work load on predictive VO_2max tests is an indicator of lower fitness and thus if a new trainee has an elevated heart rate due to conditions other than his/her fitness level, that trainee may be incorrectly classified (40). Further, the use of heart rate to predict VO_2max can be subject to errors, as great as 30% among the very fit, probably due to the asymptotic nature of the relationship between heart rate and VO_2 (185).

In summary, most studies that have examined associations between attrition and physical fitness have demonstrated very strong relationships between low aerobic fitness or low muscular endurance and higher attrition risk. Those studies that have not shown relationships have used inappropriate statistical techniques, and/or have used heart rate to predict aerobic capacity. The use of heart rate may be subject to measurement errors and inappropriate elevation due to stress. Performance measures of fitness may reflect both motivation and physiological capability and may be more appropriate tests to use when attrition is of interest.

d. Attributable Risk and Associations Among Fitness, Injuries and Attrition

In an attempt to more fully explore associations between attrition, physical fitness, injury and educational status we calculated attributable risk of discharge and injuries from past studies. In this case, attributable risk is that proportion of discharge or injury that could be ascribed to a component of physical fitness (131). For example, if we compare the most fit 25% to the least fit 25% we can estimate the reduction in discharge or injury risk if the least fit reached the level of the most fit. One weakness with this analysis is that the relationship has to be "causal" (low fitness has to cause higher discharge or injury, not just be associated with it in some unspecified manner) and this has not been demonstrated for the risk factors we will discuss (3). However, this analysis can provide some insight into the relative strength of particular risk factors.

Tables 36 and 37 show the attributable risk of discharge for men and women, respectively, from one BCT study (145); Tables 38 and 39 show the attributable risk of discharge for men and women, respectively, from another BCT study (148). For men, injury accounted for the largest proportion of the attrition risk. PUs, SUs and the 2-mile run also accounted for appreciable proportions of the attrition risk. Holding a GED (as opposed to a high school diploma) had attributable risk similar to that of the fitness measures. For women, more of the attrition risk could be attributed to the fitness measures, especially the 2-mile run, than could be attributed to injury or the GED. This may not be surprising since women have less fitness relative to men and BCT will be more physically taxing for women.

Table 36. Attributable risk (AR) of Discharge by Selected Fitness Risk Factors for Men (From Reference Number 145)

Risk Factor	Relative Risk of Discharge (95% Confidence Interval)	Prevalence of Risk Factor Among Discharged	Attributable Risk of Discharge
Lowest performance quartile, first diagnostic APFT run (19.18-31.58 minutes/2 miles)	1.67 (1.07-2.62)	0.36	0.14
Lowest performance quartile, initial APFT pushups (0-22 reps/2 minutes)	2.22 (1.46-3.37)	0.44	0.24
Lowest performance quartile, initial APFT situps (0-32 reps/2 minutes)	1.84 (1.20-2.82)	0.41	0.19
Highest quartile of body mass index (26.81-38.12 m/kg^2)	1.02 (0.66-1.59)	0.26	0.01
General Educational Development (GED)	1.82 (1.16-2.86)	0.37	0.17
Injury (one or more) during basic training	3.30 (2.20-4.96)	0.66	0.46

Table 37. Attributable risk (AR) of Discharge by Selected Fitness Risk Factors for Women (From Reference Number 145)

Risk Factor	Relative Risk of Discharge (95% Confidence Interval)	Prevalence of Risk Factor Among Discharged	Attributable Risk of Discharge
Lowest performance quartile, initial APFT run (23.49-28.68 minutes/2 miles)	2.27 (1.49-3.46)	0.43	0.24
Lowest performance quartile, initial APFT pushups (0-2 reps/2 minutes)	1.79 (1.16-2.76)	0.43	0.19
Lowest performance quartile, initial APFT situps (0-22 reps/2 minutes)	1.70 (1.09-2.64)	0.37	0.15
Highest quartile of body mass index (25.02-33.21 m/kg^2)	1.55 (1.09-2.21)	0.34	0.12
General Educational Development (GED)	2.15 (1.27-3.64)	0.18	0.10
Injury (one or more) during basic training	1.17 (0.82-1.68)	0.67	0.10

Table 37. Attributable risk (AR) of Discharge by Selected Fitness Risk Factors for Men (Previously Unpublished Data From Reference Number 148)

Risk Factor	Relative Risk of Discharge (95% Confidence Interval)	Prevalence of Risk Factor Among Discharged	Attributable Risk of Discharge
Lowest performance quartile, first diagnostic APFT run (minutes/2 miles)	1.60	0.35	0.13
Lowest performance quartile, initial APFT pushups (reps/2 minutes)	1.19	0.26	0.04
Lowest performance quartile, initial APFT situps (reps/2 minutes)	1.25	0.29	0.06
Highest quartile of body mass index (m/kg^2)	1.44	0.32	0.10
Injury (one or more) during basic training	2.05	0.43	0.22

Table 38. Attributable risk (AR) of Discharge by Selected Fitness Risk Factors for Women (Previously Unpublished Data From Reference Number 148)

Risk Factor	Relative Risk of Discharge (95% Confidence Interval)	Prevalence of Risk Factor Among Discharged	Attributable Risk of Discharge
Lowest performance quartile, first diagnostic APFT run (minutes/2 miles)	2.10	0.41	0.21
Lowest performance quartile, initial APFT pushups (reps/2 minutes)	1.42	0.32	0.09
Lowest performance quartile, initial APFT situps (reps/2 minutes)	1.75	0.31	0.13
Highest quartile of body mass index (m/kg^2)	1.36	0.31	0.08
Injury (one or more) during basic training	1.30	0.57	0.14

Table 39 shows attributable risk, relative risk, and prevalence of injury for low aerobic fitness and high BMI among basic trainees. This comparison again emphasizes that fitness has a much higher attributable risk of injury among men than among women.

Table 39. Attributable Risk, Relative Risk, and Prevalence of Injury for Selected Physical Fitness Risk Factors (From Reference Number 137)

Gender	Risk Factor	Attributable Risk of Injury	Relative Risk of Injury (95% CI[a])	Prevalence of Risk Factor Among Injured
Men	Lowest Performance Quartile, First APFT Run	0.26	2.5 (1.2-5.5)	0.44
	Highest Quartile of Body Mass Index	0.21	2.4 (1.4-4.0)	0.35
Women	Lowest Performance Quartile, First APFT Run	0.07	1.3 (0.9-1.6)	0.31
	Highest Quartile of Body Mass Index	0.08	1.4 (1.0-1.8)	0.31

[a]CI=Confidence Interval

e. Considerations in Selecting Physical Fitness Tests

The civilian and military job performance studies reviewed above all sought to validate physical fitness tests for use in selecting or screening individuals seeking civilian employment or entrance to military service. In the

United States, physical job performance and other types of employment tests must meet Equal Employment Opportunity Commission (EEOC) guidelines on fairness for test design and implementation as well as scientific standards for validity and reliability (114). In Canada and the European countries guidelines similar to those of the U.S. EEOC must also be met. According to the EEOC guidelines selection standards that may have an adverse impact on a group within society, such as women, must be (114):

1. Based on a job analysis
2. Indicative of the ability to perform critical job-related duties or tasks
3. Scientifically valid and reliable

Criteria have been established for determination of the validity of tests aimed at predicting job performance (61). In addition to the validity and reliability, tests to predict job performance must be practical (feasible to administer and measure) and affordable. Many physical fitness test validation studies experienced problems demonstrating one type of validity or another for a variety of reasons. Several criteria for validity apply to job selection tests. These include content validity, criterion-related validity and construct validity.

Content validity applies to job performance tasks. Content validity exists when the job performance tests are actual work tasks or close simulations of such tasks.

Criterion-related validity refers to the accuracy with which physical fitness, cognitive, psychological or other such tests estimate or predict the ability to perform an identified job-related task. The validity of such predictive tests is determined by the degree of correlation or association of the predictor tests with the actual job performance tasks. Simple or multiple correlation coefficients are generally used to express the relationship of job performance measures that can be measured on a continuous scale, such as the time to complete a critical job task or the total weight lifted from one position to another, with continuous physical test measures, such as VO_2max or upright pull isometric strength. For job performance measures that are dichotomous (i.e., metrics with only two outcomes), such as pass and fail or injured and not injured, different statistical tests measuring degree of association are necessary (e.g., chi square tests, logistic regression, or survival analysis).

Construct validity exists when it can be shown that a characteristic or set of characteristics, the construct, is associated with the ability to perform essential job tasks. Factor analysis is frequently used to identify clusters of characteristics or measures that are related to the ability to perform a job task.

(1) Examples of Problems with Previous Validation Studies

A number of the job-related physical fitness studies previously cited illustrated the types of problems that can adversely affect the demonstration of test validity, reliability or otherwise hinder test acceptance.

(a) Lack of Content Validity. Few examples of lack of content validity could be found in the studies reviewed. Most of the military studies discussed did not have a problem with content validity since they were based on detailed job analyses (198,199,217) or identification of common tasks that all soldiers need to perform ([Lee, 1992 #2313,253,255,256,260). One of the military studies, the diver study by Marcinik et al. (179) exhibited a form of the problem with content validity. The study attempted to use an existing Navy physical fitness test to predict job performance of diving tasks. The fitness test lacked job performance criterion validity because criterion job tasks were not established before the attempt to apply the fitness tests. Four of the five fitness tests were dry land activities (i.e., push-ups, sit-ups, pull-ups and 1.5 mile run). The fifth test was a 500 yard swim. On the other hand three of the four criterion job performance tasks involved swimming or other underwater activities. The fitness tests and the job performance criteria were poorly correlated. This emphasizes the importance of establishing job performance tasks before selecting fitness tests.

(b) Lack of Job Criterion-Related Validity. Several studies had difficulty demonstrating criteria-related validity. Surprisingly, Lee et al. (169) reported a very low (multiple $r = 0.18$ and $r^2 = 0.03$) and non-significant correlation between a criterion task involving a 16 kilometer march (25 kg load) and several physical fitness predictors that included a measure of cardiorespiratory endurance ($V0_2$ max), and a number of measures of upper and lower body anaerobic power. This failure to demonstrate a correlation was most likely a problem with the measurement used to assess the criterion (dependent) variable, the 16 km loaded march and the statistical analysis used to determine correlation. Distance marched was the criterion measure of performance but 68 of the 88 soldiers studied (77%) completed the entire 16 km distance. The criterion task thus amounted to what was really a pass or fail test rather than a test with a continuous number. As a consequence the statistical analysis of correlation with multiple regression was not appropriate. The preferred statistical test would have been a logistic regression using pass or failure as the dependent measure.

(c) Lack of Construct Validity. While factor analysis is frequently used to establish construct validity of the selection tests for physical fitness such as Rayson et al. (217) did in validating the British Army fitness tests, some studies the lack of construct validity is evident without such formal testing. The study of naval divers (179) above provides an example of such an obvious lack of content validity of the fitness tests with the job performance criterion tasks. In this study all but one of the job performance tasks involved swimming (e.g., swimming 200 ft with a 10 kg weight, treading water) or handling equipment in water (e.g., carrying scuba tanks 450 m or pulling 100 feet of "umbilical" line up 50 feet). On the other hand, only one of the five fitness tests involved swimming, a 500 yd swim, the others were typical dry land muscle endurance (push-ups, sit-ups, and pull-ups) and aerobic (1.5 mile- run) tests. Not surprisingly, the authors indicated regression results of the fitness test scores were not predictive of the job performance tasks.

(d) Lack of Reliability of Criterion Measures or Physical Fitness Measures. Rayson et al. (217) reported that one of the four criterion military job performance tasks selected for the British Army, a jerry can carrying task was unreliable. The task involved carrying two jerry cans, one in each hand, and each weighing 20 kg, up and down a 30 meter course at constant pace of 1.5 m/sec (3.3 mph) for as long as possible- until the jerry cans could not be held, the pace could not be maintained or the soldier voluntarily stopped. The test measure was duration (time) of carrying in seconds. Performance on a repeat of the carry tasks was reported to be 17% lower than the first trial. The design of this task poses several problems that could possibly account for this lack of repeatability. The jerry can weights were close to the maximal hand grip strength and total lifting capability for the women and keeping up with an external pacer.

(e) Failure to Evaluate Other Criterion Measures. Many investigators simply do not examine other indicators of job performance, such as injury or attrition or other failure rates (38,169,242,260). While some studies reviewed collected data on injuries and attrition (198,222), they did not utilize them in making decisions about selecting fitness tests. Myers commented that the correlations of injury and discharge with physical fitness measures were significant, but too small (-.14 to -.21) to be of practical concern and dismissed them for consideration in selecting tests. This typifies a wide spread problem with this kind of research – the reliance on statistical tests that assume that the predicted (dependent) and predictor (independent) variables are continuous. For dichotomous variables such as being injured or not injured or discharged or retained in service the more appropriate statistical tests of association are chi squares, logistic regression, survival analysis or similar analytic tools for dichotomous outcomes. A more subtle problem occurs when predicted variable appears to be continuous such as time or distance to complete a march actual has a limit that many or most of the subjects can achieve. In such instances the test is actually a pass/fail test and should be analyzed with statistics appropriate for dichotomous outcome (predicted, dependent) variables.

(f) Impractical, Overly Complicated or Unaffordable Fitness Test Procedures. Sharp et al. (229) employed an in-cadence stair stepping test to predict VO_2 max from heart rate. However, in their recommendations they excluded testing for stamina because of the expense and because too many personnel would be required to administer the test. The test was complex requiring stepping in time to a metronome at several rates and requiring the measurement and recording of heart rate after each successively more strenuous stepping rates. Thus, personnel, cost and complexity were all issues with this test.

(2) Criteria for Selecting Physical Fitness Tests for the Army

Government regulations/guidelines, scientific standards and administrative considerations all factor into the selection of physical fitness screening for the military. From the preceding discussion a number of criteria can be recommended in selecting fitness tests for use by the Army as follows:

Validity
- Physiologic/scientific validity
- Job tasks performance validity
- Construct validity

Reliability/Repeatability

Non-Discriminatory in Nature

Association with Occupational Indicators
- Job performance
- Injury risks
- Attrition/job failure risks

Administrative Practicality
- Ease of administration
- Reasonable personnel requirements
- Low cost
- Short time to conduct
- Low/minimal health risk
- Easy to standardize
- Equipment readily available

Each physical fitness test can be assessed on these criteria. One approach is to score each test under consideration using these criteria, and then rank them based on their scores. This would result in a more objective process.

6. RECOMMENDATIONS FOR AN ENTRY-LEVEL PHYSICAL FITNESS TEST. Our review of the literature on physical fitness testing and pre-employment/pre-enlistment screening suggests that there are at least 3 courses of action. The rationale for each is discussed below.

a. Courses of Action 1 – Keep Current Entrance Criteria

Course of Action 1 (COA1) is to keep the current Reception Station Physical Fitness Test but move testing to the MEPS or to the recruiter. The Reception Station Physical Fitness Test was described in the Introduction and consists of PUs, SUs and a 1-mile run. It has several advantages. It requires a minimum amount of equipment. It is understood by recruiters and military personnel in the MEPS because the same test items are administered as part of the biannual APFT taken by all Army personnel. No further training of recruiters or MEPS personnel is required. All three tests items have been shown to be related to injuries and attrition although the association between SUs and injuries is not as strong as that between injuries and the other two tests (133,135,144,145,158,223,238). There are some associations with some job-related military tasks (242,243,253,256). However, studies examining relationships between PUs, SUs or a 1-mile run and military job performance included the fitness test items as part of a multiple regression equation and the relationship of the single fitness test item were not included. COA1 measures muscular endurance and cardiorespiratory endurance but does not measure muscle strength.

Standards have been established for entry into BCT and these standards are shown in the Introduction. We evaluated the relationship between the fitness criteria in Table 1 to injuries, on-time completion of training, and discharges. To do this we examined existing data from a previous study (148) in which individuals took the entry-level physical fitness test but then began basic training regardless of whether or not they passed the test. We tracked injuries, discharges and attrition from training for any reason (newstarting or discharge). Table 40 shows the results. Individuals not passing the entry-level physical fitness test were more likely to experience an injury of any type, more likely to have a serious injury that removed them from training, less likely to complete BCT in 9 weeks, and more likely to be discharged. These data suggest that if the entry-level physical fitness test is administered in the same way as in the reception station, and the cutpoints remain as in Table 1, the test is likely to discriminate between those who do and do not get injured or complete BCT.

Table 40. Comparison of Recruits Passing and Not Passing the Entry-Level Physical Fitness Test

		Any Injury	Serious Injury (Removal from Training)	Do Not Complete Training With Peers (9 Weeks)	Discharged
Men	Passed ELPFT[a] Test (%)	20.1	1.6	12.9	6.6
	Did Not Pass ELPFT[a] Test (%)	37.5	6.3	40.6	18.8
	p-value	0.02	<0.01	<0.01	<0.01
	Risk Ratio (Not Pass/Pass)	1.9	3.9	3.2	2.8
Women	Passed ELPFT[a] Test (%)	39.8	5.5	22.2	11.8
	Did Not Pass ELPFT[a] Test (%)	64.4	8.2	47.9	21.9
	p-value	<0.01	0.07	<0.01	0.01
	Risk Ratio (Not Pass/Pass)	1.6	1.5	2.2	1.9

[a]ELPFT=Entry-Level Physical Fitness Test

b. Course of Action 2 – Recommendations Based on the Literature

Course of Action 2 (COA2) suggests a physical fitness test battery based on findings in the literature. Two assumptions are made: a) that the major components of physical fitness should be measured, and b) that the fitness tests should be related to criterion measures like job performance, attrition and/or injury.

Ideally, a general test of physical fitness would measure all the fitness components described in Table 3. However, few studies have validated tests of coordination, balance or flexibility against criterion measures that might be related to critical aspects of job performance (10,115) or other factors that might be of interest from a military perspective. Most studies have concentrated on tests of muscle strength, muscular endurance, cardiorespiratory endurance and/or body composition and have shown that these fitness measures are associated with various aspects of job performance, injuries, and work/training attrition (10,11,14,15,22,23,38,56,77,83,91,115,145,152,162,169,198,210,218, 220,222,225,227,229,234,238,242,243,248,254,267,271). Thus, the literature provides guidance for testing these components of physical fitness. Body fat limits have already been established for entry to service (8) so any measure of body composition based on selected criteria would have to be reconciled with this existing requirement. Because of these factors we limited COA2 to a consideration of tests of muscular strength, muscular endurance and cardiorespiratory endurance.

Muscle strength tests that have been repeatedly demonstrated to be related to simulated military performance tasks include isometric back extension or flexion (38,218,222,234,252,253), the IDL (14,15,218,222,248,254) and the isometric 38-cm upright pull (218,222,229,248). The only study on the IDL and injury did not show an association (48) but another did show some relationship between attrition and low IDL performance (254).

Muscular endurance tests have not been included in US Army validation studies involving job performance because of an assumption made early in the validation process that there was a close relationship between absolute muscular strength and absolute muscular endurance (260). In general, it is the case that stronger individuals also tend to have greater absolute muscular endurance (34,251). Some foreign military studies that have included muscular endurance tests in military task validation have found relationships with PU performance (242,243,253,256), pull-up performance (218), SU performance (253) and dynamic shoulder extension endurance (38,234). Much more work has been done relating muscular endurance tests to injuries and attrition. Tests related to injuries and attrition include PUs (135,145,158,238), SUs (144,223,238), and pull-ups (77). Relationships between injuries and PUs are more consistent than between SUs and injuries, at least in BCT (135,145,158,238). Few women can perform pull-ups (41,46) so an alternate test like the flexed arm hang would be required and tests would differ for men and women. Based on these considerations, the most appropriate muscular endurance test appears to be PUs. PUs have the most consistent relationship with injuries and attrition; the relationship with job performance appears to be weaker but some relationships have been established.

Studies that have examined cardiorespiratory endurance have shown associations between job performance and tests involving VO_2max prediction from aerobic shuttle runs (210,217,220), step tests (229,248), and bicycle ergometer tests (248). In addition, attrition and injuries have been associated with maximal effort runs at distances of 1 mile (133) 1.5 miles (219,220), 3000 meters (105), 2 miles (135,144,145,158,223,238) and 12-min (253,255) For simplicity, and because virtually all of these tests are associated with the physiological criterion measure of VO_2max (see Tables 6, 7 and 8), it would seem that any of these tests would be appropriate for a test of cardiorespiratory endurance. If space is limited or there is no access to a track, the innovative step test mentioned earlier could be used if it could be validated.

Table 41 shows the best options for COA2 based on the literature. The largest amount of support is for a test that incorporates the IDL, PUs and 2-mile run. However, a 1-mile run would be sufficient to evaluate cardiorespiratory fitness. A 1-mile run also decreases the possibility of injury, compared to a 2-mile run, since longer running distances have been associated with higher injury risk (136,161,181). A shorter run may also be less stressful for individuals who are not accustomed to prolonged maximal efforts of this sort.

Table 41. Options for a Pre-Accession Physical Fitness Tests Based on Military Task Performance, Injuries and Attrition from Service in the Literature

Muscle Strength	Muscular Endurance	Cardiorespiratory Endurance
Dynamic IDL	PUs	1-Mile Run
Isometric 38-cm Upright Pull	Pull-ups	1.5-Mile Run
Isometric Back Extension	Dynamic Shoulder Endurance	2-mile Run
Isometric Back Flexion	Pull-ups	Aerobic Shuttle Run
		Innovative Step Test

COA2 recommends a test incorporating the IDL, PUs and a 1-mile run. The innovative step test could replace the 1-mile run but tests of validity and reliability would have to be conducted first. The passing criteria for PU and the 1-mile run remain the same as in COA1. The criteria for passing the IDL are based on MOS as shown in Table 29. For MOS that have light, medium or moderate lifting requirements as defined in AR 611-201 (7), the requirement is to lift 40 lbs. For MOS having heavy or very heavy lifting requirements as defined in AR 611-201 (7), the requirement is to lift 70 lbs.

c. Course of Action 3 – Determine and Validate A Physical Fitness Test

We recommend Course of Action 3 (COA3) as the most comprehensive, rational, and legally defensible. It complies with the EEOC guidelines on employee selection procedures and takes advantage of information and techniques garnered from past military and civilian studies on pre-employment testing. COA3 involves 6 steps: 1) determining a set of critical military criteria (i.e., job performance, attrition, injury, NCO ratings), 2) determining a battery of physical fitness tests that measure the fitness components associated with these criteria, 3) obtaining performance data on a representative sample of soldiers, 4) validating and cross-validating the fitness measures against the military criteria, 5) selection of fitness test scores that represent acceptable performance on the criteria, and 6) periodic re-evaluation of the fitness tests to account for technological changes in equipment and material and for changes in the level of physical fitness of potential military recruits.

(1) Selection of Military Criteria

The first step in COA3 is to select the criteria against which to validate the fitness tests. This should be determined by a panel of military leaders in conjunction with individuals who will be performing the testing to assure that the critical criteria are chosen and that these criteria are measurable and understandable. Examples of criteria might be specific types of job performances, attrition from training, injuries, and/or NCO ratings.

Job performance is a commonly used criteria in the literature and one that is specifically mandated in the EEOC guidelines (61). The first step in determining job performance measures is a job analysis. There are several examples of job analyses in the literature that involve military (14,199,216,234,260) and civilian groups (10,11,83,130) that have been reviewed here. A job analysis was conducted previously in the US Army (9) but this analysis is at least 20 years old and the pace of technological changes dictates that a current job analysis should be performed. Job analysis involves the systematic collection of information to describe the tasks that are involved in the job. Procedures involved in the job analysis include general information gathering to guide more detailed investigation, then surveys, interviews, observation, and physical measurements (216). For general information gathering useful documents identified in the literature include soldier training publications (STPs) and Army Occupational Surveys (also called Army Data Analysis Requirements and Structure Program) (154). Surveys and interviews could be conducted with subject matter experts who actually perform the jobs. Observation and physical measurements may be necessary to quantify the physical demands of the task (216). Since the interest here is in the physical dimensions of the job, the physical activities would be emphasized. Once the assumed physically demanding tasks have been identified, they should be verified with the people actually performing the jobs to assure the correct tasks have been selected.

Past studies on the US Army, the British Army and the Canadian Forces have suggested that the wide variety of specific tasks in various MOS can be reduced to a relatively small number of general or critical tasks that are common to many MOS. These tasks have included single lifts to specific heights, repetitive lifting, pushing, pulling, lifting and carrying, road marching, and casualty evacuation (14,221,234,242,260). Studies in the civilian sector have also found that a range of complex jobs can also be reduced to some simple or critical tasks (10,11,271). The selected tasks would involve continuous measures and should require maximal performances so limiting physical fitness factors can be appropriately identified.

In addition to criterion job-related task performance, other appropriate criteria might include injuries and attrition from service. Injury data can be obtained by direct screening of medical records or from injury data collected by the Defense Medical Surveillance System (DMSS) and there are several examples in the literature of how this can be done (133,135,148,158). Attrition from service can also be collected from the DMSS or directly from unit records as has been done in past studies (103,151,152,158).

(2) Selection of Fitness Tests

Once the criterion measures are selected the fitness tests can then be determined. This selection would be based on established or logically based assumptions regarding which fitness components (Table 3) are related to the criterion tasks. If the criterion is a job performance task, that task could be broken down into individual activities and the components of fitness required for those activities could be identified. As an example, consider a soldier required to load a small truck with boxes over a 15 minute period. Task elements might include obtaining boxes from a central location, placing them on a cart, pushing the cart to the truck and lifting the boxes into a truck. This would require back, arm and leg muscular strength and muscular endurance (to lift the boxes and push the cart) as well as cardiorespiratory endurance (to sustain the work rate). Another common task performed by many soldiers is casualty evacuation over a short distance. This task can be broken down into activities involving lifting the casualty onto the litter, lifting the litter, carrying the litter and lowering the litter. Important fitness components might include upper torso and back strength (to get the casualty onto the litter), hand grip strength or endurance (to hold and carry the litter), and muscular endurance of the upper body and legs (to transport the litter). Many authors have provided appropriate tests for different components of physical fitness (69,70,126) and many of those tests have been reviewed here.

Selection of appropriate fitness tests that might be related to injury and attrition can be guided by the literature. Many past investigations reviewed here demonstrate that a number physical fitness components are related to injury.

(3) Obtaining Soldier Data

The next step would be obtaining the data to validate the physical fitness tests. Teves et al. (248) and Rayson et al. (217,218) present paradigms that can be applied here. Recruits could be tested three times: prior to BCT (Phase 1), at the conclusion of BCT (Phase 2) and at the conclusion of AIT (Phase 3). In Phase 1, recruits entering BCT would be given the selected physical fitness tests. Strictly for testing purposes, it would be prudent to give these tests in the Reception Station prior to BCT rather than in the MEPS because it would be much easier to follow a recruit once he/she is assigned to a single BCT post rather than tracking recruits through several posts. In Phases 2 and 3, recruits would be administered the physical fitness battery and criterion job performance tasks. Giving the fitness battery a second and third time would provide a look at changes in the specific components of physical fitness measured. The criterion task performances during Phases 2 and 3 would be related to the fitness measures in Phase 1. Injuries during BCT and AIT would be tracked through the DMSS. Attrition from training (discharges from service and newstarting/recycling) would be tracked through records in the BCT and AIT units.

(4) Validation and Cross-Validation of the Fitness Tests

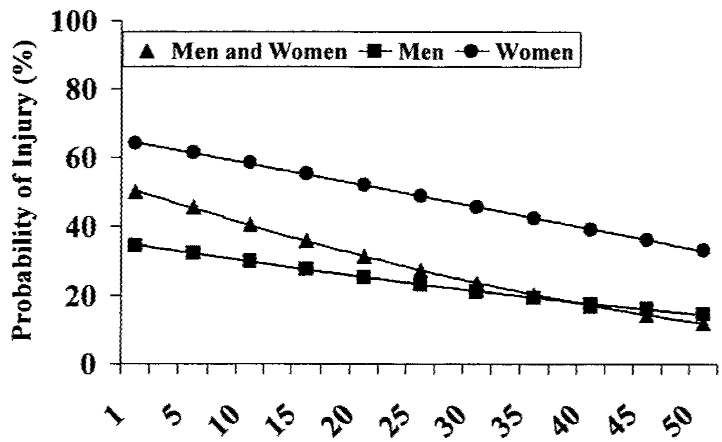
Once the data is collected the analysis can begin. Multiple regression would be the primary statistical tool used to determine the set of physical fitness measures most related to the criterion task performance. For dichotomous variables like injuries or attrition, logistic regression or survival analysis would be the primary statistical tools. The solider sample would be split into two for validation and cross-validation purposes. Predictive models derived on the validation sample would be tested on the cross-validation sample. The multiple correlation coefficient would describe the strength of the relationship between the fitness measures and the criterion tasks. The standard error of estimate would describe the error of prediction. Errors of prediction could also be calculated using the Bland/Altman Method (24).

(5) Determination of Fitness Criteria

Determination of pre-accession entry standards would be based on cut-scores that define who will be accessed into service and who will not. Cutpoints can be determined by a number of methods described by Gebhardt (79) and Hodgdon (114). For criterion that are continuous, a simple or multiple linear regression combined with an analysis of the prediction error might be appropriate. As an example, consider a criterion task that involves lifting and carrying a soldier on a litter 100 yards as rapidly as possible. The fitness measure might involve hand grip strength since this measure highly related to litter carriage performance (149). A simple linear regression can be used to describe the relationship between carriage time and grip strength. Assume the critical time to transport the soldier is 2 minutes. The grip strength associated with this time can be determined from a regression plot of hand grip strength and time. The standard error of estimate must also be considered since this defines the prediction error.

For models that involve dichotomous (pass/fail) criteria a logistic regression model is appropriate. For example, consider a criterion task that involves whether or not an individual is injured during BCT. The fitness test might involve an assessment of cardiorespiratory endurance such as the time on a 1-mile run. Once the logistic regression equation is developed, curves showing the probability of injury can be developed (114). Figure 3 shows an example of this using previously unpublished data from another study (148).

Figure 3. Probability of Injury Based on Number of Push-ups On Entry to Basic Combat Training



Another method described in detail by Gephardt (79) involves the use of expectancy and contingency tables. Expectancy tables show the relationship between a criterion measure (job performance, injuries, attrition) and a fitness score (i.e., 1-mile run) or set of fitness scores (from multiple regression or logistic regression equation). Test score distributions can be set in equal units (i.e., quintiles, deciles) or absolute units and the criterion task performance plotted against these. Figure 4 shows an example (using a graph rather than a table) using data from a previous study (148). In this example, the criterion measure is discharge from service during BCT and the physical performance task is a 1-mile run on entry to BCT. If the selection criteria is set at a run faster than 13.0 minutes, then 7% of recruits fail the test. The number of correctly classified people can be determined using a contingency table as shown in Table 42. The number of individuals correctly classified (with a 13-minute cutpoint) can be determined using the formula:

$$\text{Correct Decisions} = \frac{\text{True Passers \& No discharge} + \text{True Failures \& Discharges}}{\text{Entire Sample}}$$

In this case:

$$\text{Correct Decisions} = \frac{710+16}{852} = 85\%$$

From the contingency table it can be seen that 5% (44/852) of all individuals would be falsely rejected while 10% (82/852) would be falsely accepted. This is only an example with a single variable and multiple variables could be used in conjunction with a multiple regression equation. There are also other methods for determining cutpoints (79,114)

Figure 4. Association of 1-Mile Run Time with Discharge in BCT (Women)

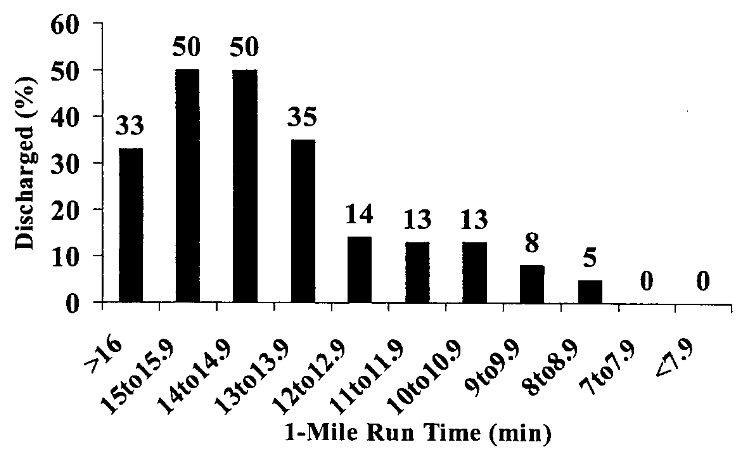


Table 42. Contingency Table for One-Mile Run and Discharge Status

Discharge Status	1-Mile Run	
	>13 min (failures)	<13 min (passers)
Not Discharged	False Rejection=44	True Acceptance=710
Discharged	True Rejection=16	False Acceptance=82

Another consideration is the fairness of the test. The EEOC has defined unfairness as a condition where the members of one race, sex or ethnic group typically obtain a lower score on a selection test and that test score does not reflect differences in criterion job performance (61,125). Fairness can be established statistically by constructing different regression equations for the subgroups of interest, comparing the standard errors of prediction (39) and testing subgroups for equality of regression slopes and intercepts (114,125,229). Using gender as an example, where it can be demonstrated that the gender-specific slopes are parallel and the intercepts are coincident gender-free models can be developed. Where the gender-specific slopes are parallel but the intercepts are not coincident, gender would have to be included as a variable in the model. Where gender-specific slopes and intercepts differ, separate gender-specific models would have to be developed (229). However, physiological interpretations of the data are also important and blind application of statistical principles can lead to misinterpretation of data (114). In addition to examination of slopes, residual variance of the two groups should be examined for heterogeneity (229).

(6) Periodic Re-evaluation

The pace of technological change and possible changes in the physical fitness of American youth dictates that periodic re-evaluation of the fitness tests should be performed. When the new job analysis is conducted the appropriateness of the criterion tasks would be determined. If necessary, the criterion tasks could be changed. Whether or not the criterion tasks are changed, a new sample of recruits should be tested to account for potential changes in the physical fitness level of these recruits. A review of the literature (159) suggests that some components of physical fitness have changed in as short a period as 20 years. For example, the VO_2max of male recruits has not changed while that of female recruits has improved from at least 1975 to 1998. Performance has declined on endurance running tasks in a similar time period. It may be that youth and recruits are not as proficient at applying their physiological capability to performance tasks like timed runs, possibly because of factors such as reduced experience with running, lower motivation, and/or environmental factors. Limited data on Army recruits demonstrate an increase in strength from 1978 to 1998. Data on muscular endurance is not consistent. There is strong evidence that body weight and body mass index (BMI) have increased, presumably due to an increase in caloric intake. Most physical fitness trends can be modeled using linear regression and there is little reason to think the trends cited above will not continue into the future (159). The other steps involved in the process would also have to be repeated (validation, cross-validation and determining cut scores).

7. SUMMARY. The CAR requested we review the literature on pre-enlistment physical fitness screening and recommend courses of action for a physical fitness test for pre-accession screening. We reviewed the literature on the concept of physical fitness to achieve a thorough understanding of the concept. We then reviewed the variety of tests that assess physical fitness. Civilian and military literature involving pre-employment testing was reviewed to understand previous work world-wide. We also reviewed the literature on associations between fitness and injuries and attrition for service. Our review found that measures of physical fitness components were associated with the performance of military tasks as well as attrition and injuries. Finally, courses of action for a pre-enlistment physical fitness test were suggested.

COA1 is to keep the current pre-accession test involving PUs, SUs, and a 1-mile run. Men could enter service if they could perform 13 PUs, 17 SUs and run a mile in 8.5 minutes. Women could enter service if they could perform 3 PUs, 17 SUs and run a mile in 10.5 minutes. These tests are related to injuries and attrition but the relationship to military job performance is weaker and the test battery lacks a test of muscular strength.

COA2 is based on studies performed in the literature that have examined job performance, attrition from service, and injuries. It involves an IDL, PUs and a 1-mile run. The passing criteria for PU and the 1-mile run remain the same as in COA1. The criteria for passing the IDL are based on MOS. For MOS that have light, medium or moderate lifting requirements as defined in AR 611-201, the requirement is to lift 40 lbs. For MOS having heavy or very heavy lifting requirements as defined in AR 611-201, the requirement is to lift 70 lbs. Individual test items in this test battery is related to job performance, injuries and attrition.

We recommended COA3 as the most rational, logical, and legally defensible. It complies with the EEOC guidelines on employee selection procedures and takes advantage of information and techniques garnered from past military and civilian studies on pre-employment/pre-enlistment testing. COA3 involves a research project encompassing 6 steps: 1) determining a set of critical military criteria (i.e., job performance, attrition, injury, NCO ratings), 2) determining a battery of physical fitness tests that measure the fitness components associated with these criteria, 3) obtaining performance data on a representative sample of soldiers, 4) validating and cross-validating the fitness measures against the military criteria, 5) selection of fitness test scores that predict acceptable performance on the criterion tasks, and 6) periodic re-evaluation of the criterion tasks and soldier sample. In the long term the Army will need COA3 since Army tasks, equipment, and personal characteristics change over time.

8. CONCLUSIONS.

This review has shown that physical fitness is strongly associated with job performance, injuries and attrition from service. The findings are reproducible across many studies and generally when contrary evidence is found there are problems with experimental design or statistical analysis. The attributable risk of injury and attrition, especially in men, is great enough to warrant routine pre-enlistment screening for physical fitness along with health/medical history and cognitive ability.

Several studies show that the current entry-level physical fitness test possesses some validity since individuals who do not pass the test are more likely to be injured or to attrite from service. However, the current physical fitness entrance test could be immediately improved by eliminating the SU and replacing it with the IDL. In the long term, an entry-level physical test should be developed through a comprehensive research program that involves well established methods of relating physical fitness tests to criterion measures important to the military like job performance, injuries, and attrition. A physical fitness test battery established from these research procedures would have a strong rational basis, be legally defensible, and would place testing of the physical capability of potential recruits on a footing similar to cognitive ability testing which has been performed since WWI (62).

APPENDIX A
References

1. President's Council on Physical Fitness and Sports. Physical Fitness Research Digest 1:1-3. 1971.
2. American College of Sports Medicine. Guidelines for Exercise Testing and Prescription (4th ed). Philadelphia: Lea and Febiger, 1991.
3. Ahlbom A and Norell S. Introduction to Modern Epidemiology. Chestnut Hill MA: Epidemiological Resources Inc, 1984.
4. Ahlborg G, Felig P, Hagenfeldt L, Hendler R and Wahren J. Substrate utilization during prolonged exercise in man. J Clin Invest 53:1080-1090. 1974.
5. Albanese CV, Diessel E and Genant HK. Clinical application of body composition measurements using DXA. J Clin Densit 6:75-85. 2003.
6. Anderson GS. A comparison of predictive tests of aerobic capacity. Can J Sports Sci 17:304-308. 1992.
7. Military Occupational Classification and Structure. Army Regulation (AR) 611-201. Washington DC: Headquarters, Department of the Army, 1994.
8. Standards of Medical Fitness. Washington DC: Department of the Army. Army Regulation No. 40-501, 2003.
9. Military Occupational Specialty (MOS) Physical Demand Analysis. Washington DC: Department of the Army, Deputy Chief of Staff for Personnel, 1982.
10. Arnold JD, Rauschenberger JM, Souble WG and R MG. Validation and utility of a strength test for selection of steelworkers. J Appl Psychol 67:588-604. 1982.
11. Arvey RD, Landon TE, Nutting SM and Maxwell SE. Development of physical ability tests for police officers: a construct validation approach. J Appl Psychol 77:996-1009. 1992.
12. Astrand PO and Rhyming I. A nonogram for calculations of aerobic capacity from pulse rates during submaximal work. J Appl Physiol 7:218-221. 1954.
13. Astrand PO and Rodahl K. Textbook of Work Physiology. New York: McGraw Hill, 1977.
14. Ayoub MM, Denardo JD, Smith JL, Bethea NJ, Lambert BK, Alley LR and Duran BS. Establishing physical criteria for assigning personnel to Air Force jobs. Lubbock TX: Institute for Ergonomics Research, University of Texas. Technical Report No. AFOSR-TR-83-0941, 1982.
15. Ayoub MM, Jiang BC, Smith JL, Selan JL and McDaniels JW. Establishing a physical criterion for assigning personnel to US Air Force jobs. Am Ind Hyg Assoc J 48:464-470. 1987.
16. Bar-Or B. The Wingate anaerobic test. An update on methodology, reliability and validity. Sports Med 4:381-394. 1987.
17. Bar-or O, Dotan R and Inbar O. A 30-sec all out ergometer test - its reliability and validity for anaerobic capacity. Israeli J Med Sci 13:326-327. 1977.
18. Barry AJ and Cureton TK. Factorial analysis of physique and performance in prepubescent boys. Res Quart 32:283-295. 1961.
19. Bass RI. An analysis of the components of tests of semicircular canal function and of static and dynamic balance. Res Quart 10:33-52. 1939.

20. Bassett DR and Howley ET. Limiting factors for maximal oxygen uptake and determinates of endurance performance. Med Sci Sports Exerc 32:70-84. 2000.
21. Baumgartner TA and Zuidemia MA. Factor analysis of physical fitness tests. Res Quart 43:443-450. 1972.
22. Bernauer EM and Bonanno J. Development of physical profiles for specific jobs. J Occ Med 17:27-33. 1975.
23. Bertina FM. Physical selection criteria for the Royal Netherlands Army. Brussels, Belgium: North Atlantic Treaty Organization. Technical Proceedings No. AC/243 (Panel 8) TP/13, 1997.
24. Bland JM and Altman DG. Statistical method for assessing agreement between two methods of clinical measurement. Lancet 1(8476):307-310. 1986.
25. Blazevich AJ, Gill N and Newton RU. Reliability and validity of two isometric squat tests. J Strength Condit Res 16:298-304. 2002.
26. Borchardt JW. A cluster analysis of static strength tests. Res Quart 39:258-264. 1968.
27. Bosco C, Komi PV, Tihanyi J, Fekete G and Apor P. Mechanical power test and fiber composition of human leg extensor muscles. Eur J Appl Physiol 51:129-135. 1983.
28. Bouchard C, Shephard RJ, Stephens T, Sutton JR and McPherson BD, in *Exercise Fitness and Health. A consensus of current knowledge*, edited by Bouchard C, Shephard RJ, Stephens T, Sutton JR, and McPherson BD (Human Kinetics, Champaign IL, 1990).
29. Brady JI, Knight DR and Berghage TE. Relationship between measures of body fat and gross motor proficiency. J Appl Psychol 62:224-229. 1977.
30. Brooks GA. Lactate doesn't necessarily cause fatigue: why are we suprised? J Physiol 536:1. 2001.
31. Brozek J, Grande F and Anderson JT. Densitometric analysis of body composition: revision of some quantative assumptions. Ann NY Acad Sci 110:113-140. 1963.
32. Burke EJ. Validity of selected laboratory and field tests of physical working capacity. Res Quart 47:95-104. 1976.
33. Buskirk ER, Underwater weighing and body density. A review of procedures in *Techniques for the Measurement of Body Composition*, edited by Brozek J and Henschel A (National Academy of Science Press, Washington DC, 1961).
34. Caldwell LS. Measurement of static muscle endurance. J Eng Psychol 3:16-22. 1964.
35. Carlson BR and Kroll W. The use of analysis of variance in estimating reliability of isometric elbow flexion strength. Res Quart 41:129-134. 1970.
36. Carpenter A. An analysis of the relationship of the factors of velocity, strength, and dead weight to athletic performance. Res Quart 12:34-39. 1941.
37. Caspersen CJ, Powell KE and Christenson GM. Physical activity, exercise and physical fitness: definitions, and distinctions for health-related research. Pub Health Rep 100:126-131. 1985.
38. Chahal HS. Army physical fitness performance standards based on body composition, muscular strength and endurance. Edmonton, Alberta: University of

Alberta, Department of Physical Education and Sports Studies. Doctoral Dissertation, 1993.
39. Chatterjee S and Price B. Regression Analysis by Example. New York: John Wiley &Sons, Inc., 1991.
40. Chin DL, Blackwood GV and Gackstetter GD. Ergometry as a predictor of Basic Military Training Success. Mil Med 161:75-77. 1996.
41. Clarke HH. Endurance of the arm and shoulder-girdle muscles. Physical Fitness Research Digest 8:1-2. 1978.
42. Clarke HH. Objective strength tests of affected muscle groups involved in orthopedic disabilities. Res Quart 19:118-147. 1948.
43. Coleman JW. The differential measurement of the speed factor in large muscle activity. Res Quart 8:123-130. 1937.
44. Cooper KH. A means of assessing maximal oxygen intake. Correlation between field and treadmill testing. JAMA 203:201-204. 1968.
45. Corbin CB, Dowell LJ, Lindsey R and Tolson H. Concepts in Physical Education. Dubuque, IA: W.C. Brown Company, 1983.
46. Cotten DJ and Marwitz B. Relationship between 2 flexed-arm hangs and pull-ups for college women. Res Quart 40:415-416. 1969.
47. Cousins GF. A factor analysis of selected wartime fitness tests. Res Quart 26:277-288. 1955.
48. Cowan D, Jones BH, Tomlinson JP, Robinson J, Polly D, Frykman P and Reynolds K. The epidemiology of physical training injuries in the U.S. Army infantry trainees: methodology, population and risk factors. United States Army Research Institute of Environmental Medicine, Natick MA. Technical Report No. T4/89, 1988.
49. Cumbee FZ. A factorial analysis of motor coordination. Res Quart 15:412-428. 1954.
50. Cumbee FZ and Harris CW. The composite criterion and its relation to factor analysis. Res Quart 24:127-134. 1953.
51. Cumbee FZ, Meyers M and Peterson G. Factorial analysis of motor coordination variables for third and fourth grade girls. Res Quart 28:100-108. 1957.
52. Cureton KJ, Effects of experimental alterations in excess weight on physiological responses to exercise and physical performance. In *Body Composition and Physical Performance. Applications for Military Services*, edited by Marriott BM and Grumstrup-Scott J (National Academy Press, Washington, D.C., 1992).
53. Daniels WL, Vogel JA and Kowal SM. Guidelines for aerobic fitness training in the US Army. Natick MA: US Army Research Institute of Environmental Medicine. Technical Report No. T5/79, 1979.
54. Darakjy S, Hauret KG, Canada SE, Knapik JJ, Wells J, Hodebecke EL, Kenyon M, Marin RE and Bullock SH. Injuries and injury risk factors among armor battalion soldiers at Ft Riley, Kansas. Med Sci Sports Exerc 35:S278. 2003.
55. Davis CTM. Limitations to the prediction of maximum oxygen intake from cardiac frequency measurements. J Appl Physiol 24:700-706. 1968.

56. Davis PO, Dotson C and Maria DLS. Relationship between simulated fire fighting tasks and physical performance measures. Med Sci Sports Exerc 14:65-71. 1982.
57. Dempster P and Aitkens S. A new air displacement method for the determination of human body composition. Med Sci Sports Exerc 27:1692-1697. 1995.
58. Dickinson RV. The specificity of flexibility. Res Quart 39:792-794. 1968.
59. Doolittle TL and Bigbee R. The twelve-minute run-walk: a test of cardiorespiratory fitness of adolescent boys. Res Quart 491-495:1968.
60. Doolittle TL, Dominic JAC and Doolittle J. The reliability of selected cardiorespiratory endurance field tests with adolescent female populations. Am Correct Ther J 23:135-138. 1969.
61. Equal Employment Opportunity Commission. Uniform Guidelines on employee selection procedures. Federal Register 43:(38289-28315). 1978.
62. Eitelberg MJ, Laurence JH, Waters BK and Perelman LS. Screening for service: aptitude and educational criteria for military entrance. Alexandria VA: Human Resources Research Organization. Technical Report No. FR-PRD-83-24, 1984.
63. Elbel ER. Relationship between leg strength, leg endurance and other body measurements. J Appl Physiol 2:197-207. 1949.
64. Farrell PA, Wilmore JH, Coyle EF, Billings JE and Costell DL. Plasma lactate accumulation and distance running performance. Med Sci Sports Exerc 11:338-344. 1979.
65. Faulkner JA, Claflin DR and McCully KK, Power output of fast and slow fibers from human skeletal muscles. In *Human Muscle Power*, edited by Jones NL, McCartney N, and McComas AJ (Human Kinetics, Champaign IL, 1986).
66. Fitts RH and Widrick JJ. Muscle mechanics: adaptations with exercise training. Med Sports Sci Rev 24:427-473. 1996.
67. Fleishman EA. Factor analysis of physical fitness tests. Educ Psychol Meas 23:647-651. 1963.
68. Fleishman EA. Relating individual differences to the dimensions of human tasks. Ergonomics 21:1007-1019. 1978.
69. Fleishman EA. The Structure and Measurement of Physical Fitness. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1964.
70. Fleishman EA and Quaintance MK. Taxonomies of Human Performance. New York: Academic Press Inc., 1984.
71. Fox K. The reliability and validity of selected physical fitness tests for high school girls. Res Quart 30:430-437. 1959.
72. Friedl KE and Vogel JA. Validity of percent body fat prediction from circumferences: classification of men for weight control regulations. Mil Med 162:194-200. 1997.
73. Fuchs RJ, Theis CF and Lancaster MC. A nonogram to predict lean body mass in men. Am J Clin Nutr 31:673-678. 1978.
74. Fukunaga T, Kawakami Y, Kuno S, Funato K and Fukashiro S. Muscle architecture and function in humans. J Biomech 30:457-463. 1997.

75. General Accounting Office. Job opportunities for women in the military: progress and problems. Washington DC: General Accounting Office. Report No. FPCD-76-26, 1976.
76. General Accounting Office. Military attrition: better data, coupled with policy changes, could help the services reduce early separations. Washington DC: General Accounting Office. Report No. GAO/NSIAD 98-213, 1998.
77. General Accounting Office. Military recruiting. DOD could improve its recruiter selection and incentive systems. Washington DC: General Accounting Office. Report No. GAO/NAIAD-98-58, 1998.
78. Gastin PB. Energy system interactions and relative contributions during maximal exercise. Sports Med 31:725-741. 2001.
79. Gebhardt DL, Establishing performance standards. In *The Process of Physical Fitness Standards Development*, edited by Constable S and Palmer B (Human Systems Information and Analysis Center, Wright-Patterson Air Force Base, 2000).
80. Getchell LH, Kirkendall D and Robbins G. Prediction of maximal oxygen uptake in young adult women joggers. Res Quart 48:61-67. 1977.
81. Gibson ASC, Broomhead S, Lambert MI and Hawley JA. Prediction of maximal oxygen uptake from a 20-m shuttle run as measured directly in runners and squash players. J Sports Sci 12:331-335. 1998.
82. Gilchrist J, Jones BH, Sleet DA and Kimsey CD. Exercise-related injuries among women: strategies for prevention from civilian and military studies. MMWR 49:13-33. 2000.
83. Gledhill N and Jamnik VK. Development and validation of a fitness screening protocol for firefighter applicants. Can J Sports Sci 17:199-206. 1992.
84. Going SB, Densitometry. In *Human Body Composition*, edited by Roche AF, Heymsfield SB, and Lohman TG (Human Kinetics, Champaign Il, 1996).
85. Gollnick PD and Hermansen L, in *Exerc Sports Sci Rev*; *Vol. 1*, edited by Wilmore JH (Academic Press, New York, 1973), p. 1-43.
86. Gollnick PD, K.Piehl and Saltin B. Selective glycogen depletion pattern in human muscle fibers after exercise of varying intensity and at varying pedal rates. J Physiol 241:45-47. 1974.
87. Goodwin LD. Changing concepts of measurement validity. J Nurs Educ 36:102-107. 1997.
88. Grant S, Cobett K, Amjad AM, Wilson J and Aitchison T. A comparison of methods of predicting maximum oxygen uptake. Br J Sports Med 29:147-152. 1995.
89. Green HJ, Muscle power: fiber type, recruitment, metabolism and fatigue. In *Human Muscle Power*, edited by Jones NJ, McCartney N, and Mccomas AJ (Human Kinetics, Champaign, IL, 1986).
90. Gregurek R, Vukusic H and Tocilj G. Changes in pulse rate among the civilian population during air-raid alerts in the city of Zagreb. Mil Med 163:850-852. 1998.
91. Gunderson EK, Rahe RH and Authur RJ. Prediction of performance in stressful underwater demolition training. J Appl Psychol 56:430-432. 1972.

92. Gutin B. A model of physical fitness and dynamic health. J Phy Ed Recreation 51:48-51. 1980.
93. Guyton AC and Hall JE. Textbook of Medical Physiology. Philadelphia: W.B. Sanders, 1996.
94. Hagan RD, Smith MG and Gettman LR. Marathon performance in relation to maximal aerobic power and training indices. Med Sci Sports Exerc 13:185-189. 1981.
95. Hall DM. Motor fitness tests for farm boys. Res Quart 13:432-443. 1942.
96. Harman E, Frykman P and Kraemer W. Maximal cycling force and power at 40 and 100 RPM. Nat Strength Cond Ass J 8:71. 1986.
97. Harman EA and Frykman PN, The relationship of body size and composition to the performance of physically demanding military tasks. In *Body Composition and Physical Performance*, edited by Marriott BM and Grumstrup-Scott J (National Academy Press, Washington DC, 1992), p. 105-118.
98. Harman EA, Rosenstein MT, Frykman PN, Rosenstein RM and Kraemer WJ. Estimation of human power output from vertical jump. J Appl Sports Sci Res 5:116-120. 1991.
99. Harris JE. The differential measurement of force and velocity for junior high school girls. Res Quart 8:114-121. 1937.
100. Harris ML. A factor analytic study of flexibility. Res Quart 40:62-70. 1969.
101. Harrison MH, Bruce CL, Brown GA and Cochrane LA. A comparison of some indirect methods for predicting maximal oxygen uptake. Aviat Space Environ Med 51:1128-1133. 1980.
102. Hauret KG, Darakjy S, Canada S and Knapik JJ. Injury incidence and risk factors for male military police (Army). Med Sci Sports Exerc 35:S279. 2003.
103. Hauret KG, Knapik JJ, Lange JL, Heckel HA, Coval DL and Duplessis DH. Outcomes of Fort Jackson's Physical Training and Rehabilitation Program in Basic Combat Training - return to training, graduation, and two-year retention. Mil Med 169:562-567. 2004.
104. Heir T and Eide G. Injury proneness in infantry conscripts undergoing a physical training programme: smokeless tobacco use, higher age, and low levels of physical fitness are risk factors. Scand J Med Sci Sports 7:304-311. 1997.
105. Heir T and Glomsaker P. Epidemiology of musculoskeletal injuries among Norwegian conscripts undergoing basic military training. Scand J Med Sci Sports 6:186-191. 1996.
106. Hempel WE and Fleishman EA. A factor analysis of physical proficiency and manipulative skill. J Appl Psychol 39:12-16. 1955.
107. Henry FM. Factorial structure of speed and static strength in a lateral arm movement. Res Quart 31:440-447. 1960.
108. Herrin GD, Jaraiedi M and Anderson CK. Prediction of overexertion injuries using biomechanical and psychophysical models. Am Ind Hyg Assoc J 47:322-330. 1986.
109. Hesterman JL. The erosion of the enlisted force: a study of attrition. Maxwell Air Force Base AL: Air Command and Staff College Air University. Master's Thesis, 1999.
110. Highmore G. A factor analysis of athletic ability. Res Quart 27:1-11. 1956.

111. Hodgdon JA. Body composition in the military services: standards and methods. San Diego CA: Naval Health Research Center. Technical Report No. 90-21, 1990.
112. Hodgdon JA and Beckett MB. Prediction of percent body fat for U.S. Navy men from body circumference and height. San Diego, CA: Naval Health Research Center. Report No. 84-11, 1984.
113. Hodgdon JA and Beckett MB. Prediction of percent body fat for U.S. Navy women from body circumference and height. San Diego, CA: Naval Health Research Center No. 84-29, 1984.
114. Hodgdon JA and Jackson AS, Physical test validation for job selection. In *The Process of Physical Fitness Standards Development*, edited by Constable S and Palmer B (Human Systems Information and Analysis Center, Wright-Patterson Air Force Base, 2000).
115. Hogan J. The structure of physical performance in occupational tasks. J Appl Psychol 76:495-507. 1991.
116. Holloszy JO, in *Exerc Sports Sci Rev*; *Vol. 1*, edited by Wilmore JH (Academic Press, New York, 1973), p. 45-71.
117. Hopkins WG and Walker NP. The meaning of "physical fitness". Prev Med 17:764-773. 1988.
118. Hortobagyi T and Katch FI. Reliability of muscle mechanical characteristics for isokinetic and isotonic squat and bench press exercise using a multifunction computerized dynamometer. Res Quart 61:191-195. 1990.
119. Hutto LE. Measurement of the velocity factor and of athletic power in high school boys. Res Quart 9:109-128. 1938.
120. Ikai M and Fukunaga T. Calculation of muscle strength per unit of cross-sectional area of human muscle by means of ultrasounic measurement. Int Z Angew Einschl Arbeitsphysiol 26:26-32. 1968.
121. Ikai M and Fukungaga T. A study on training effect on strength per unit cross-sectional area of muscle by means of ultrasonic measurement. Int Z Angew Einschl Arbeitsphysiol 28:173-180. 1970.
122. Ismail AH and Cowell CC. Factor analysis of motor aptitude of preadolescent boys. Res Quart 32:507-513. 1961.
123. Jackson A, Watkins M and Patton RW. A factor analysis of twelve selected maximal isotonic strength performances on the Universal Gym. Med Sci Sports Exerc 12:274-277. 1980.
124. Jackson AS. Factor analysis of selected muscular strength and motor performance tests. Res Quart 42:164-172. 1971.
125. Jackson AS. Preemployment physical evaluation. Exerc Sports Sci Rev 22:53-90. 1994.
126. Jackson AS. Types of physical performance tests. In *The Process of Physical Fitness Standards Development*, edited by Constable S and Palmer B (Human Systems Information and Analysis Center, Wright-Patterson Air Force Base, 2000).
127. Jackson AS and Frankiewicz RJ. Factorial expressions of muscular strength. Res Quart 46:206-217. 1976.

128. Jackson AS and Pollock ML. Generalized equations for predicting body density of men. Br J Nutr 40:497-504. 1978.
129. Jackson AS, Pollock ML and Ward A. Generalized equations for predicting body density of women. Med Sci Sports Exerc 12:175-182. 1980.
130. Jamnik VK and Gledhill N. Development of fitness screening protocols for physically demanding occupations. Can J Sports Sci 17:222-227. 1992.
131. Jekel JF, Elmore JG and Katz DL. Epidemiology, Biostatistics, and Preventive Medicine. Philadelphia: W.B. Saunders, 1996.
132. Jette M. A Comparison between predicted VO2max and the Astrand procedure and the Canadian Home Fitness Test. Can J Appl Sports Sci 4(3):214-218. 1979.
133. Jones BH, Bovee MW, Harris JM and Cowan DN. Intrinsic risk factors for exercise-related injuries among male and female Army trainees. Am J Sports Med 21:705-710. 1993.
134. Jones BH, Bovee MW and Knapik JJ, Associations among body composition, physical fitness, and injuries in men and women Army trainees. In *Body Composition and Physical Performance*, edited by Marriott BM and Grumstrup-Scott J (National Academy Press, Washington, D.C., 1992), p. 141-173.
135. Jones BH, Cowan DN, Tomlinson JP, Robinson JR, Polly DW and Frykman PN. Epidemiology of injuries associated with physical training among young men in the Army. Med Sci Sports Exerc 25:197-203. 1993.
136. Jones BH and Knapik JJ. Physical training and exercise-related injuries. Surveillance, research and injury prevention in military populations. Sports Med 27:111-125. 1999.
137. Jones BH, Manikowski R, Harris JR, Dziados J, Norton S, Ewart T and Vogel JA. Incidence of and risk factors for injury and illness among male and female Army basic trainees. United States Army Research Institute of Environmental Medicine. Technical Report No. T19/88, 1988.
138. Katch FI and McArdle WD. Prediction of body density from simple anthropometric measurements in college-age men and women. Hum Biol 45:445-454. 1973.
139. Katch FI, McArdle WD, Czula R and Pechar GS. Maximal oxygen intake, endurance running performance, and body composition in college women. Res Quart 44:301-312. 1973.
140. Katch FI and Michael ED. Prediction of body density from skin-folds and girth measurements of college females. J Appl Physiol 25:92-94. 1968.
141. Klesius SE. Reliability of the AAHPER youth fitness test items and relative efficiency of the performance measure. Res Quart 39:809-811. 1968.
142. Kline GM, Porcari JP, Hintermeister R, Freedson PS, Ward A, Mccarron RF, Ross J and Rippe JM. Estimation of VO2max from a one-mile track walk, gender, age, and body weight. Med Sci Sports Exerc 19:253-259. 1987.
143. Knapik JJ. A review of the literature on attrition from the military services: risk factors and interventions to reduce attrition. Aberdeen Proving Ground: US Army Center for Health Promotion and Preventive Medicine. Technical Report No. 12-HF-01Q9A-04, 2004.

144. Knapik JJ, Ang P, Reynolds K and Jones B. Physical fitness, age and injury incidence in infantry soldiers. J Occ Med 35:598-603. 1993.
145. Knapik JJ, Canham-Chervak M, Hauret K, Hoedebecke E, Laurin MJ and Cuthie J. Discharges during US Army Basic Combat Training: injury rates and risk factors. Mil Med 166:641-647. 2001.
146. Knapik JJ, Canham-Chervak M, Hoedebecke E, Hewitson WC, Hauret K, Held C and Sharp MA. The Fitness Training Unit in Basic Combat Training: physical fitness, training outcomes, and injuries. Mil Med 166:356-361. 2001.
147. Knapik JJ, Cuthie J, Canham M, Hewitson W, Laurin MJ, Nee MA, Hoedebecke E, Hauret K, Carroll D and Jones BH. Injury incidence, injury risk factors, and physical fitness of U.S. Army basic trainees at Ft Jackson SC, 1997. Aberdeen Proving Ground, MD: U.S. Army Center for Health Promotion and Preventive Medicine. Epidemiological Consultation Report No. 29-HE-7513-98, 1998.
148. Knapik JJ, Darakjy S, Scott S, Hauret KG, Canada S, Marin R, Palkoska F, VanCamp S, Piskator E, Rieger W and Jones BH. Evaluation of two Army fitness programs: the TRADOC Standardized Physical Training Program for Basic Combat Training and the Fitness Assessment Program. Aberdeen Proving Ground, MD: US Army Center for Health Promotion and Preventive Medicine. Technical Report No. 12-HF-5772B-04, 2004.
149. Knapik JJ, Harper W and Crowell HP. Physiological factors in stretcher carriage performance. Eur J Appl Physiol 79:409-413. 1999.
150. Knapik JJ, Hauret K, Bednarek JM, Arnold S, Canham-Chervak M, Mansfield A, Hoedebecke E, Mancuso J, Barker TL, Duplessis D, Heckel H, Peterson J and 2001 Staff of the Army Physical Fitness School. The Victory Fitness Program. Influence of the US Army's emerging physical fitness doctrine on fitness and injuries in Basic Combat Training. Aberdeen Proving Ground, MD: US Army Center for Health Promotion and Preventive Medicine. Epidemiological Consultation Report No. 12-MA-5762-01, 2001.
151. Knapik JJ, Hauret KG, Arnold S, Canhan-Chervak M, Mansfield AJ, Hodedbecke EL and McMillian D. Injury and fitness outcomes during implementation of Physical Readiness Training. Int J Sports Med 24:372-381. 2003.
152. Knapik JJ, Hauret KG, Lange JL and Jovag B. Retention in service of recruits assigned to the Army Physical Fitness Test Enhancement Program in Basic Combat Training. Mil Med 168:490-492. 2003.
153. Knapik JJ, Meredith C, Jones B, Fielding R, Young V and Evans W. Leucine metabolism during fasting and exercise. J Appl Physiol 70:43-47. 1991.
154. Knapik JJ, Pontbriand Rd, Harper W, Tauson R, Swoboa J and Doss W. Feasibility of MOS task analysis and redesign to reduce physical demands in the U.S. Army. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory, Human Research and Engineering Directorate. Technical Report No. ARL-TR-1594, 1997.
155. Knapik JJ and Ramos MU. Isokinetic and isometric torque relationships in the human body. Arch Phy Med Rehab 61:64-67. 1980.

156. Knapik JJ, Scott SJ, Sharp MA, Hauret KG, Darakjy S, Rieger WR, Plakoska FA, VanCamp SE and Jones BH. Guidance for ability group run speeds and distances in Basic Combat Training. Aberdeen Proving Ground MD: US Army Center for Health Promotion and Preventive Medicine. Technical Report No. 12-HF-5772a-03, 2003.
157. Knapik JJ, Sharp MA, Canham ML, Hauret K, Cuthie J, Hewitson W, Hoedebecke E, Laurin MJ, Polyak C, Carroll D and Jones B. Injury incidence and injury risk factors among US Army Basic Trainees at Ft Jackson, SC (including fitness training unit personnel, discharges, and newstarts). Aberdeen Proving Ground MD: US Army Center for Health Promotion and Preventive Medicine. Epidemiological Consultation Report No. 29-HE-8370-99, 1999.
158. Knapik JJ, Sharp MA, Canham-Chervak M, Hauret K, Patton JF and Jones BH. Risk factors for training-related injuries among men and women in Basic Combat Training. Med Sci Sports Exerc 33:946-954. 2001.
159. Knapik JJ, Sharp MA, Jones BH, Darakjy S, Jones S, Hauret KG and Piskator G. Secular tends in the physical fitness of American youth, young adults and Army recruits. Aberdeen Proving Ground: US Army Center for Health Promotion and Preventive Medicine. Technical Report No. 12-HF-01Q9B-04, 2004.
160. Knapik JJ, Wright JE and Vogel JA. Measurement of isometric strength in an upright pull of 38 cm. Natick, MA: U.S. Army Research Institute of Environmental Medicine. Technical Report No. T3/81, 1981.
161. Koplan JP, Powell KE, Sikes RK, Shirley RW and Campbell CC. An epidemiologic study of the benefits and risks of running. JAMA 248:3118-3121. 1982.
162. Kowal DM, Vogel JA, Sharp D and Knapik JJ. Analysis of attrition, retention and criterion task performance of recruits during training. Natick MA: US Army Research Institute of Environmental Medicine. Technical Report No. T2/82, 1982.
163. Krahenbuhl GS, Panagrazi RP, Petersen GW, Burkett LN and Schneider MJ. Field testing of cardiorespiratory fitness in primary school children. Med Sci Sports Exerc 10:208-213. 1978.
164. Kroll W. Reliability of a selected measure of human strength. Res Quart 33:410-415. 1962.
165. Larson LA. A factor analysis of motor ability variables and tests, with tests for college men. Res Quart 12:499-517. 1941.
166. Larson LA. A factor and validity analysis of strength variables and tests with a test combination of chinning, dipping and vertical jump. Res Quart 11:82-96. 1940.
167. Laurence JH, Naughton J and Harris DA. Attrition revisited: identifying the problem and its solution. Alexandria VA: Human Resources Research Organization. ARI Research Note No. 96-20, 1996.
168. Leach DA. The measurement of cardiorespiratory endurance and the standard of evaluation for Army personnel in the 40-45 age category. Carlisle Barracks PA: US Army War College. Report No. 1983.
169. Lee SW. Task-related aerobic and anaerobic physical fitness standards for the Canadian Army. Edmonton Alberta Canada: University of Alberta,

Department of Physical Education and Sports Studies. Doctoral Dissertation No. 1992.
170. Leger L and Gadoury C. Validity of the 20 m shuttle run with 1 min stages to predict VO2max in adults. Can J Sports Sci 14:21-26. 1989.
171. Leger LA and Lambert J. A maximal multistage 20-m shuttle run test to predict VO2max. Eur J Appl Physiol 49:1-12. 1982.
172. Leger LA, Seliger V and Brassard L. Backward extrapolation of VO2max values from O2 recovery curves. Med Sci Sports Exerc 12:24-27. 1980.
173. Lehmann M, Berg A, Kapp R, Wessinghage T and Keul J. Correlations between laboratory testing and distance running performance in marathoners of similar performance ability. Int J Sports Med 4:226-230. 1983.
174. Lieber RL and Friden J. Functional and clinical significance of skeletal muscle architecture. Clin Orthop Related Res 383:140-151. 2000.
175. Linden W and McEachern HM. A review of physiological prestress adaptation: effects of duration and context. Int J Psychophysiol 2:239-245. 1985.
176. Lohman TG, Dual energy X-ray aborptiometry. In *Human Body Composition*, edited by Roche AF, Heymsfield SB, and Lohman TG (Human Kinetics, Champaign, IL, 1996), p. 63-78.
177. Maksud MG, Cannistra C and Dublinski D. Energy expenditure and VO2max of female athletes during treadmill exercise. Res Quart 47:692-697. 1976.
178. Maksud MG and Coutts KD. Application of the Cooper twelve-minute run-walk test to young males. Res Quart 42:54-59. 1971.
179. Marcinik EJ, Hyde DE and Talyor WF. The relationship between the U.S. Navy Fellet Diver Physical Screening Test and job task performance. Aviat Space Environ Med 66:320-324. 1995.
180. Marsh HW. The multidimensional structure of physical fitness: invariance over age and gender. Res Q Exerc Sport 64:256-273. 1993.
181. Marti B, Vader JP, Minder CE and Abelin T. On the epidemiology of running injuries. The 1984 Bern Grand-Prix study. Am J Sports Med 16:285-294. 1988.
182. Maughan RJ. Relationship between muscle strength and cross-sectional area. Sports Med 1:263-269. 1984.
183. Maughan RJ and Leiper JB. Aerobic capacity and fractional utilization of aerobic capacity in elite and non-elite male and female marathon runners. Eur J Appl Physiol 52:80-87. 1983.
184. Maughan RJ, Watson JS and Weir J. Strength and cross-sectional area of human skeletal muscle. J Physiol 1983:37-49. 1983.
185. McArdle WD, Katch FI and Katch VL. Exercise Physiology: Energy, Nutrition and Human Performance. Philadelphia: Lea & Febiger, 1991.
186. McCloy CH. A factor analysis of tests of endurance. Res Quart 27:213-216. 1956.
187. McCloy CH. The measurement of general motor capacity and general motor ability. Res Quart 5:46-61. 1934.
188. McCloy E. Factor analysis methods in the measurement of physical abilities. Res Quart 6:114-121. 1935.

189. McCraw LW. A factor analysis of motor learning. Res Quart 20:316-335. 1949.
190. McCutcheon MC, Sticha SA, Giese MD and Nagle FJ. A further analysis of the 12-minute run prediction of maximal aerobic power. Res Q Exerc Sport 61:280-283. 1990.
191. McNaughton L, Hall P and Cooley D. Validation of several methods of estimating maximal oxygen uptake in young men. Perc Mot Skills 87:575-584. 1998.
192. Mello RP, Murphy MM and Vogel JA. Relationship between the Army two mile run test and maximal oxygen uptake. U.S. Army Research Institute of Environmental Medicine. Technical Report No. T3/85, 1984.
193. Metheny E. Studies of the Johnson Test as a test of motor educability. Res Quart 9:105-127. 1938.
194. Mitchell JH, Sproule BJ and Chapman CB. The physiological meaning of the maximal oxygen uptake test. J Clin Invest 37:538-547. 1957.
195. Mueller WH and Malina RM. Relative reliability of circumferences and skinfolds as measures of body fat distribution. Am J Phy Anthropol 72:437-439. 1987.
196. Murphy MM, Knapik JJ, Vogel JA and Drews F. Relationship of anaerobic power capacity to performance during a 5-day sustained combat scenario. Natick MA: US Army Research Institute of Environmental Medicine. Technical Report No. T5/84, 1984.
197. Myers DC, Gebhardt DL, Crump CE and Fleishman EA. The dimensions of human physical performance: factor analysis of strength, stamina, flexibility, and body composition measures. Human Perf 6:309-344. 1993.
198. Myers DC, Gebhardt DL, Crump CE and Fleishman EA. Validation of the military entrance physical strength capacity test. Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences. Technical Report No. 610, 1984.
199. Myers DC, Gebhart DL and Fleishman EA. Development of physical performance standards for Army jobs: the job analysis methodology. Washington, D.C.: Advanced Research Resources Organization. Technical Report No. 446, 1980.
200. Myers J and Ashley E. Dangerous curves. A perspective on exercise, lactate and the anaerobic threshold. Chest 111:787-795. 1997.
201. Nicks DC and Fleishman EA. What do physical fitness tests measure? A review of factor analytic studies. Educ Psychol Meas 23:77-95. 1962.
202. Novaco RW, Cook TM and Sarason IG, in *Stress Reduction and Prevention*, edited by Meichenbaum D and Jaremko ME (Plenum Press, New York NY, 1983).
203. Paliczka VJ, Nichols AK and Boreham CAG. A multi-stage shuttle run as a predictor of running performance and maximal oxygen uptake in adults. Br J Sports Med 21:163-165. 1987.
204. Pate RR. A new definition of youth fitness. Physician Sportsmed 11:77-83. 1983.

205. Patton JF, Daniels WL and Vogel JA. Aerobic power and body fat of men and women during Army Basic Training. Aviat Space Environ Med 51:492-496. 1980.
206. Pette D and Sapmer C. Metabolic subpopulations of muscle fibers a quantitative study. Diabetes 28:25-29. 1979.
207. Phillips M. Study of a series of physical education tests by factor analysis. Res Quart 20:60-71. 1949.
208. Pollock ML, Hickman T, Kendrick Z, Jackson A, Linnerud AC and Dawson G. Prediction of body density in young and middle-aged men. J Appl Physiol 40:300-304. 1976.
209. Pollock ML, Laughridge EE, Coleman B, Linnerud AC and Jackson A. Prediction of body density in young and middle-aged women. J Appl Physiol 38:745-749. 1975.
210. Pope RP, Herbert RH, Kirwan JD and Graham BJ. Predicting attrition in basic military training. Mil Med 164:710-714. 1999.
211. Powers S and Howley ET. Exercise Physiology. New York: McGraw-Hill, 2001.
212. Ramos MU and Knapik JJ. Instrumentation and techniques for the measurement of muscular strength and endurance in the human body. United States Army Research Institute of Environmental Medicine, Natick MA. Technical Report No. T2-80, 1979.
213. Ramsbottom R, Brewer J and Williams C. A progressive shuttle run to estimate maximal oxygen uptake. Br J Sports Med 22:141-144. 1988.
214. Rarick L. An analysis of the speed factor in simple athletic activities. Res Quart 8:89-105. 1937.
215. Rasch PJ and Wilson D. The correlation of selected laboratory tests of physical fitness with military endurance. Mil Med 129:256-258. 1964.
216. Rayson M, Job analysis. In *The Process of Physical Fitness Standards Development*, edited by Constable S and Palmer B (Human Systems Information Analysis Center, Wright-Patterson Air Force Base, 2000).
217. Rayson M, Holliman D and Belyavin A. Development of a physical selection procedure for the British Army. Phase 2: relationship between physical performance tests and criterion tasks. Ergonomics 43:73-105. 2000.
218. Rayson M, Holliman DE, Nevola RV and Birch CL. Physical selection standards for the British Army: Phase 5 validation. Farnborough, Hampshire, England: Center for Human Sciences. Technical Report No. PLSD/CHS5/CR96/021, 1996.
219. Rayson M and Wilkinson D. Physical selection standards for recruits: Amended PSS(R) loaded road march models. Clifton, Bristol, United Kingdom: Optimal Performance. Repo, 2003.
220. Rayson M and Wilkinson D. Potential risk modifiers for training outcomes and injury in single entry recruits: body mass, composition and aerobic fitness. Farnham, Surry, England: Optimal Performance. Technical Report dated 22 April 2003.
221. Rayson MP, Bell DG, Holliman DE, Llewellyn M, Nevola VR and Bell RL. Physical selection standards for the British Army. Franborough, Hampshire,

England: Army Personnel Research Establishment. Technical Report No. 94R036, 1994.
222. Rayson MP, Pynn H, Rothwell A and Nevill AM. Validation of job-related fitness tests. Med Sci Sports Exerc 32:S126. 2000.
223. Reynolds KL, Heckel HA, Witt CE, Martin JW, Pollard JA, Knapik JJ and Jones BH. Cigarette smoking, physical fitness, and injuries in infantry soldiers. Am J Prev Med 10:145-150. 1994.
224. Ribisl PM and Kachadorian WA. Maximal oxygen intake prediction in young and middle-aged males. J Sports Med Phy Fit 9:17-22. 1969.
225. Robertson DW. Development of an occupational strength test battery. San Diego CA: Navy Personnel Research and Development Center. Technical Report No. NPRDC-TR-82-42, 1982.
226. Roche AF, Heymsfield SB and Lohman TG. Human Body Composition. Champaign, IL: Human Kinetics, 1996.
227. Schonfeld BR, Doerr DF and Convertino VA. An occupational performance test validation program for firefighters at the Kennedy Space Center. J Occ Med 32:638-643. 1990.
228. Seashore HG. Some relationships of fine and gross motor abilities. Res Quart 13:259-274. 1942.
229. Sharp DS, Wright JE, Vogel JA, Patton JF, Daniels WL, Knapik JJ and Kowal DM. Screening for physical capacity in the U.S. Army: an analysis of measures predictive of strength and stamina. U.S. Army Research Institute of Environmental Medicine. Technical Report No. T8/80, 1980.
230. Sharp MA, Patton JF, Knapik JJ, Smutok MA, Hauret K, Mello RP, Ito M and Frykman PN. A comparison of the physical fitness of men and women entering the US Army during the years 1978-1998. Med Sci Sports Exerc 34:356-363. 2002.
231. Shaver LG. Maximum aerobic power and anaerobic work capacity prediction from various running performances of untrained college men. J Sports Med Phy Fit 15:147-149. 1975.
232. Shim AL, Bailey ML and Westings SH. Development of a field test for upper body power. J Strength Condit Res 15:192-197. 2001.
233. Sills FD. A factor analysis of somatotypes and of their relationship to achievement in motor skills. Res Quart 21:424-437. 1950.
234. Singh M, Lee SW, Wheeler GD, Chahal P, Oseen M and Couture RT. Development of the Forces Mobile Command Army Physical Fitness Standards for field units. Edmonton, Alberta, Canada: Department of Physical Education and Recreation, University of Alberta. Report No. 1991.
235. Siri WE, Body composition from fluid spaces and density: analysis of methods. In *Techniques for measuring body composition*, edited by Brozek J and Henschel A (National Academy of Sciences National Research Council, Washington, D.C., 1961).
236. Siri WE. The gross composition of the body. Advances in Biology and Medicine 4:239-280. 1956.
237. Sjodin B and Svedenhag J. Applied physiology of marathon running. Sports Med 2:83-99. 1985.

238. Snoddy RO and Henderson JM. Predictors of basic infantry success. Mil Med 159:616-622. 1994.
239. Snook SH, Campanelli RA and Hart JW. A study of three preventive approaches to low back injury. J Occ Med 20:478-481. 1978.
240. Spangenburg EE and Booth FW. Molecular regulation of individual skeletal muscle fiber types. Acta Physiol Scand 178:413-424. 2003.
241. Staron RS. Human skeletal muscle fiber types: delineation, development and distribution. Can J Appl Physiol 22:307-327. 1997.
242. Stevenson JM, Bryant JT, Andrews GM, Smith JT, French SL, Thomson JM and Deakin JM. Development of physical fitness standards for Canadian Armed Forces younger personnel. Can J Sports Sci 17:214-221. 1992.
243. Stevenson JM, Deakin JM, Andrews GM, Bryant JT, Smith JT and Thomson JM. Development of physical fitness standards for Canadian Armed Forces older personnel. Can J Appl Physiol 19:75-90. 1994.
244. Strickland MK, Petersen SR and Bouffard M. Prediction of maximal aerobic power from the 20-m multi-stage shuttle run test. Canadian Journal of Applied Physiology 28:272-282. 2003.
245. Taggart P, Carruthers M and Somerville W. Some effects of emotion on the normal and abnormal heart. Curr Probl Cardiol 7:1-29. 1983.
246. Talmadge RJ, Roy RR and Edgerton VR. Muscle fiber type and function. Curr Opin Rheumatol 5:695-705. 1993.
247. Tate MM. Utility of demographic and psychological measures for the prediction of Army separation. Los Angeles, CA: University of Southern California. Doctoral Dissertation, 1986.
248. Teves MA, Wright JE and Vogel JA. Performance on selected candidate screening test procedures before and after Army Basic and Advanced Individual Training. Natick, MA: U.S. Army Research Institute of Environmental Medicine. Technical Report No. No. T13/85, 1985.
249. Thomas CW. Fitness Training Unit (FTU) entry and exit criteria. Ft Monroe VA: US Army Training and Doctrine Command. Memorandum No. 10 September 1999, 1999.
250. Army Training and Doctrine Command. A review and analysis of the Army's Trainee Discharge Program. Fort Monroe VA: US Army Training and Doctrine Command. Report, 1984.
251. Tuttle W, Janney C and Salzano J. Relation of maximal back and leg strength to back and leg endurance. Res Q Exerc Sport 26:96-106. 1955.
252. Van der Doelen LHM, Dijk J, Visser T and Veenstra BJ. Physiological aspects of the military task lifting. LTG Knoop Barracks, Utrecht, Netherlands: Training Medicine and Training Physiology Team/NATCO. TGFT Report No. 96-104, 1996.
253. vanDijk J, Visser T, van der Doelen LHM and Veenstra BJ. Physiological aspects of the military task weightload marching. LTG Knoop Barracks, Utrecht, Netherlands: Training Medicine and Training Physiology Team/NATCO. TGTF Report No. 96-103, 1996.

254. VanNostrand SJ, Thompson ME and Captain GJ. Evaluation of the Military Entrance Physical Strength Capacity Test (E-MEPSCAT). Bethesda MD: US Army Concepts Analysis Agency. Technical Report No. CAA-SR-85-23, 1985.
255. Visser T, Dijk J, van der Doelen LHM and Veenstra BJ. Physiological aspects of the military task digging. LTG Knoop Barracks, Utrecht, Netherlands: Training Medicine and Training Physiology Team/NATCO. TGTF Report No. 96-106, 1996.
256. Visser T, Dijk J, van der Doelen LHM and Veenstra BJ. Physiological aspects of the military task load carrying. LTG Knoop Barracks, Utrecht, Netherlands: Training Medicine and Training Physiology Team/NATCO. TGTF Report No. 96-105, 1996.
257. Vogel JA. A review of fitness as it pertains to the military service. U.S. Army Research Institute of Environmental Medicine. Technical Report No. No. T14/85, 1985.
258. Vogel JA and Friedl KE, Army data: body composition and physical performance. In *Body Composition and Physical Performance. Applications for Military Services*, edited by Marriott BM and Grumstrup-Scott J (National Academy Press, Washington, D.C., 1992), p. 89-103.
259. Vogel JA, Kirkpatrick JW, Fitzgerald PI, Hodgdon JA and Harman EA. Derivation of anthropometry based body fat equations for the Army's weight control program. U.S. Army Research Institute of Environmental Medicine. Technical Report No. No. 17-88, 1988.
260. Vogel JA, Wright JE, Patton JF, Dawson J and Eschenbach MP. A system for establishing occupationally-related gender-free physical fitness standards. Natick MA: US Army Research Institute of Environmental Medicine. Technical Report No. T5/80, 1980.
261. Wardynski C, Halford L, Koss T, Frick M, Peterson P and Bardo R. Understanding and managing entry level separations in the training base. West Point, NY: Department of Social Sciences, Office of Economic and Manpower Analysis, 1999.
262. Washington S. Relationship between body mass index and musculoskeletal system and connective tissue disorders, US Army, 1990-1999. MSMR 6(7):2-10. 2000.
263. Wendler AJ. A critical analysis of test elements used in physical education. Res Quart 9:64-76. 1938.
264. Westerblad H, Allen DG and Lannergren J. Muscle fatigue: lactic acid or inorganic phosphate the major cause. News Physiol Sci 17:17-21. 2002.
265. Westphal KA, Friedl KE, Sharp MA, King N, Kramer TR, Reynolds KL and Marchitelli LJ. Health, performance and nutritional status of U.S. Army women during basic combat training. Natick, MA: U.S. Army Research Institute of Environmental Medicine. Technical Report No. T96-2, 1995.
266. Wiley JF and Shaver LG. Prediction of maximal oxygen intake from running performances of untrained young men. Res Quart 43:89-93. 1972.
267. Willford HN, Duey WJ, Olsoh MS, Howard R and Wang N. Relationship between fire fighting suppression tasks and physical fitness. Ergonomics 42:1179-1186. 1999.

268. Wilmore JH and Behnke AR. An anthropometric estimate of body density and lean body weight in young men. J Appl Physiol 27:25-31. 1969.
269. Wilmore JH and Behnke AR. An anthropometric estimate of body density and lean body weight in young women. Am J Clin Nutr 23:267-273. 1970.
270. Wilmore JH and Costill DL. Physiology of Sports and Exercise. Champaign, Il: Human Kinetics, 1994.
271. Wilmore JH and Davis JA. Validation of a physical abilities field test for the selection of state traffic officers. J Occ Med 21:33-40. 1979.
272. Wright HF, Dotson CO and Davis PO. A simple technique for measurement of percent body fat in man. US Navy Med 72:23-27. 1981.
273. Wright HF, Dotson CO and P.O.Davis. An investigation of assessment techniques for body composition of women Marines. US Navy Med 71:15-26. 1980.
274. Wyndham CH, Strydom NB, Graan CH, VanRensburg AJ, Rogers GG, Greyson JS and VanderWalt WH. Walk or jog for health II. Estimating the maximal aerobic capacity for exercise. South Af Med J 45:6-9. 1971.
275. Young CM, Martin MEK, Tensuan R and Blondin J. Predicting specific gravity and body fat in young women. J Am Diet Ass 40:102-107. 1962.
276. Zuidema MA and Baumgartner TA. Second factor analysis study of physical fitness tests. Res Quart 42:247-256. 1974.
277. Zwiren LD, Freedson PS, A W, Wilke S and Rippe JM. Estimation of VO2max: a comparative analysis of five exercise tests. Res Q Exerc Sport 62:73-78. 1991.

Appendix B
Uniform Guidelines On Employee Selection Procedures

The Equal Employment Opportunity Commission (EEOC), Civil Service Commission, Department of Labor, and the Department of Justice adopted guidelines for employee selection procedures in 1978. These guidelines have been revised and the latest revision is dated 1 July 2003 (61). These guidelines have also been summarized by Hodgdon and Jackson (114).

The guidelines indicate that an employee selection procedure has adverse impact if the selection rate for any race, sex, or ethnic group is less than 4/5 (80%) of the group with the highest selection rate. Adverse impact is generally implied unless the employer can show that the selection procedures is justified because of the nature of the job. Such justification can be established through validity studies that show the selection procedure is specifically linked to the job in objective and measurable ways.

The guidelines define acceptable validity studies as those involving criterion-related validity, content validity or construct validity that are consistent with professional standards (87). Evidence of criterion-related validity is that the test is predictive of critical or important elements of the job. To determine criterion-related validity a set of critical job elements (e.g., task performance, injuries, employee rating) are selected and the relationship between these job elements are established using correlational and regression analysis techniques. Evidence of content validity is that the test is linked to important elements of the job. To determine content validity, the job is examined and specific job tasks or simulations of these tasks are developed and used in the selection process. Evidence of construct validity is that the tests are related to a particular trait (e.g., physical fitness) that is important for successful performance of a job. To determine construct validity it must be demonstrated that a particular characteristic or set of characteristics (e.g., components of physical fitness) are required for successful job completion.

General Guidelines

In addition to validity requirements there are several other standards that must be met. Any selection procedure that has adverse impact must have documentation showing that technical validation standards have been met (described below). The validation studies must be carried out under conditions that assure accuracy with administration under standardized conditions. Caution is advised against using tests that that can be learned in a brief orientation period and have adverse impact. If a particular method has a greater adverse impact than another method the user should have evidence to support the greater validity of the selected method. Where cutoff scores are used they should be consistent with acceptable proficiency on the job. If it is expected that the applicant will progress to a higher job level automatically or in a timely manner, the selection procedure can be used to assist in selecting for the higher level job; if there is not automatic or timely progression to higher job levels, the tests must evaluate the entry level position. If a test has been used that is not fully validated users may continue using those tests as long as there is some evidence of validity and there is a plan to fully validate the test in a timely manner. Whenever the validity of a test has been demonstrated additional studies need not be performed unless dictated by a review of alternative valid selection procedures that might have less adverse impact.

Employers may use selection procedures that have not been validated to eliminate adverse impact or as part of affirmative action programs. In circumstances where validation studies cannot be performed, tests should be as job-related as possible and designed to reduce or eliminate adverse impact. Validation studies that are not conducted by the employer are permitted as long as the selection tests meet professional validation requirements, the employer's job is similar to the job involved in the validation test, and the validation study includes a consideration of adverse impact. Cooperative studies among employers, labor organizations, and employment agencies are encouraged. Unacceptable substitutes for validation studies include the general reputation of a test, assumptions of validity based on name, promotional literature, frequency of use, testimonials, and other non-empirical or anecdotal accounts. Employment services/agencies must conform to the guidelines in the same manner as the actual employer. Applicants who were denied equal treatment because of prior discriminatory practices must be afforded the opportunities that existed for other employees during the period of discrimination and allowed to qualify under less stringent procedures unless the user can demonstrate that the increased standards are required by business necessity. There should be opportunities for retesting. The use of validated selection procedures does not relieve employers of affirmative action obligations.

Technical Standards For Validity Studies

In general, validity is the extent to which a test measures what it purports to measure (87). The guidelines prescribe minimum technical standards for studies involving criterion-related validity, content validity, and construct validity.

Minimum Technical Standards For Criterion-Related Validation

The employer must determine if it is technically feasible to do a criterion-related validation study in their employment context. The number of people needed for the study should be determined based on selection procedure, potential sample available, and the employment situation. Jobs can be grouped if they have similar major work behaviors. There is no requirement to hire or promote workers to conduct a criterion-related validation study.

The job should be examined to determine tasks that are relevant. Relevant tasks are those that represent critical or important job duties. The possibility of bias needs to be considered carefully.

The sample subjects should be representative of the market of recruits normally available in the labor market for the job. It should include races, sexes, and ethnic groups normally in the relevant job market.

The degree of relationship between the criterion measure and the tests should be examined using acceptable statistical procedures. Generally, a relationship significant at an alpha level of 0.05 ($p<0.05$) meets the guideline criteria.

Employers should review tests to assure they are adequate for operational use. Generally, the test will be appropriate for use in proportion to the size of the correlation coefficient between the test and the criteria and to the extent critical aspects of the job are covered by the criterion. Low correlations and criterions that consider only limited job aspects will be subject to close review.

Employers should avoid techniques that over-inflate validity. These include reliance on a few selection procedures or few performance criteria when many performances are required on the job. The use of optimal weight statistics involving a single sample tend to over-inflate validity estimates. Tests should involve large samples and cross-validation.

Employers should be “fair” in selection procedures. Unfairness results when a particular selection process has an adverse impact on a particular group and the differences in scores are not reflected in measures of job performance.

Minimum Technical Standards For Content Validity Studies

Employers should determine if it is appropriate to conduct a content validity study in the particular employment context. Selection procedures based on content validity can be supported to the extent that it is a representative sample of the job. Content validity strategies are not appropriate for knowledge, and skills that are to be learned on the job.

There should be a job analysis that includes all the important work behaviors required for successful job performance. The tasks selected for measurement should be critical and/or important work behaviors constituting most of the job.

To demonstrate content validity, the employer should show that the behaviors are a representative sample of the tasks involved in the job or that the tests provide a representative sample of the work products of the job. The closer the content and context of the test to the work behaviors the stronger is the basis for showing content validity.

Statistical reliability is a matter of concern for tests of content validity. Whenever feasible the reliability (repeatability) of the tests should be determined.

Where a measure of success in the training program is used as a selection tool, it must be shown that there is a relationship between the content of the training program and the content of the job.

If an employer can show that a higher score on a test is likely to result in better job performance, the results may be used to rank persons who exceed some minimum level. Where a test based only on content validity is used to rank personnel the test should measure aspects of performance which differentiate among levels of job performance.

Minimum Technical Standards for Construct Validity Studies

Construct validation studies are more complex than criterion-related or content validity studies and particular care must be taken to assure that the standards are met.

There should be a job analysis to show the critical and important work behaviors required for successful job performance and the constructs believed to underlie successful performance of these work behaviors. Each construct should be named and defined to distinguish it from other constructs.

A selection procedure should then be identified or developed that measures the construct. The employer should show that the construct is validly related to critical and/or import work behaviors.

Claims of construct validity without a criterion-related validity study will be accepted only in cases where a criterion-related study has been conducted and meets the standards for transportability of criterion-related studies.

Appendix C
Acknowledgements

We would also like to thank Ms Sara Canada for the technical review and Jeanette England for the editorial review. Ms Carol Pace assembled the final manuscript for publication.