RELIABILITY, VALIDITY AND PERFORMANCE DETERMINANTS OF THE MILITARY SIMULATION TEST FOR THE ASSESSMENT OF ANAEROBIC PERFORMANCE IN SOLDIERS

Joonas Helén

Master's Thesis
Science of Sport Coaching and Fitness Testing
Faculty of Sport and Health Sciences
University of Jyväskylä
Fall 2020
Supervisor: Keijo Häkkinen

ABSTRACT

Helén, J. 2020. Reliability, validity and performance determinants of the military simulation test for the assessment of anaerobic performance in soldiers. Faculty of Sport and Health Sciences, University of Jyväskylä, Master's Thesis in Science of Sport Coaching and Fitness Testing, 64 pp.

Introduction: Measurements of physical fitness have been identified as one of the most important research priorities in the military. Currently, the physical fitness tests of The Finnish Defence Forces (FDF) do not include a valid and reliable anaerobic test. In order to develop job-related task simulation tests, the purpose of the study was to determine the reliability and validity of a high-intensity military simulation test (MST), which is designed to assess the task-specific anaerobic performance of soldiers. The MST includes typical combat-related maneuvers while wearing a 20 kg combat load.

Methods: 22 male cadets (22.5 ± 0.8 years) volunteered to participate. The MST was performed three times with a minimum of 48 hours of recovery between trials to determine reliability. To study the validity and the contribution of different physical characteristics on performance, the MST was compared with measurements of body composition, muscular strength and endurance, maximal oxygen uptake (VO_2max), and anaerobic performance.

Results: The mean MST times for the first, second, and third trial were 108.5 ± 8.6, 103.3 ± 6.6 and 98.4 ± 6.5 s, respectively. Repeated measures ANOVA showed significant improvement in the MST time between every trial ($p < 0.05$ for both). The intraclass correlation coefficient was 0.58 (95 % CI: 0.03–0.84) between the first and second trial and 0.74 (95 % CI: −0.06–0.94) between the second and third trial. The MST performance was significantly correlated with several measured variables. The strongest relationships were observed with mean and peak power in the 60-second Bosco jump test ($r = -0.65$, $p < 0.01$; $r = -0.68$, $p < 0.01$, respectively), maximal isometric strength of the upper body ($r = -0.67$, $p < 0.01$), standing long jump ($r = -0.66$, $p < 0.01$), sit-ups ($r = -0.69$, $p < 0.01$), and VO_2max ($r = -0.63$, $p < 0.01$).

Discussion: The improvement in MST time in every trial suggested a significant learning effect, especially between the first and second trials. Due to the complex nature of MST, several familiarization trials are required to reach statistically stable results. In addition, the MST may not be considered as a valid task-specific anaerobic test for soldiers, as the performance in MST relies on the simultaneous participation of all three energy pathways, and a combination of neuromuscular, anaerobic, and anaerobic performance characteristics.

Key words: physical employment standards, anaerobic testing, military anaerobic performance

CONTENTS

1 INTRODUCTION

Physical demands in operational environments and measurements of physical fitness have been identified as the most important research priorities related to soldiers' health and performance (Lovalekar et al. 2018). The physical demands are absolute and are not altered by age, anthropometrics, or gender. Generic fitness tests, such as push-ups, sit-ups, curl-ups, and running tests impose a physiological bias against heavier, not fatter, individuals and do not measure absolute fitness in military occupations (Vanderburgh 2007). In addition, most of these tests are conducted in sportswear, which is not specific to most military tasks requiring external load carrying. The use of military-specific task simulation tests is recommended (Reilly et al. 2015).

Currently, the physical fitness tests of The Finnish Defence Forces (FDF) include a 12-minute running test to assess aerobic capacity, and standing long jump, sit-ups and push-ups to assess muscular strength and endurance. Waist circumference is measured to evaluate body composition and health risks. However, there is a lack of a valid and reliable military-specific anaerobic test, which could be a useful tool for establishing physical employment standards (PESs). There is a need for a broad group of field-expedient fitness tests to properly evaluate military performance, at least in the most demanding military occupational specialties (MOSs).

Pihlainen et al. (2018) developed the military simulation test (MST) to measure occupationally relevant anaerobic performance. The MST includes similar maneuvers and tasks as soldiers are expected to be able to perform with high intensity. The mean MST performance time was 148 ± 22 seconds. The strongest correlations were with the countermovement jump (CMJ) performed with combat gear ($r = -0.66$) and the dead mass ratio ($r = -0.66$), which was calculated by dividing body mass by fat mass. In addition, significant correlations were found with standing long jump, CMJ without combat gear, sit-ups, body fat, push-ups, isometric strength of the upper and lower extremities, 3 000 m run, skeletal muscle mass, and pull-ups. The stepwise multivariate regression analysis showed that 66 % of the variance in the MST time was explained by the CMJ with combat gear, 3 000 m running test and skeletal muscle mass. The MST was concluded to be a promising military-specific assessment method of muscle power of the lower extremities and endurance capacity. The concurrent validity against generic anaerobic fitness tests and the reliability of the MST was not studied.

The purpose of this study was to evaluate the contribution of body composition, maximal oxygen uptake (VO_2max), maximal strength and strength endurance, and anaerobic performance to performance in the MST. The main focus was on associations with the MST and three anaerobic fitness tests (the 60-second Wingate cycle test, 300 m run, and the 60-second Bosco jump test) in order to evaluate the MST as a valid military-specific anaerobic test. In addition, the reliability of the MST was studied to determine the number of trials required to obtain statistically stable results.

2 PHYSIOLOGICAL REQUIREMENTS OF MILITARY PERSONNEL

Soldiers are exposed to many physiological and environmental stressors in military operations (figure 1). Challenges are, for example, maintenance of physical fitness and body composition, load carriage, environmental extremes (heat, cold, and altitude), and injuries. The effect of these stressors must be fully understood to ensure the use of proper risk mitigation strategies. (Nindl et al. 2013.) The physiological changes during deployment are fairly well documented. The most typical change is the decrease in aerobic performance, whereas strength levels are usually maintained or improved. There are contradictory findings about changes in body composition. (Sharp et al. 2008; Lester et al. 2010; Warr et al. 2012; Rintamäki et al. 2012; Fallowfield et al. 2014; Nagai et al. 2016.)

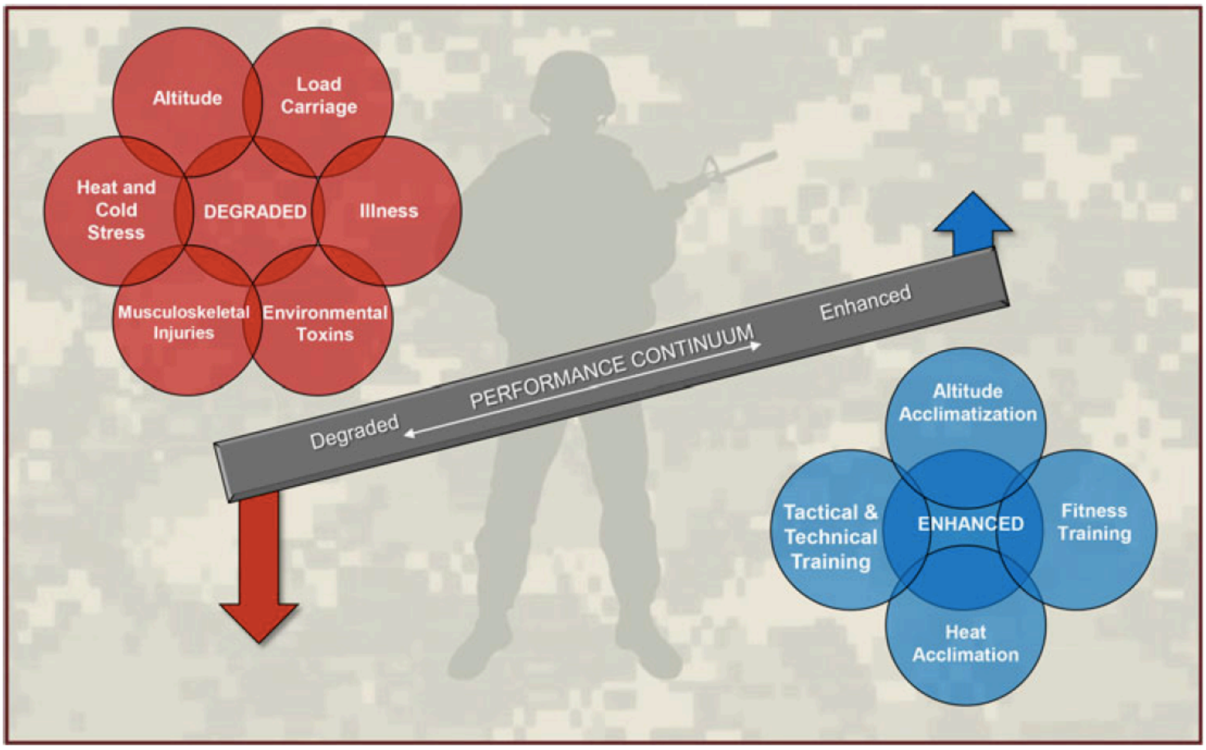

FIGURE 1. Factors affecting soldier performance during deployment. Adapted from Nindl et al. (2013).

2.1 Physical performance characteristics

The modern-day battlefield places high demands on a soldier’s physical performance. Military personnel must have the ability to face a wide spectrum of different tasks that require muscular strength, power and aerobic capacity. The requirements depend on the warfighter’s MOS, the mission of the warfighter’s unit and whether the warfighter is currently deployed. (Kraemer et al. 2017, 506.) Traditionally, aerobic endurance has been emphasized in the physical training of military personnel. One of the reasons for this is the easy implementation of such training programs for large numbers of soldiers. However, the role of maximal strength and power is increasingly being recognized as one of the most important factors in preparedness for combat. Common tasks that require high levels of strength and power are, for example, load carriage (LC), sprinting, lifting and casualty evacuation. The modern-day operational environment has been described as an “anaerobic battlefield”. (Kraemer & Szivak 2012; Friedl et al. 2015; Nindl et al. 2015.)

In 2013, The National Strength and Conditioning Association’s tactical strength and conditioning program sponsored the second Blue Ribbon Panel on military physical readiness: military physical performance testing. This meeting brought together 20 subject matter experts (SMEs). The SME panel rated 9 common military tasks by the degree to which different fitness components were required in these tasks (table 1). Muscular strength, power, and muscular endurance received the highest scores. (Nindl et al. 2015).

TABLE 1. Assessment by the expert panel of the physical characteristics required in different military tasks. Adapted from Nindl et al. (2015).

	Strength	Power	Muscular endurance	Body composition	Coordination	Balance	Agility	Flexibility	Aerobic fitness	Speed	Reaction time
Jump or leap over obstacles	**7.5**	**9.0**	4.0	6.4	6.9	5.7	6.5	5.9	2.6	5.7	4.0
Move with agility-coordination	4.7	5.4	5.5	5.8	**9.5**	**8.4**	**9.8**	6.1	4.1	6.5	6.6
Carry heavy loads	**8.8**	6.2	**7.5**	5.2	3.7	5.0	2.9	3.3	5.5	2.2	1.6
Drag heavy loads	**9.2**	**7.4**	**7.4**	5.2	4.5	4.8	3.3	3.8	5.2	2.7	1.6
Run long distances	3.8	3.1	6.9	6.9	3.2	3.2	3.0	3.2	**9.9**	4.0	1.4
Move quickly for short distances	6.0	**7.8**	5.0	6.2	7.0	6.4	**7.8**	4.4	4.0	9.3	6.0
Climb over obstacles	**8.3**	6.5	5.7	6.7	7.0	6.1	6.0	5.9	3.9	4.1	2.2
Lift heavy objects off ground	**9.7**	**7.7**	5.4	5.5	4.8	5.1	2.7	5.0	3.0	2.3	1.6
Load/stow/mount hardware	**7.7**	6.0	6.3	5.0	5.7	5.3	3.4	4.9	3.6	2.6	2.2
Overall mean	**7.3**	6.6	6.0	5.9	5.8	5.5	5.0	4.7	4.6	4.4	3.0

Aerobic fitness, however, can be considered to be as important as strength. Military operations often require long-term moderate-intensity physical activity, for example, tactical road marches, preparing fighting positions, filling and emplacing sandbags, constructing emplacements, loading and unloading trucks, evacuating casualties over long distances, erecting camouflage, land navigation and moving over, through and around obstacles. Soldiers with a higher VO_2max are able to perform these tasks with a relatively lower intensity, fatigue more slowly and recover faster. Aerobic fitness is also important for injury prevention and overall health. (Friedl et al. 2015.)

Pihlainen et al. (2014) studied cardiorespiratory responses during military tasks in field conditions. 15 male conscripts participated in the study. The VO_2max of the subjects was 48.1 ± 4.4 $mL \cdot kg^{-1} \cdot min^{-1}$, measured by a progressive running test protocol on a treadmill with indirect calorimetry. Unloaded (M1) and loaded marching (M2), artillery field preparation (AFP), and digging of defensive positions (D) were monitored. O_2 consumption during military tasks was measured with a portable indirect calorimetry system. The relative oxygen uptake (% VO_2max) during M1, M2, AFP, and D were 42 ± 7 %, 47 ± 6 %, 37 ± 6 %, and 51 ± 9 %, respectively. The authors concluded that the minimum requirement of VO_2max for army soldiers seems to be 45 to 50 $mL \cdot kg^{-1} \cdot min^{-1}$. Epstein et al. (1988) demonstrated that when the LC work intensity is above 50 % of VO_2max, energy expenditure is gradually elevated, which will lead to fatigue over time. Therefore, high levels of aerobic fitness are important, as it allows soldiers to work at a lower percentage of VO_2max. (Pihlainen et al. 2014.)

Excessive endurance training, such as basic military training, may not be optimal for the development of strength and power. This is called the interference effect, first described by Hickson (1980). Nindl et al. (2007) reported physical performance decrements after the intensive 8-week U.S. Army Ranger training course. 50 young healthy male soldiers experienced an approximately 20 % decline in strength and power, significant losses in fat and fat-free mass, and negative alterations in circulation concentrations of testosterone, IGF-1, and cortisol. Santtila et al. (2009) found that a large amount of endurance-based military training interfered with strength adaptations during basic training. Furthermore, basic training as such significantly improved VO_2max levels, whereas additional benefits with added endurance training were not observed. The authors concluded that strength training should play an important role in military training, with some decreases in the amount of endurance training. Groeller et al. (2015) studied the effects of military-specific physical training during initial employment training, which is the period immediately after basic training. The additional training was not sufficient to elicit positive adaptations or even maintenance of physical performance. All these aforementioned studies suggest that it is difficult to improve physical fitness during military training, especially strength and power. This is due to the excessive amount of physical stress, and/or insufficient exercise stimulus and the lack of individual training programs. (Nindl et al. 2007; Santtila et al. 2009; Groeller et al. 2015.)

Body composition is associated with physical performance capabilities, but assessing a soldier's performance through body composition is problematic. Traditional military fitness tests (e.g. running, push-ups, sit-ups) favor soldiers with lower body mass, because it is possible to achieve excellent results with a low level of absolute fitness (Vanderburgh 2007). Using body mass index (BMI) to assess military performance can be misleading, as a higher BMI has been found to be beneficial in some physical performance characteristics, such as muscular power and muscular strength (Pierce et al. 2017). BMI may misclassify individuals with large muscle mass as overweight or obese, because the body fat percentage is not taken into account (Ode et al. 2007).

The purpose of body composition standards in soldiers is not to predict performance, but to motivate fitness and nutrition habits that promote military readiness. The body composition of soldiers must be considered from a health, performance, and military appearance point of view (figure 2). The assessment of body fat is necessary to distinguish large but not obese individuals. The measurement of waist circumference is a common method, and it is more related with health risks than total body fat, body weight, or body size (BMI) (Friedl 2004; Friedl 2012.)

FIGURE 2. The optimal body composition of soldiers falls in between the extremes of military appearance and health point of view. NHLBI = National Heart, Lung, and Blood Institute (U.S.). Adapted from Friedl (2012).

Women are typically shorter, have lower body mass, lower bone mass (20 %), greater body fat (20 %) and less lean body mass (30 %) than men. Men also have a larger proportion of fast-twitch muscle fibers. Understanding sex and age-specific differences in physical performance is important for the development of appropriate training programs. Male soldiers outperform their female counterparts in cardiorespiratory and muscular performances, which is concerning especially amongst the most demanding MOSs, such as infantry. An increasing number of previously restricted combat-centric occupations have been opened to women, which presents new challenges to the physical demands. A proper physical fitness assessment will help to identify specific needs to improve military readiness. (Greeves 2015; Dada et al. 2017.) Resistance training appears to be particularly important for reducing gender differences and improving military performance in women. Untrained women showed significant improvements in occupationally relevant military tasks after 6 months of periodized progressive resistance training. (Kraemer et al. 2001.)

Nindl et al. (2017) studied the effects of a 24-week physical training program on women's military occupational performance. The training program included resistance training, long-distance running, backpacking, and specialized drills. The results suggested that women can significantly improve physical performance in demanding military tasks following 6 months of combined strength and endurance training. Also, the interference effect on strength gains was minimal, which could be explained by the adequate duration of the training period. Hendrickson et al. (2010) reported similar findings. In recreationally active women, 8 weeks of non-linear, periodized combined strength and endurance training resulted in similar improvements in strength and power when compared to strength training alone. Aerobic capacity also improved to the same extent when compared to endurance training alone. The benefits of concurrent training have also been reported in male soldiers (Kraemer et al. 2004). In conclusion, combined training seems to be superior to improve military performance when compared to strength or endurance training alone.

Allison et al. (2017) found greater ankle inversion and eversion strength, anaerobic and aerobic capacity, and agility to predict ground combat MOS school graduation in female Marines. The purpose of the study was to determine which demographic, musculoskeletal, and physiological characteristics predict graduation from infantry and vehicle ground combat

MOS schools in female Marines, as these MOS schools were not previously open to women. The physical demands in these occupations are high, and female Marines may have a higher risk of injury and discharge. These findings support existing knowledge about the importance of aerobic capacity, and provide new evidence of the importance of ankle strength, anaerobic capacity, and agility. All of these are modifiable characteristics, which can be improved with specialized training programs.

2.2 Load carriage

Soldiers are often required to carry heavy external loads in military operations. LC places high demands on physical performance, decreases mobility, and increases the risk for injury. Prolonged LC is also found to decrease cognitive performance in soldiers (Eddy et al. 2015). The amount and weight of equipment has increased throughout history, which is due to the need of body armour, firepower, and communications. The total weight of a modern warfighter's combat gear can exceed 50 kg. (Knapik et al. 2004.)

The decline in soldier's mobility is a critical element during high-intensity tactical combat movements. During such tasks, survival is influenced by susceptibility and vulnerability to enemy fire. Wearing body armour equipment reduces vulnerability, but in contrast, increases the susceptibility of being hit. Billing et al. (2015) studied the impact of an external load on susceptibility during tactical combat movements. Participants completed a break contact simulation (five 30-m sprints every 44 seconds) and a fire and movement simulation (sixteen 6-m bounds every 20 seconds). Both simulations were completed in five different load conditions ranging from 9.8 to 30.3 kg. Exposure time significantly increased with added load in both simulations, each kilogram representing a 0.8 and a 1.1 % increase during break contact and fire and movement, respectively. However, the results presented large inter-individual differences in response to external load, suggesting that tailored training is needed to optimize performance in tactical combat movements. (Billing et al. 2015.)

There are several potential risk management strategies for LC. Due to the nature of military tasks, eliminating the soldier's load is not a viable solution at this stage. In terms of engineering controls, load reduction through technological advances has not been successful. Replacing heavier equipment with lighter alternatives is not considered an option, as there is no known substitution for load weight. However, unit and subunit commanders can reduce the level of risk at a given load using risk modifiers (speed of march, terrain, distance, and duration). Administrative controls are also needed to prescribe work procedures to reduce the soldier's exposure to risks. In addition, an evidence-based conditioning program needs to be developed, with progression to meet task requirements. (Orr & Pope 2015.) Taipale et al. (2015) studied the effects of active and passive recovery on maximal isometric leg extension force and hormone concentrations after LC, and found no difference between these recovery protocols. The authors concluded that LC performed at an intensity below anaerobic threshold does not require specialized recovery protocols. (Taipale et al. 2015.) However, the duration of the LC was only 50 minutes, so these findings cannot be generalized for all LC tasks.

Aerobic training alone seems to be insufficient to improve performance in LC tasks. Improvements in the 3.2 km loaded-run performance have not been demonstrated with aerobic training alone in men (Kraemer et al. 1987; 2004) and in women (Kraemer et al. 2001). However, there are conflicting findings. A 13 % improvement has been reported in the 3.2 km LC time with aerobic endurance training alone in women. The effect of the training might have been lower than this, as the control group also showed a 7 % improvement, suggesting that there was a large learning effect. (Hendrickson et al. 2010.) The most effective way to improve LC performance seems to be in combined strength and endurance training (Kraemer et al. 2004; Knapik et al. 2012; Hendrickson et al. 2010).

Adding one weekly progressive LC exercise is found to augment the largest overall improvements. This also follows the specificity principle of exercise (i.e. the training routine is similar to the desired outcome). Furthermore, it appears that upper body strength and muscular endurance is more important than lower body strength and muscular endurance. Hip-belts in backpacks can effectively distribute some of the weight to the lower body, but most of the weight rests on the shoulders. This suggests that resistance training for enhancing LC performance should focus on the upper body. (Knapik et al. 2012.) In contrast, leg lean mass is found to be a key anthropometric characteristic in stretcher, jerry can, and kettle bell carry tasks (Beck et al. 2017). It is important to note that the nature of these tasks differs from the prolonged LC, as the median carry distances in this study ranged from 265 m to 650 m. When considering the most important physical characteristics in LC, the duration of the tasks must be taken into account, as the demands are different in short carry tasks (e.g. stretcher carry), the 3.2 km loaded run, and prolonged long-distance LC.

As a part of military operation planning, the assessment of LC capacity is challenging. Numerous personal and task-related factors affect the energy cost and/or relative intensity of the LC task. Existing guidelines for the maximum acceptable work duration are based upon unloaded conditions and their accuracy for LC tasks are yet to be validated. Despite the fact that personal characteristics (e.g. muscle mass, VO_2max) are important in LC, task characteristics (e.g. walking speed, terrain, external load) are more likely to affect the maximum acceptable work duration. Compared to personal characteristics, task characteristics are also more modifiable during military operations. However, the operational environment and mission requirements may change rapidly. The nature of military operations and the lack of valid guidelines makes precise planning of LC tasks difficult. Even with some limitations, using the existing maximal acceptable work duration model as a robust assessment in LC tasks, is likely to far outweigh having no planning tool at all. (Drain et al. 2016.)

2.3 Injury prevention

Military personnel are exposed to a high risk of both combat and non-combat related injury. When a soldier becomes injured, it always affects the entire unit, because an essential element of the team is now missing. It is important to understand the types of injuries involved in order to create appropriate risk management strategies. Musculoskeletal injury, typically in the lower back and lower extremities is the most common type of injury in both deployed and non-deployed military personnel. The most common causes of such injuries are overuse, LC and vigorous physical training. (Nindl et al. 2013; Jones et al. 2017.)

Within a military context, several physiological factors are associated with the risk of injury. The causes of injury are often multifactorial, which makes it difficult to distinguish independent variables. Lower aerobic capacity has been identified as a risk factor for injuries in many studies (Knapik et al. 2001; Knapik et al. 2004). Rappole et al. (2017) concluded that older age and poor aerobic fitness are strong predictors of injuries in operational U.S. Army soldiers. The study was also the first to suggest that BMI is less significantly associated with injury than age or fitness. However, interpretation of this result is difficult, since these three variables are interrelated.

Jones et al. (2017) examined the combined effects of physical fitness and body composition on the risk of training-related musculoskeletal injuries among U.S. Army trainees. The highest risk occurred in males and females who had both the lowest levels of aerobic fitness and the lowest levels of BMI. The most aerobically fit soldiers with an average or over-average BMI were at the lowest risk of injury. The authors recommend that recruits with extremely low BMI should be monitored, as this group could represent individuals with reduced muscle mass levels and higher risks of injury.

Unlike aerobic fitness, the association of muscle strength to injury risk is not that well established. However, the importance of strength in injury prevention cannot be overlooked, as it is a key factor in military performance (Nindl et al. 2015). Occupational task demands may exceed the work capacity of the soldiers resulting in injury. Carriage and lifting of external loads are found to be associated with the highest incidence of injury during deployment, which is also the main reason for the loss of duty days. (Roy et al. 2012a; 2012b.) In this context, resistance training is an essential part of injury prevention, as it will increase the soldier’s ability to handle external loads (Friedl et al. 2015).

Nagai et al. (2017) examined physiological and neuromuscular characteristics as predictors of preventable musculoskeletal injuries. A total of 491 soldiers of the U.S. Army’s 101st Airborne Division participated in the study. The baseline laboratory testing included body composition, aerobic capacity, anaerobic power/capacity, muscular strength, flexibility, static balance, and landing biomechanics. Soldiers with a previous injury history or soldiers who sustained non-preventable injuries were excluded from the study. In addition, female soldiers were also excluded due to the small sample size, resulting in a total of 275 male soldiers for statistical analysis. Within the year following the baseline testing, 162 soldiers remained injury free and 113 soldiers developed at least one injury. The study showed static balance and peak anaerobic power to be significant predictors of musculoskeletal injuries, being possibly the first paper to report such results. However, due to several limitations of the study, the authors recommend interpreting the results with some caution. Similar demographics and physical fitness levels of the participants and the small sample size might limit the generalizability of the results.

Female sex is considered a risk factor for injuries in military setting. In general, the reason for this is considered to be women's lower average physical capacity, which is why they have to perform the same tasks with a relatively higher intensity (Nindl et al. 2016). For example, during basic combat training women have been found to have twice the risk of being injured as compared to men (Knapik et al. 2001). Gender, after controlled for fitness, is reported not to be significantly associated with injuries during basic combat training. This finding suggests that physical fitness, particularly cardiovascular fitness, is the key risk factor for injuries. (Bell et al. 2000.) In contrast, Roy et al. (2012b) concluded that female sex was indeed a risk factor for injury during deployment. In this study, more injuries were observed among women despite the fact that the loads worn by female soldiers were proportionally less than those by men. In addition, women performed less physically demanding tasks.

Wardle & Greeves (2017) conducted a systematic review of the most effective (physical training-related) injury prevention strategies for military personnel. The aim of the study was to update the existing evidence base reported by Bullock et al. (2010). Sixty-one articles were selected, and categorized into six strategies: conditioning, footwear modifications, bracing, physical activity volume, physical fitness, and leadership/supervision/awareness. In combination with the study by Bullock et al. (2010), the findings support a reduction in physical activity volume, and bracing for high-risk activities to reduce injury rates. Using multiaxial, neuromuscular, proprioceptive and agility training did not seem to be an effective strategy to reduce injuries, although no negative effects of such programmes were observed. Footwear modifications showed no beneficial effects on injury prevention in military cohorts. A high level of pre-accession physical fitness and leadership/supervision/awareness were deemed to be important contributors in reducing injury rates. While improving physical fitness, excessive time on foot should be avoided. With the existing evidence, current injury prevention strategies seem to be suitable for both men and women. (Wardle & Greeves 2017.)

2.4 Environmental factors

Military personnel must be able to operate in different types of extreme conditions. Exposure to heat, cold, and high altitude affects physical performance and increases injury risk (Margolis et al. 2014; Taylor 2015). Proper prevention and mitigation strategies are needed to maintain combat readiness. When possible, an acclimatization period allows the body to make physiological adjustments to better tolerate heat and altitude. However, the body cannot acclimatize to cold environments. (Nindl et al. 2013.)

Heat illness can occur during prolonged intense activity, especially in hot and humid conditions. These illnesses include exercise associated muscle cramping, heat exhaustion, or exertional heatstroke. Heat exhaustion and muscle cramps are typically related to fatigue, body water and/or electrolyte depletion, and/or a failure of the thermoregulatory system. Exertional heatstroke is the most severe form of heat illness and is defined as a core body temperature above 40 °C associated with central nervous system disturbances and multiple organ system failure (Epstein & Yanovich 2019). An acclimatization period of four to 14 days is recommended to reduce the risk of heat illness. (Armstrong et al. 2007; Parsons et al. 2019.)

Lower physical fitness, higher BMI (obesity), and female sex are associated with an increased risk for heat illness (Bedno et al. 2010; Alele et al. 2020) Sex differences in males and females in thermoregulation are mainly related to anthropometric factors and fitness, rather than sex, *per se* (Corbett et al. 2020) Hyperthermia is not generally the primary cause of heat exhaustion in soldiers, but indeed of cardiovascular insufficiency. During thermal stress, elevated cardiac output is required to maintain oxygen delivery, blood pressure regulation, and waste removal. With the depletion of blood volume due to excessive sweating, homeostatic compensation is challenged and can fail to sustain mean arterial pressure, resulting in systematic hypotension (Taylor 2015; Epstein & Yanowich 2019)

Nevertheless, preventive measures are similar to both hyperthermia and its precursors, cardiovascular insufficiency, and hypotension. Pre-exposure and auxiliary cooling are effective strategies, while wearing undergarments that wick away sweat offer minimal support (van den Heuvel et al. 2010; Caldwell et al. 2012). Novel heat mitigation strategies in the military are, for example, ingestion of ice slurry, arm immersion cooling, and microclimate cooling. Personal microclimate cooling refers to garments worn under protective clothing. Different systems have been developed, such as air-cooled, liquid-cooled, ice-packed vests, and phase change materials. In a military context, liquid cooled garments are considered as the most beneficial. All of these 3 methods are effective and suitable for use in different situations (Lee et al. 2015). Other strategies include acclimatization, proper hydration, and limiting the duration of continuous work periods if possible (Alele et al. 2020). Personal protection and equipment must come first, impeding optimal heat evaporation from the skin. (Taylor 2015.)

Cold stress affects physical performance capabilities, morale, and manual performance tasks that require fine motor skills. Exposure to extremely low temperatures or cold-wet environments may cause cold injuries and/or hypothermia. Although the body has various physiological thermoregulatory systems to maintain internal temperature, countermeasures to avoid cold stress mainly include the proper use of protective equipment. (Nindl et al. 2013.) Although the majority of cold adaptation is behavioural, some physiological adaptations may occur; if a person is habituated to cold, shivering and vasoconstriction are blunted and the sensations of cold less intense (Mäkinen 2007).

At a high altitude, less oxygen is available to the body because of the low barometric pressure. Hence, physical performance is decreased, especially aerobic capacity (Fulco et al. 1998; Coffman et al. 2020). Acute mountain sickness is the most common altitude sickness, with minor symptoms such as headache, fatigue, and nausea, while others can be life threatening and need immediate medical attention. The most important risk mitigation strategy to prepare soldiers for high-altitude operational environment is acclimatization. However, graded ascent is not always possible due to mission time constraints (Nindl et al. 2013; Lechner et al. 2018.)

3 ANAEROBIC PERFORMANCE TESTING

3.1 Interaction of energy systems during exercise

Various methods to quantify anaerobic energy release are not precise. Many of the concepts have often led to the misconception that the energy systems operate in discrete time periods. Aerobic energy release can be quantified, as there is a direct relationship between VO_2max and aerobic production of ATP. In contrast, the existing methods to measure anaerobic metabolism provide only an assessment of anaerobic energy production during exercise. Several attempts have been made to establish valid tests. Historically, the peak blood lactate concentration has been used as an indicator of anaerobic energy release. (Gastin 2001.) However, this method provides only an estimate of anaerobic energy production, as blood lactate concentration has been demonstrated to be lower than muscle lactate concentration (Jacobs & Kaiser 1982; Tesch et al. 1982). Other potential methods include muscle biopsy samples, power output from ergometric tests, and the concept of oxygen deficit and oxygen debt. All of these methods have several limitations, but can provide an acceptable estimation of anaerobic metabolism. (Gastin 2001.)

All forms of biologic work need the breakdown of the energy-rich compound adenosine triphosphate (ATP). There are three metabolic pathways to resynthesize ATP. The anaerobic system is divided into the immediate (alactic) energy system and the glycolytic (lactic) system. The aerobic (oxidative phosphorylation) system has a large capacity to produce energy in the presence of oxygen, but with a relatively slow rate. The anaerobic system is capable of producing ATP at high rates, responding to the energy demands of intense exercise. (Gastin 2001; McArdle et al. 2014, 134.)

The immediate energy system refers to energy derived from intramuscular high-energy phosphates ATP and phosphocreatine (PCr). Each kilogram of skeletal muscle contains 3 to 8 mmol of ATP and 4 to 5 times more PCr. The immediate energy system is dominant in short-duration maximal efforts lasting less than 10 seconds. In order to continue exercise at high intensity, energy to phosphorylate ATP comes mainly from stored muscle glycogen via the anaerobic glycolysis, which is the dominant energy pathway in maximal exercises lasting less than 75 seconds. After this, aerobic energy production (oxidative phosphorylation) is the dominant source of energy (table 2). (Gastin 2001; McArdle et al. 2014, 134–145.)

TABLE 2. Estimates of anaerobic and aerobic energy contribution during selected periods of maximal exercise. Modified from Gastin (2001).

Duration of exhaustive exercise (s)	% Anaerobic	% Aerobic
0–10	94	6
0–15	88	12
0–20	82	18
0–30	73	27
0–45	63	37
0–60	55	45
0–75	49	51
0–90	44	56
0–120	37	63
0–180	27	73
0–240	21	79

Many authors describe short-duration exercise as 'anaerobic' and longer efforts as 'aerobic'. Chamari & Padulo (2015) identified the incorrect use of these terms, as every effort relies on the simultaneous participation of all three energy pathways. All-out efforts lasting less than 1 s to around 6 s are not dependent only on the phosphagen pathway, but also partially on glycolysis. Therefore, these efforts cannot be described as purely anaerobic alactic exercises. In addition, longer all-out efforts of less than 1 minute cannot be described as anaerobic or anaerobic lactic, as they involve a large proportion of energy derived from oxidative phosphorylation. Furthermore, purely aerobic exercises do not exist when a minimum of intensity is put into the efforts. The use of the following terms is suggested (Chamari & Padulo 2015):

1. 'Explosive efforts': all-out exercises with a duration of up to 6 s (predominance of 'phosphagens' pathway').

2. ‘High intensity efforts’: all-out efforts lasting from 6 s to 1 min (predominance of the ‘glycolytic pathway’ in addition to the ‘phosphagen’s pathway’ and ‘oxidative phosphorylation’).

3. ‘Endurance intensive efforts’: exercise with a duration exceeding 1 min (predominance of ‘oxidative phosphorylation’).

3.2 Anaerobic tests

As reviewed by Vandewalle et al. (1987), there is no anaerobic test which enables the accurate measurement of the different components of anaerobic metabolism. Anaerobic tests are divided into tests measuring anaerobic power and anaerobic capacity. The term power means the rate at which work is done, and anaerobic capacity is described as the maximal amount of ATP resynthesized via anaerobic metabolism (Green 1994). Anaerobic power reflects energy-output capacity of the immediate (alactic) energy system, while anaerobic capacity depends more on glycolytic (lactic) metabolism. It is difficult to distinguish the contribution of alactic and lactic energy production, as these metabolisms act simultaneously at the very beginning of maximal effort. In addition, the involvement of aerobic metabolism and the role of mechanical efficiency limit the validity of anaerobic tests. (Vandewalle et al. 1987.)

The anaerobic tests include force-velocity tests, jumping tests, running tests, staircase tests, and cycle ergometer tests. Furthermore, the measurement of maximal accumulated oxygen deficit (MAOD) is a widely used method to determine anaerobic capacity. These tests can be performed by a variety of test protocols, depending on the purpose of the test. In athletes, the tests must be sport-specific and reliable enough to detect changes in anaerobic exercise performance. The magnitude of the training effects can be smaller than the coefficient of variation of the test. (Vandewalle et al. 1987.)

3.2.1 Force-velocity tests

The term ‘strength’ in human performance has been described in various ways. Knuttgen & Kraemer (1987) proposed that ‘strength’ should refer to the maximal force a muscle or muscle group can generate at a specified velocity during concentric, isometric, or eccentric muscle contraction. If the velocity of muscle contraction is high, the produced force is relatively low. In contrast, when the force is high, the velocity is low. This force-velocity relationship was first introduced by Hill (1938). Maximal power is achieved with an optimal load (figure 3). Despite some limitations, it is possible to measure maximal anaerobic power with protocols such as cycle ergometer tests, or staircase tests. In order to achieve maximal power, it is necessary to assess the relationship between force and velocity. (Hill 1950; Vandewalle et al. 1987.)

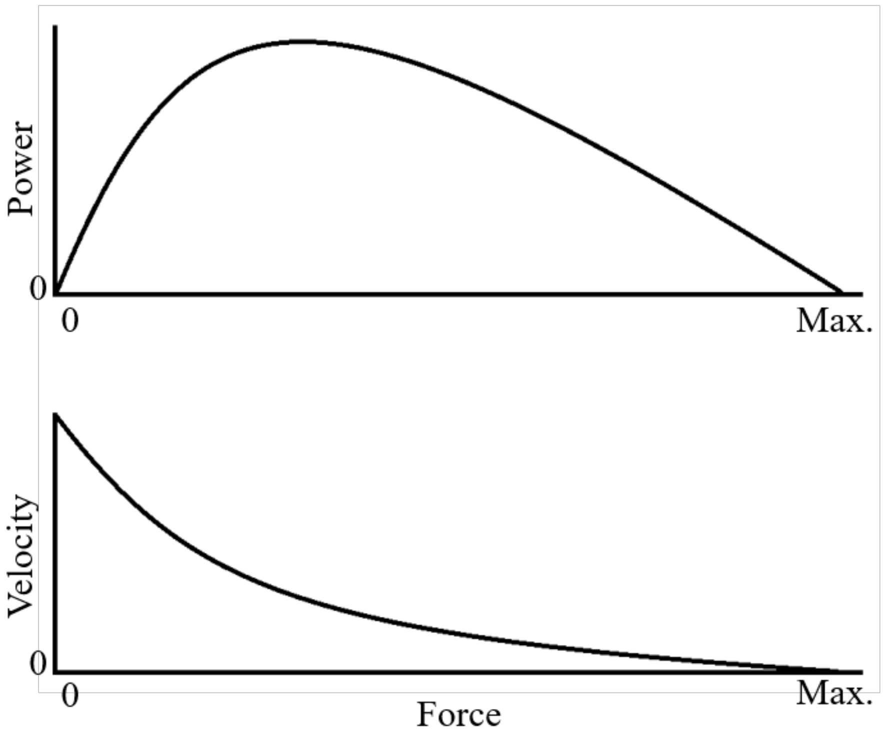

FIGURE 3. Force-power relationship and force-velocity relationship. Adapted from Hill (1950).

However, the measurement of muscular 'power' with short-term movements, like a maximal effort vertical jump, has faced criticism. The association with jump height and true mechanical power of the working muscles has been found to be weak. Dynamic short-term activities like jumping, throwing, and striking should be considered as a neuromuscular (anaerobic) performance. The term 'power' as a performance variable is suggested to be used only with true measures of mechanical power. (Knudson 2009.)

3.2.2 Wingate anaerobic test

One of the most popular tests to measure anaerobic performance is the Wingate anaerobic test (WAnT). The test is conducted using a cycle ergometer and consists of cycling at maximal effort for 30 seconds. The resistance is calculated according to the body weight of each individual. Three variables are determined from the test: the highest 5-second power output (peak power), mean power, and the rate of power decrease (fatigue index). The original test was developed at the Wingate Institute in Israel during the early 1970s (Inbar et al. 1996). Since then, it has undergone several modifications such as extended duration (Lericollais et al. 2010) and the use of different resistances (Jaafar et al. 2014).

Beneke et al. (2002) examined the metabolic profile of WAnT. The aim of the study was to measure the contribution energetics with respect to working efficiency and performance. Aerobic energy production was measured using a spirometric system. The energy produced from anaerobic alactic metabolism was estimated from the fast component of EPOC and the energy equivalent of O_2. Net energy produced from anaerobic lactic acid metabolism was determined from the net lactate accumulation, body mass, and O_2-lactate equivalent. An increase of 1 mM in blood lactate concentration is assumed to reflect energy released by the consumption of 3 ml O_2 per kg of body mass (di Prampero & Ferretti 1999). Fractions of the energy from the aerobic, anaerobic alactic acid, and anaerobic lactic acid metabolism were 18.6 ± 2.5 %, 31.1 ± 4.6 %, and 50.3 ± 5.1 %, respectively ($p < 0.01$). Energy from the anaerobic lactic acid metabolism explained 83 % and 81 % of the variance of peak power, and mean power, respectively. The authors concluded that the maximal rate of glycolysis is reached between the 5^{th} and 10^{th} second of WAnT, and that the anaerobic lactic acid metabolism is the dominant energy source up until to the end of the test. However, the contribution of aerobic metabolism was shown to have a significant role. At WAnT termination, VO_2 was between 60 % and 80 % of VO_2max. (Beneke et al. 2002.)

As with many other anaerobic tests, the validity of WAnT has been criticized. It is likely that the 30-second work duration is too short to fully exhaust the anaerobic energy sources, and that the involvement of aerobic metabolism is not accounted for. Performance in WAnT is affected by individual physical characteristics. The ability to maintain power (fatigue index) during WAnT could be more related to the anaerobic capacity. Individuals who are able to produce a high rate of energy anaerobically (peak power) are not necessarily able to produce a large amount of energy anaerobically (mean power). A higher proportion of type II muscle fibers has a significant correlation with peak power during WAnT. (Minahan et al. 2007.)

Jaafar et al. (2014) compared the effects of two different braking forces (8.7 and 11 % BM) on the reliability of WAnT performance, peak blood lactate, peak heart rate, and the rate of perceived exertion (RPE). Sixteen active male subjects participated in the study. The reliability of the measured variables was not affected by the braking forces. Peak power, mean power, and fatigue index were enhanced with 11 % BM. Previous studies suggest that the mean peak pedal rate of around 120 rpm is optimal for maximal power output. In this study, the mean peak pedal rate at 8.7 and 11 % BM was 122 ± 16 and 144 ± 17, respectively, indicating that a lower load was not optimal. In conclusion, a braking force of 11 % BM is recommended in active men.

3.2.3 Maximal accumulated oxygen deficit

Measuring the maximal accumulated oxygen deficit (MAOD) is considered as the gold standard test to determine anaerobic capacity (Noordhof et al. 2010). The method is based on the estimation of the linear relationship between individual O_2 uptake and treadmill speed (figure 4). This relationship is extrapolated to the O_2 demand in supramaximal exercise intensities. The MAOD is calculated from the difference between the estimated O_2 demand

and the actual O_2 uptake, which represents the energy that is derived anaerobically. (Medbø et al. 1988.)

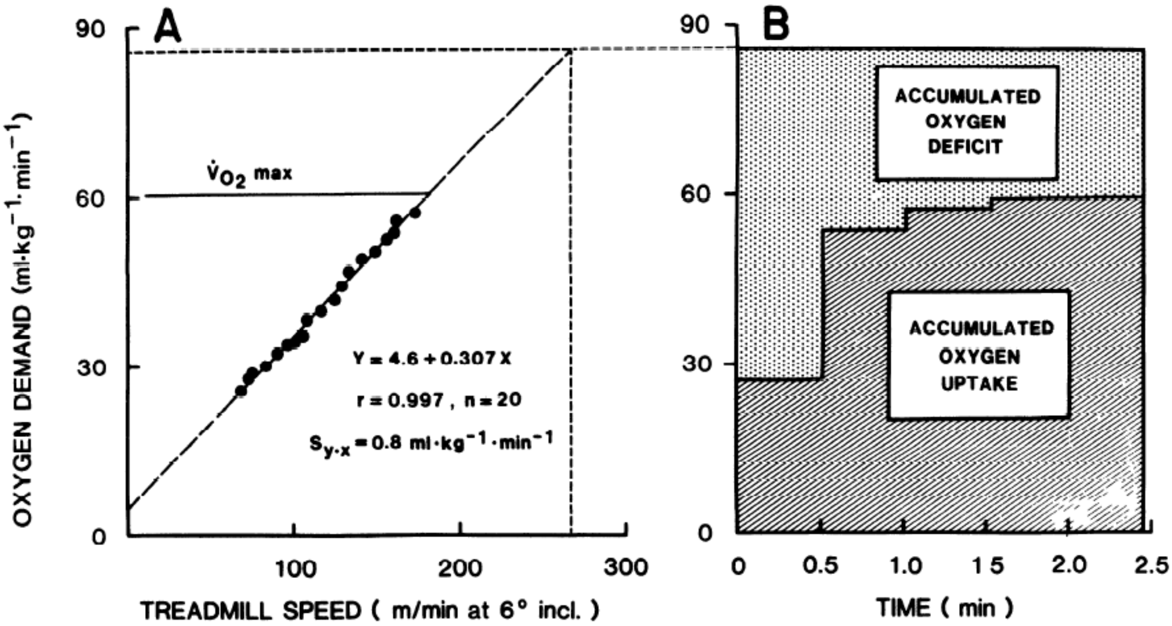

FIGURE 4. Principles for determining O_2 deficit. A: relationship between treadmill speed and O_2 demand. B: accumulated O_2 deficit is calculated as difference between O_2 demand and O_2 uptake. Adapted from Medbø et al. (1988).

A review by Noordhof et al. (2010) evaluated the validity and reliability of the MAOD. Previous studies have shown that the MAOD may not be a reliable measure of anaerobic capacity. The authors identified a limitation in the original protocol used by Medbø et al. (1988) to determine the linear relationship between O_2 uptake and exercise intensity, as O_2 uptake shows a secondary increase after 3 minutes of exercise at the highest submaximal exercise intensities, which will lead to a non-linearity of the O_2 uptake-intensity relationship. Another concern is related to the inclination of the treadmill, as the secondary increase in O_2 uptake has shown to be significantly higher during uphill running in comparison with horizontal running. Furthermore, using different supramaximal exercise protocols may affect the test outcome. In order to improve the validity and reliability of the MAOD, the authors suggest the following methodology. It is necessary to implement relatively short (4 min) exercise bouts at different intensities for the construction of a linear relationship. It is advisable to use horizontal running instead of uphill running, and the supramaximal exercise protocol should be specific for the athlete's event.

The limitation of the wide use of the MAOD is the time required to apply several sub- and supramaximal exercise sessions. Zagatto et al. (2016) studied the validity and reliability of using a single supramaximal exercise session ($MAOD_{ALT}$) to determine MAOD. The participants underwent 8 exhaustive supramaximal running efforts with intensities ranging from 100 to 150 % VO_2max. The energy derived from the phosphagen metabolic pathway was estimated from the fast component of EPOC, and the O_2-lactate equivalent was used to estimate the involvement of the glycolytic metabolic pathway. The $MAOD_{ALT}$ results were compared to the MAOD result which was measured with the original protocol, and showed no statistically significant difference. The $MAOD_{ALT}$ at 115 % VO_2max presented the highest correlation with MAOD, and was concluded to be a valid and reliable method for assessing anaerobic capacity in a single supramaximal exercise session.

3.2.4 Jumping tests

The maximum vertical jump has been used in order to assess neuromuscular performance. Several different types of jumps have been applied, including the countermovement jump, squat jump, drop jump, and continuous jumping. As with other tests, the tests represent only an assessment of muscle power, since muscle power cannot be directly measured in vivo. The muscle action can affect the measured outcome, as the squat jump includes concentric-only muscle action, while other types of jumps include a stretch-shortening cycle of muscle action (Markovic & Jaric 2005). Maximal muscular power is achieved with an optimal load, according to the force-velocity relationship (Hill 1938). Jaric & Markovic (2013) presented evidence that the optimal load in maximum vertical jumps is one's own body mass, regardless of the strength of the lower limb muscles. In contrast, Matic et al. (2015) showed that in drop jumps, the drop height should be adjusted based on maximal muscle strength, as stronger individuals generated maximal values of power output on a higher drop height compared with the weaker individuals. (Markovic & Jaric 2007; Jaric & Markovic 2013.)

The muscle action in cycling is mainly concentric, and that is why WAnT is not a specific anaerobic test for most athletes. Many sports require a combination of eccentric and concentric muscular actions (i.e. the stretch-shortening cycle, SSC) such as running and jumping. Bosco et al. (1983) developed a simple test for the measurement of mechanical power during a series of vertical rebound jumps. The test consists of repeated vertical jumps for 60 seconds. The flight time is measured with a digital timer, and the number of jumps is counted. Mechanical power is calculated from the measured parameters. The relationship between this mechanical power and the 60-second Wingate test (r = 0.87) and 60 m dash (r = 0.84) were high. The reproducibility is also high (r = 0.95). Therefore, this test seems to be suitable to assess anaerobic performance in athletes experiencing the stretch-shortening cycle in their sport.

In a long (> 45 s) anaerobic test, the participant may consciously limit the performance intensity. In a 30-second test, the produced power is generally greater than during the first 30 seconds in the longer test. Based on these limitations, Dal Pupo et al. (2014) examined a modified 30-second continuous vertical jumping test. The test showed high test-retest reliability for the maximal jump height (ICC = 0.94), mean vertical jump height (ICC = 0.98) and fatigue index (ICC = 0.87). The mean height of the first four jumps correlated with the peak power of the 30-second WAnT (r = 0.57), and the mean jump height with the mean power of the WAnT (r = 0.70). The fatigue index of the jump test showed a moderate correlation with the fatigue index of the WAnT (r = 0.43). The 30-second jumping test may be more feasible than longer tests, as it offers more operational facilities with less discomfort.

3.2.5 Staircase tests

Margaria et al. (1966) developed a staircase test to measure anaerobic power. After a short 2-meter run on a flat surface, subjects are required to run up ordinary stairs (around 17.5 cm), 2 steps at a time. The time to cover an even number of jumps is measured. In order to determine power, the time taken from the fourth to the sixth jump (70 cm height) is recorded.

Several modifications of the test have been proposed. Kalamen (1968) suggested a protocol where running on a flat surface is 6 m and 3 steps are climbed at a time instead of 2. A study by Huskey et al. (1989) supported the advantage of a longer approach distance. Using an additional weight equal to 40 % of BM has been demonstrated to improve power output by 16 % (Caiozzo & Kyle 1980). The standard formula to calculate the power output in the Margaria staircase test is suggested to overestimate the power output of larger subjects and underestimate the power of smaller ones. A modified formula is proposed to provide a body size independent index of power, which allows establishing standards and a comparison of individuals. (Nedeljkovic et al. 2007.)

Hetzler et al. (2010) examined a modified Margaria-Kalamen test for football players. Subjects were required to complete a total of 25 trials with approximately 30 seconds of rest between trials. Each trial consisted of 20 steps. The results showed an increase in peak power throughout the 25 trials. The test-retest reliability was acceptable (ICC = 0.73). In conclusion, a modified test duration with multiple trials was sufficient in capturing peak velocities in athletes.

Another limitation of the anaerobic staircase test is the skill level of the individual, which may affect the measured outcome. Running stairs at maximal speed requires the skill of transforming horizontal velocity to vertical velocity and the co-ordination of foot placement on the correct steps. It is possible that eliminating the steps may reduce the skill requirement of the test and increase power output. In order to have a more accurate measure of anaerobic power, the use of a ramp instead of stairs is suggested. (Huskey et al. 1989.)

3.2.6 Running tests

Many researchers have used an all-out 300 m running test to assess anaerobic performance (Lima et al. 2011; Angeltveit et al. 2016). However, the validity and reliability of the test has not been studied. Kimura et al. (2014) examined the validity and reliability of the Hawaii anaerobic run test (HART), which consists of a 200 m sprint performed on a standard 400 m track. The duration of approximately 30 seconds seems appropriate in order to measure anaerobic power. The recorded 25 m split times were measured to calculate anaerobic performance variables (peak momentum, mean momentum, and fatigue index). The test reliability was calculated through 2 separate trials. Momentum on the HART was compared with power on the 30 s WAnT in order to determine the validity. The reliability (ICC) for peak and mean momentum was 0.98 and 0.99, respectively. High correlations were found between peak power and peak momentum (r = 0.88), and mean power and mean momentum (r = 0.94). The HART was considered to be a valid and reliable test to measure anaerobic performance.

In contrast with the findings of Kimura et al. (2014), Legaz-Arrese et al. (2011) concluded that performance in the WAnT was not significantly associated with running performance in any distance event. 116 world-class runners participated in the study. Performance in the WAnT was not simultaneously compared to running tests, but runners were classified into groups according to their best competition performance times. Distances ranged from 100 m to a marathon. The limitation of the study was the elapsed time between the WAnT and measurement of running performance. The observed non-significant association between WAnT and running performance may be due to differences in muscle action. In conclusion, the information provided by the WAnT is not useful in evaluating the anaerobic capabilities of elite runners.

Several types of intermittent running tests have been developed. Andrade et al. (2015) studied the use of the running anaerobic sprint test (RAST) as a predictor of anaerobic capacity in young soccer players. The RAST consisted of 6 maximal efforts of 35 m, separated by a passive recovery period of 10 s. The time of each effort and the participant's body mass were used to calculate power, and in order to determine validity the results were compared with MAOD and a 30 s all-out tethered running on a treadmill. The correlation between the MAOD and RAST parameters was not significant. The fifth effort of RAST was associated with the peak and mean power of a 30 s all-out test. The results suggested that RAST is an appropriate test to evaluate anaerobic power, but should not be used to measure anaerobic capacity.

Dardouri et al. (2014) developed a maximal anaerobic shuttle run test (MASRT). The test consisted of intermittent shuttle running between two lines, spaced 20 m apart. The duration of each stage was 20 s with 100 s of passive rest between stages. The initial velocity was set individually, measured from the 20 m multi-stage shuttle run test. The velocity was increased by 0.28 $km \cdot h^{-1}$ for each stage. The validity was determined by examining the relationship between MASRT and the 30 s WAnT. The results showed a good correlation between MASRT and WAnT parameters, and the authors concluded that the MASRT is a valid test for assessing anaerobic capacity. The difference in the total duration of RAST and MASRT may explain why RAST measures anaerobic power, and MASRT is a more appropriate test to evaluate anaerobic capacity. The mean duration of RAST is approximately 30 s (Andrade et al. 2015), while the total duration of MASRT is several minutes.

The maximal anaerobic running test (MART) was developed to measure anaerobic performance of sprint athletes. The MART consisted of 20-s runs on a treadmill with 100-s recovery between runs. The initial speed was 14.6 $km \cdot h^{-1}$ and was increased to 1.37 $km \cdot h^{-1}$ after each run until exhaustion. The inclination was 4° during the whole test. The correlation between most of the corresponding variables of the MART and the 30 s WAnT were low, which may be explained by the difference in muscle action and the total duration of the tests. In addition, the MART includes intermittent submaximal workloads, while the WAnT is a short all-out test. It was concluded that the MART, at least partly, measures the same anaerobic properties as the WAnT, and is a valid test to determine anaerobic work capacity during treadmill running. (Nummela et al. 1996.)

3.2.7 Military-specific anaerobic tests

Bishop et al. (1999) examined physiological determinants of performance on an indoor military obstacle course (OC) test. The 11-item OC included, for example, low crawling, running, climbing, and carrying a 4.2 kg medicine ball. The association between the OC time and body composition, upper and lower body aerobic and anaerobic power, muscular strength, and endurance was studied. Anaerobic power of the upper and lower body was measured with a 30 s WAnT. Lower body and upper body resistance were set at 7.5 % and 2.4 % of body weight, respectively. The results suggested that performance on the OC is very complex and not reliant on any single physical characteristic. (Bishop et al. 1999). The purpose of this study was not to develop a valid military-specific anaerobic test. In addition, the total duration of the OC test was 187 ± 69 seconds, which means that anaerobic metabolism is not the dominant energy source (Vandewalle et al. 1987).

The Danish Armed Forces physical fitness test protocol includes two high-intensity tests. The soldiers can choose one of the two tests. The Danish Military Speed Test includes running between two lines 20 m apart as many times as possible in 30 seconds, followed by 30 seconds of rest. The cycle is repeated 10 times. No studies have been published on the reliability and validity of this test. Similar to the Danish Military Speed Test, the Yo-Yo IR 1 consists of 20 m shuttle runs at increasing speeds, interspersed with a 10-second recovery. The speed is controlled by audio signals. (Malmberg 2011.) The Yo-Yo IR 1 is demonstrated to be a valid measure of aerobic and anaerobic fitness performance (Krustrup et al. 2003; Bangsbo et al. 2008).

Angeltveit et al. (2016) developed a new job-specific anaerobic evacuation (EVAC) test. The purpose of the test was to measure specific anaerobic performance in Special Operations Forces (SOF) operators. The EVAC test consisted of pulling a 70 kg doll wearing a 10 kg plate carrier on a course measuring 10 × 20 m (figure 5). The validity of the EVAC test was studied by comparing the results with performance in the 30 s WAnT, 300 m sprint, and MAOD. The EVAC test demonstrated a strong association with other well-established anaerobic tests. The results also showed the importance of body mass (muscle mass) in the casualty evacuation performance. The EVAC test was considered a valid test to measure specific anaerobic performance in SOF operators.

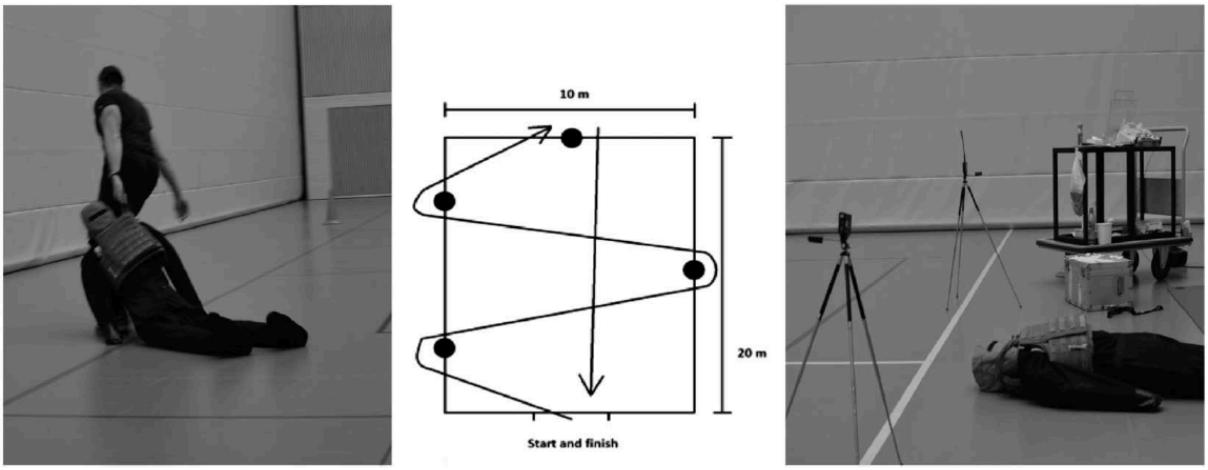

FIGURE 5. Schematic view of the EVAC test. Adapted from Angeltveit et al. (2016).

All of the aforementioned tests are carried out without wearing combat gear. LC significantly reduces the ability to perform tactical maneuvers that require power and agility (Joseph et al. 2018). This issue should be taken into account when developing a proper assessment of military-specific anaerobic performance, as soldiers are expected to be able to complete all tasks while wearing a heavy external load (Taylor et al. 2016).

Harman et al. (2008) found the vertical and horizontal jump and 3.2 km run test to be the best predictors of performance in 4 simulated battlefield tasks while wearing a standard infantry fighting load (18 kg): five 30 m sprints prone to prone, 400 m run, obstacle course, and casualty recovery. The mean times for these tests were 43–84 seconds. The 2-minute sit-up test was associated with performance in the obstacle course, suggesting involvement of abdominal muscles, for example, in the low crawl. Subjects with high body mass performed better in the casualty recovery, but not in other tests. Higher body mass seems to be an advantage when soldiers are required to move relatively heavy loads on the battlefield.

Mala et al. (2015) studied the role of strength and power in high-intensity military tasks under heavy load carriage. A one repetition maximum squat and bench press, and countermovement jumps were used to measure lower and upper body strength and power. The Army Physical Fitness Test (APFT) was also performed, which includes 2 minutes of maximal push-ups, 2 minutes of maximal sit-ups, and a 2-mile run. A high-intensity military task course was performed under unloaded (uniform, boots) and loaded (7–9 kg body armor, 30 kg rucksack) conditions. The course included a 30 m sprint, a 27 m zigzag run, and a 10 m casualty (79.4 kg) drag. All measured strength and power characteristics were associated with the loaded course time. The components of APFT were not significantly associated.

Huang et al. (2018) studied physical fitness predictors of a warrior task simulation test (WTST). The test included 9 events: a) 282-m run, b) low hurdles, c) high crawl, d) negotiating under and over hurdles, e) casualty (81.6 kg) evacuation, f) balance beam ammo carry, g) point-aim-move, h) 91 m shuttle sprint with ammo can, and i) agility sprint. The test was performed while wearing combat boots, a uniform, helmet, and dummy rifle. The mean time to complete the WTST was approximately 4 minutes. Several components of physical fitness were associated with the WTST, including anaerobic and aerobic capacity, body composition, agility, muscular endurance, and fat-free mass.

4 PHYSICAL EMPLOYMENT STANDARDS IN A MILITARY CONTEXT

4.1 Development of physical employment standards

The purpose of physical employment standards is to select individuals who are able to meet the requirements of different military occupations. The physical assessments and standards must be research-based and withstand legal scrutiny (Milligan et al. 2016). In addition, the predictive fitness tests must be valid, reliable, safe, and easy to administer. When developing and validating PESs for the military, five key challenges are addressed: (a) identification of critical tasks and minimum safe and efficient performance levels, (b) involvement of military personnel (SMEs) in the research process, (c) translation of criterion tasks to tests and standards (d) examination of bias in the tests and standards, and (e) implementation, monitoring, and revision of tests and standards. (Reilly et al. 2015; Petersen et al. 2016.)

Critical tasks are defined as the most physically demanding tasks, in which the inability to perform to a certain standard may result in injury/death or compromise a mission outcome. An example of a critical task is casualty evacuation, which can be considered a task that every soldier is required to perform. Other considerations include, for example, the effects of thermal stress (Cheung et al. 2016), load carriage (Taylor et al. 2016), and differences between the sexes (Epstein et al. 2013; Roberts et al. 2016). The physical demands of critical tasks are absolute and are not altered by age, anthropometrics, or gender. Generic fitness tests, such as push-ups, sit-ups, curl-ups, and running tests penalize heavier individuals, and do not measure absolute fitness (Vanderburgh 2007). The use of task simulation tests is recommended, when translating criterion tasks into PESs. There must be a logical relationship between the task simulation tests and the corresponding criterion tasks, which requires the opinion of SMEs in order to achieve face validity. Ideally, generic fitness tests and task simulation tests are combined into task-related predictive tests that are easier to administer than task simulation tests. (Reilly et al. 2015; Beck et al. 2016a)

In the context of developing PESs, it is concluded that all standards involve some subjective aspects. For example, no consensus exists to determine how many critical tasks should be used or in what combination in order to establish a proper task analysis. In addition, potential problems with task simulations have been identified: a) the validity of such a test is dependent on the accuracy of the simulation, b) some critical tasks are difficult to simulate in a way that allows the simulations to be undertaken in a variety of locations, and within controlled environmental conditions, c) there is no way of knowing what percentage of their maximum work rate an individual is working at when they undertake a simulation, d) it is not viable to use simulations requiring significant skill components that should be acquired as a result of performing the job. Due to these and several other limitations, predictive tests and standards are not perfect. A significant amount of research is needed to remove some of the subjective aspects associated with the development of PESs, as they must be legally defensible. (Tipton et al. 2013; Petersen et al. 2016; Milligan et al. 2016.)

4.2 Physical employment standards for the armed forces

Recently, the U.S. Army Research Institute of Environmental Medicine (USARIEM) developed the gender-neutral physical employment screening test to evaluate a recruit's ability to serve in a selected MOS. This study was titled the U.S. Army Physical Demands Study. The purpose was to identify the criterion tasks relevant to the MOS, develop reliable and valid task simulations, and establish a single combination of task-related predictive tests that would predict physical performance on the seven MOSs: infantryman, infantryman-indirect fire, combat engineer, cannon crewmember, fire support, cavalry scout, and armor crewman. Stepwise regressions were used to identify a single combination of the predictor tests that significantly predicted performance on different MOSs. After consideration of the positive and negative features of each test battery, medicine ball put, squat lift, beep test, and standing long jump were selected as the Army's Occupational Physical Assessment Test (OPAT). (Foulis et al. 2017a; Foulis et al. 2017b; Boye et al. 2017; Sharp et al. 2017)

The Canadian Armed Forces developed a new physical fitness assessment called the FORCE (Fitness for Operational Requirements of Canadian Armed Forces Employment) evaluation. Six common and essential military tasks were identified and developed into task simulation tests: building a protective shelter with sandbags, extracting a casualty from a vehicle, building a picket and wire fence, picking, digging, and escaping to cover. These task simulations were translated into simple task-related predictive tests using a 20 kg sandbag: sandbag lifts, intermittent loaded shuttles, sandbag drag, and unloaded 20 m rushes. The standards are age and gender independent. In addition to this measure of operational fitness, health-related fitness is evaluated by estimating aerobic capacity derived from the FORCE evaluation results, combined with a measure of abdominal circumference. (Gagnon et al. 2015; Stockbrugger et al. 2018)

As the inability to perform the most critical tasks may compromise a mission outcome or result in injury or death, several tests have been established for the assessment of these specific tasks. One such task is carrying a casualty on a stretcher. Bilateral jerry can carry, and kettlebell carry have shown to be appropriate tests to predict performance in this time-critical task (Beck et al. 2016b; Beck et al. 2016c). Manual material handling is a common military task, and a significant factor in the occurrence of musculoskeletal injuries. Carstairs et al. (2016) developed a box lift and place test for the assessment of manual handling performance. In the test, the participant was required to lift a 15 kg box onto a 1.5 m platform. After each successful lift, the mass of the box increased by 5 kg until the lift failed.

5 PURPOSE OF THE STUDY

The purpose of this study was to identify the contribution of body composition, maximal oxygen uptake (VO_2max), muscular strength and endurance, and anaerobic performance to performance in the MST. The main focus was on associations with the MST and three different generic anaerobic fitness tests, in order to evaluate the MST as a valid military-specific anaerobic test. In addition, the reliability of the MST was studied to determine the number of trials required to obtain statistically stable results.

Research question 1: Is the MST a reliable fitness test?

Hypothesis 1: In a complex fitness test, such as the MST, the performance in subsequent tests may be affected by learning. Any improvement observed in the MST time may be due to, for example, enhancement in motor skills or pacing. (Pandorf et al. 2003; Spiering et al. 2012.)

Research question 2: Can the MST be considered a valid anaerobic test for soldiers, or is the performance affected by the contribution of other physical characteristics?

Hypothesis 2: In many studies, the validity of an anaerobic test is established by examining the association with a "gold standard" anaerobic test alone (e.g. Dal Pupo et al. 2013; Dardouri et al. 2013; Kimura et al. 2014). While this practice is generally accepted, the role of other fitness characteristics will remain unclear. Despite the high-intensity nature of the MST, the duration of the MST exceeds the point (> 75 s) at which the aerobic metabolism is the dominant energy source (Gastin 2001). Other physical characteristics may contribute to performance to the extent that the MST may not be considered as a valid test of anaerobic performance.

6 METHODS

6.1 Experimental approach to the problem

Measurements of body composition and a wide array of physical fitness tests were administered over the course of three weeks. The tests included an assessment of strength, power, muscular endurance, aerobic capacity and anaerobic performance in order to study the associations between physical fitness characteristics and the MST. The MST was compared with three different anaerobic tests (the 60 s WAnT, 300 m run, and the 60 s Bosco jump test) to evaluate the concurrent validity. In addition, a total of three trials of the MST were conducted to study the reliability.

6.2 Subjects

Twenty-two (*n* = 22) physically active male cadets from The Finnish National Defence University were recruited and voluntarily participated in the study. The mean (± SD) age, height, body mass, and BMI were 22.5 ± 0.8 years, 181.0 ± 6.8 cm, 82.0 ± 7.2 kg, 24.9 ± 1.4 $kg \cdot m^2$, respectively. All participants were informed of the risks and benefits of the study prior to giving their written consent to participate. Participants were instructed to avoid heavy physical exercise to ensure adequate recovery between tests. The study was performed in accordance with the Declaration of Helsinki and approved by the ethical committee of the Hospital District of Helsinki and Uusimaa (Finland).

6.3 Procedures

Participants were measured on 7 different days over the course of three weeks (table 3). Body composition and different physical performance characteristics were measured during the first two weeks. The MST was performed three times during the third week. There was a minimum of 48 hours of recovery between the tests, except for the strength and VO_2max measurements which were performed on the same day.

TABLE 3. Timeline of the measurements during the 3-week study period.

	Measurement day																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
WAnT60	x																				
300 m run			x																		
CMJ60					x																
Muscular strength								x	(x)	(x)											
VO_2max								x	(x)	(x)											
MST														x		x		x			

Muscular strength = MVC_{upper}, MVC_{lower}, SLJ, sit-ups, push-ups. Muscular strength and VO_2max tests were conducted for all participants over three-day period (days 8–10).

Body mass and body composition was measured by using bioelectrical impedance analysis (InBody 720, Biospace Co., Ltd., Seoul, South Korea). The measurements were taken after an overnight fast. Body height was measured with a wall-mounted scale. Waist circumference (WC) was measured by a tape measure.

The 60-second Wingate test (WAnT60) was performed on a friction-braked cycle ergometer (model 894E, Monark AB, Vansbro, Sweden). The braking force was set at 7.5 % BM (Bar-Or 1987). Participants warmed up by cycling on a regular cycle ergometer (Ergoselect 200K, Ergoline GmbH, Bitz, Germany) at 70 rev·min^{-1} for 5 minutes at 75 W. After that the subjects performed two 5-second "all-out" sprints with the same cycle ergometer and braking force as in the test, separated by 30 s of recovery between bouts. The test started after 5 min of recovery after the warm-up. Participants were instructed to cycle as fast as possible for 60 seconds and were given verbal encouragement to maintain maximal effort. Blood lactate (BLa) was measured from the fingertip 2 and 4 minutes after the test. RPE was recorded on a scale from 6 to 20 (Borg 1982). The power output (W) was recorded every 5 seconds using a computer with manufacturer's software (Monark Anaerobic Test Software). Peak power ($WAnT60_{peak}$) was determined as the highest average power in any 5-second period. Mean power ($WAnT60_{mean}$) was calculated as the average power across the entire 60-second test.

The 300 m run was performed on an outdoor tartan-covered running track. The warm-up included a 5-minute jog with two 50-meter sprints, followed by a 5-minute rest before the test. Participants were verbally encouraged during the test. The time was measured with a stopwatch and rounded to the nearest second. BLa and RPE were recorded with the same methods as in the WAnT60. Kimura et al. (2014) showed that the 200 m run is a valid and reliable test to measure anaerobic performance. However, no studies on the validity and reliability of the 300 m run were found.

The 60-second Bosco jump test consisted of repeated vertical countermovement jumps for 60 seconds (CMJ60). The warm-up included cycling on a cycle ergometer (Ergoselect 200K, Ergoline GmbH, Bitz, Germany) at 70 rev·min-1 for 5 minutes at 75 W, followed by two sets of 5 continuous vertical jumps similar to the test. Hands were placed on the hips (akimbo) and the participants were instructed to maintain an approximately 90° knee angle in the bottom position of each jump (Bosco et al. 1983). Participants were verbally encouraged during the test. The flight time and the number of jumps were measured with an infrared timer (Optojump, Microgate, Bolzano, Italy). Mechanical power was calculated from the measured parameters. Peak power ($CMJ60_{peak}$) was determined as the highest average power in any 5-second period. Mean power ($CMJ60_{mean}$) was calculated as the average power across the entire 60-second test. BLa and RPE were recorded with the same methods as in the WAnT60 and 300 m run.

Maximal isometric force of the lower (MVC_{lower}) and upper (MVC_{upper}) extremities were measured in a seated position with an electromechanical force dynamometer (University of Jyväskylä, Finland) (Häkkinen et al. 1998). In the lower extremity measurement, similar to leg press, feet were placed hip-width apart against a force plate and the seat was adjusted to maintain a knee angle of 107°. In the upper body measurement, similar to bench press, the bar was set at shoulder height and the seat was adjusted to maintain an elbow angle of 90°. In both measurements, the participants exerted maximal force in three attempts separated by a minimum of 30 seconds of recovery. The best results (N) were selected for analysis.

Standing long jump (SLJ) was used for the assessment of explosive force production of the lower extremities (Markovic et al. 2004). Feet were placed from hip to shoulder-width apart. While standing, the participants were instructed to descend into a squat position while bringing their arms back, and to maximally extend their hips while swinging their arms forward. After proper technique instructions and 5–7 warm-up jumps, the participants performed three maximal jumps on a 10 mm thick plastic mattress (Fysioline Co, Tampere, Finland). The takeoff and landing were performed bilaterally, and falling backwards was not allowed. The best result in centimeters was kept for analysis.

Sit-ups were used to measure strength endurance of the abdominal and hip flexor muscles (Viljanen et al. 1991). In the starting position, the participant laid on his back while his feet were supported by an assistant. The knees were flexed at a 90° angle. Fingers were crossed behind the back of the head and the elbows pointed upward. One repetition was counted when the upper body was raised from the ground and the elbows touched the knees or were brought to the side of the knees at the same level. The result was determined as the number of successful repetitions in 60 seconds.

Push-ups were used to measure upper body strength endurance. Hands were placed shoulder-width, fingers pointing forward and so that thumbs could reach the shoulders. Feet were placed together or at most hip-width apart. In the starting position the arms were extended, and feet, trunk, shoulders and head were in the same line. One repetition was counted when the participant flexed his arms and descended his torso so that the angle of elbows was 90° or below and returned back to the starting position by flexing his arms. The result was determined as the number of successful repetitions in 60 seconds.

VO_2max was measured by using a modified Balke treadmill exercise protocol (Balke & Ware 1959). The warm-up consisted of 5 minutes of running on the treadmill (h/p/cosmos pulsar® 3p, h/p/cosmos sports & medical gmbh, Germany) with initial speed of the test, which was set between 8–10 $km \cdot h^{-1}$ based on the participant's last 12-minute running test result. The inclination was set to 3°. After 2 minutes of rest, the test started. The speed increased 1 $km \cdot h^{-1}$ every minute until exhaustion. Oxygen consumption was measured breath-by-breath using a gas analyzer (Oxycon Pro™, Jaeger, Hoechberg, Germany). VO_2max was determined as the highest average 20 s VO_2 value.

The MST (figure 6) was performed on an artificial grass court while wearing a 20 kg fighting load, including a combat dress uniform, boots, body armor and tactical vest, helmet and assault rifle replica (Pihlainen et al. 2018). Before each trial, the participants completed the MST with submaximal intensity for warm-up and familiarization. The test started from prone position. First, the participants performed four 6.2 m combat rushes, followed by 11.3 meters of low crawl. Next, a sprint of 21.8 m was performed, followed by another 21.8 m sprint while jumping over three 40 cm high obstacles. Then, two 16 kg kettlebells were lifted, carried and lowered four times for a distance of 2.5 meters, followed by a zigzag run of 42.4 meters. In the last phase, a 123 kg sled was dragged for 24 meters before running back to the starting line. The total length of the MST was 242.5 meters. The time was recorded with a stopwatch and rounded to the nearest second. BLa and RPE were recorded with the same methods as in the anaerobic tests.

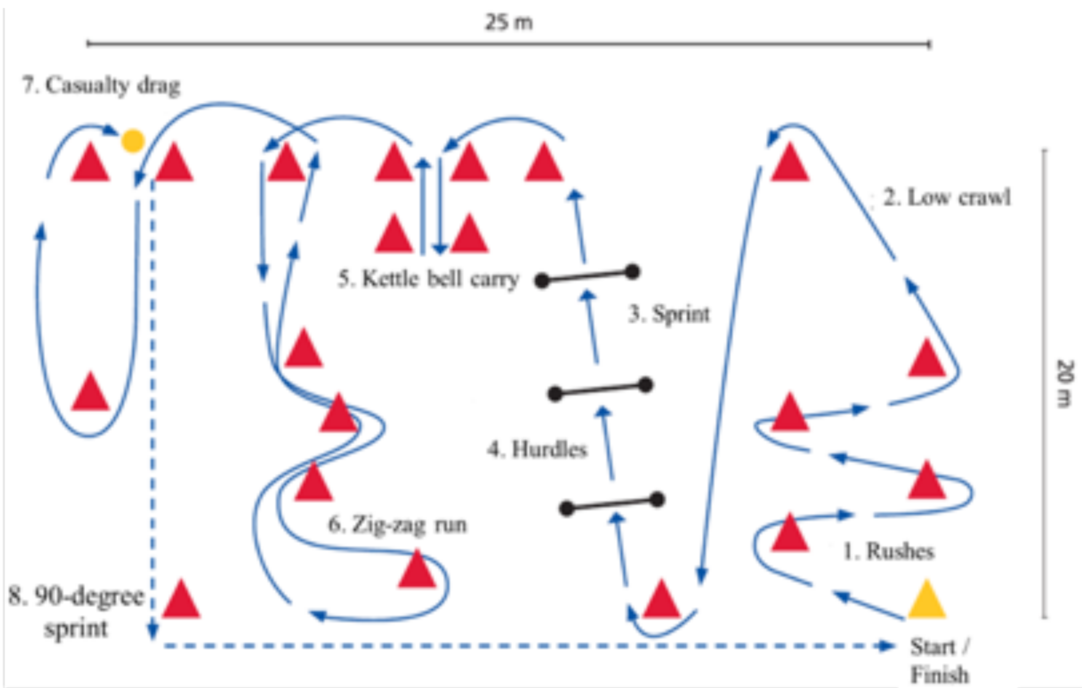

FIGURE 6. Schematic diagram of the MST.

6.4 Statistical analyses

Subjects who completed all three trials (n = 17) were included in reliability analysis. A one-way repeated measures analysis of variance (ANOVA) was used to examine differences in mean BLa, RPE, and performance times between trials (i.e. possible learning effect). Reliability for the MST was established by determining the intraclass correlation coefficient, ICC (single measures, 2-way random with absolute agreement). ICC values greater than 0.75 were considered as good reliability, those between 0.5 and 0.75 as moderate reliability, and those below 0.5 as poor reliability. Pearson's correlation coefficients and stepwise linear regression analysis were used to assess the relationships between the MST and physical fitness characteristics.

7 RESULTS

The anthropometric and physical performance characteristics of the participants are presented in tables 4 and 5. Of the 22 volunteers who participated in the study, some were unable to complete all measurements due to academic pressures. No injuries were observed as a result of participating in this study.

TABLE 4. Body composition characteristics (n = 22)

	Mean ± SD
Height (cm)	181.0 ± 6.8
Body mass (kg)	82.0 ± 7.2 kg
BMI ($kg \cdot m^2$)	24.9 ± 1.4
SMM (kg)	41.6 ± 3.6
FATM (kg)	9.9 ± 2.5
FAT%	12.0 ± 2,5
WC	81.6 ± 2.8

SMM, skeletal muscle mass; FATM, fat mass; WC, waist circumference.

TABLE 5. Physical performance characteristics.

	Mean ± SD	n
$WAnT60_{mean}$ (W)	555.6 ± 60.8	15
$WAnT60_{peak}$ (W)	907.7 ± 138.8	15
300 m (s)	44.0 ± 2.2	17
$CMJ60_{mean}$ (W)	17.0 ± 2.9	19
$CMJ60_{peak}$ (W)	21.3 ± 3.5	19
MVC_{lower} (N)	1287 ± 198	20
MVC_{upper} (N)	3514 ± 508	20
SLJ (cm)	244.4 ± 16.6	20
Sit-ups (reps/min)	52.9 ± 9.9	20
Push-ups (reps/min)	49.8 ± 11.9	20
VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$)	55.8 ± 4.6	20

MVC_{lower}, maximal voluntary contraction of the lower extremities; MVC_{upper}, maximal voluntary contraction of the upper extremities; SLJ, standing long jump.

The mean time to perform the MST was 108.5 ± 8.6, 103.3 ± 6.6, and 98.4 ± 6.5 seconds for the first, second, and third trial, respectively. A one-way repeated measures ANOVA showed significant ($p < 0.05$) improvement in the MST time between every trial. Peak blood lactate after the MST increased between the first and second trial, but decreased between second and third trial (table 6). Individual results in trials 1–3 are presented in figure 7. A moderate reliability was found between the first and second trial (ICC = 0.58, 95 % CI: 0.03–0.84, $p = 0.001$), and second and third trial (ICC = 0.74, 95 % CI: −0.06–0.94, $p < 0.001$).

TABLE 6. Time, peak blood lactate and RPE for the MST (n =17).

	Trial 1	Trial 2	Trial 3
MST time (s)	108.5 ± 8.6	103.3 ± 6.6*	98.4 ± 6.5*†
Peak blood lactate (mmol·L^{-1})	14.4 ± 2.4	17.1 ± 1.7*	15.8 ± 2.1*†
RPE (6–20)	18.2 ± 1.0	18.8 ± 0.7*	18.6 ± 1.0*

Values are mean ± SD. *Significantly ($p < 0.05$) different than trial 1. †Significantly ($p < 0.05$) different than trial 2.

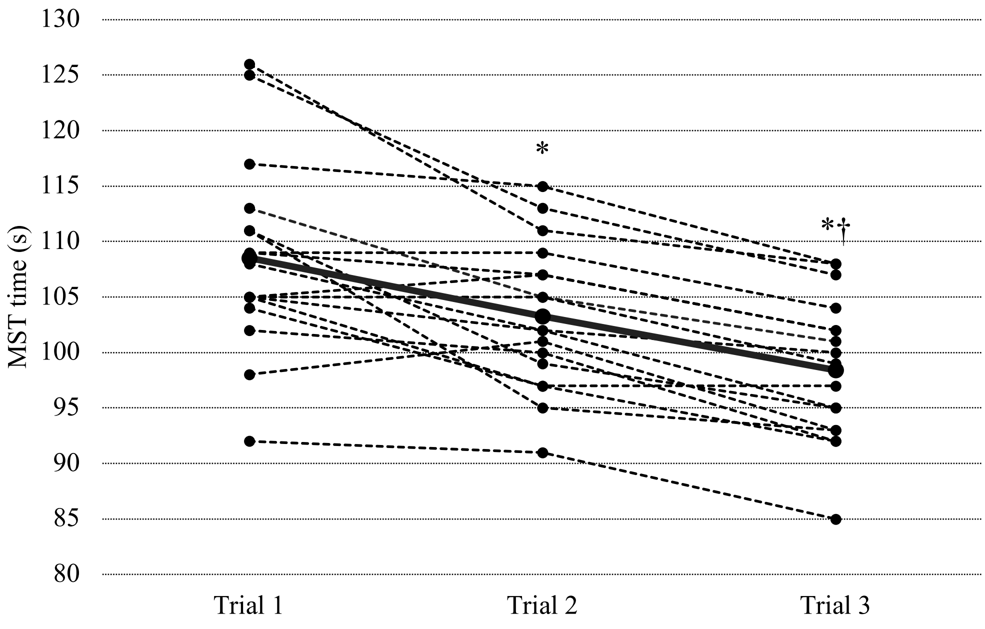

FIGURE 7. Individual (dotted lines) and mean (solid line) times in the MST. *Significantly ($p < 0.05$) different than trial 1. †Significantly ($p < 0.05$) different than trial 2.

The second trial of the MST, which had the highest number of samples, was selected for further analysis to study associations with other measured variables. Of the three anaerobic tests, 300 m run time was positively correlated ($p < 0.05$), and $WAnT60_{peak}$, $CMJ60_{peak}$, and $CMJ60_{mean}$ inversely correlated ($p < 0.05$) with the MST (table 7, figure 8). Among other physical fitness variables, MVC_{upper}, SLJ, sit-ups, and VO_2max were inversely correlated ($p < 0.05$) with the MST. Among anthropometric variables, FATM and FAT% were positively correlated ($p < 0.05$) with the MST, whereas height, body mass, BMI, SMM, and WC were not correlated with performance in the MST (table 7). The stepwise linear regression analysis showed that $CMJ60_{peak}$ explained 68 % of the variance in the MST trial 2 ($F = 24.5$, $p = 0.01$, $R^2 = 0.68$).

TABLE 7. Correlations between the measured variables and MST.

	MST (trial 2)	*n*
Height (cm)	−0.11	21
Body mass (kg)	−0.16	21
BMI ($kg \cdot m^2$)	−0.16	21
SMM (kg)	−0.38	21
FATM (kg)	0.49^{*}	21
FAT%	0.63^{*}	21
WC	−0.32	21
$WAnT60_{mean}$ (W)	−0.32	15
$WAnT60_{peak}$ (W)	-0.62^{*}	15
300 m (s)	0.58^{*}	16
$CMJ60_{mean}$ (W)	-0.65^{**}	19
$CMJ60_{peak}$ (W)	-0.68^{**}	19
MVC_{lower} (N)	−0.27	19
MVC_{upper} (N)	-0.67^{**}	19
SLJ (cm)	-0.66^{**}	19
Sit-ups (reps/min)	-0.69^{**}	19
Push-ups (reps/min)	−0.42	19
VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$)	-0.63^{**}	19

* $p < 0.05$, ** $p < 0.01$

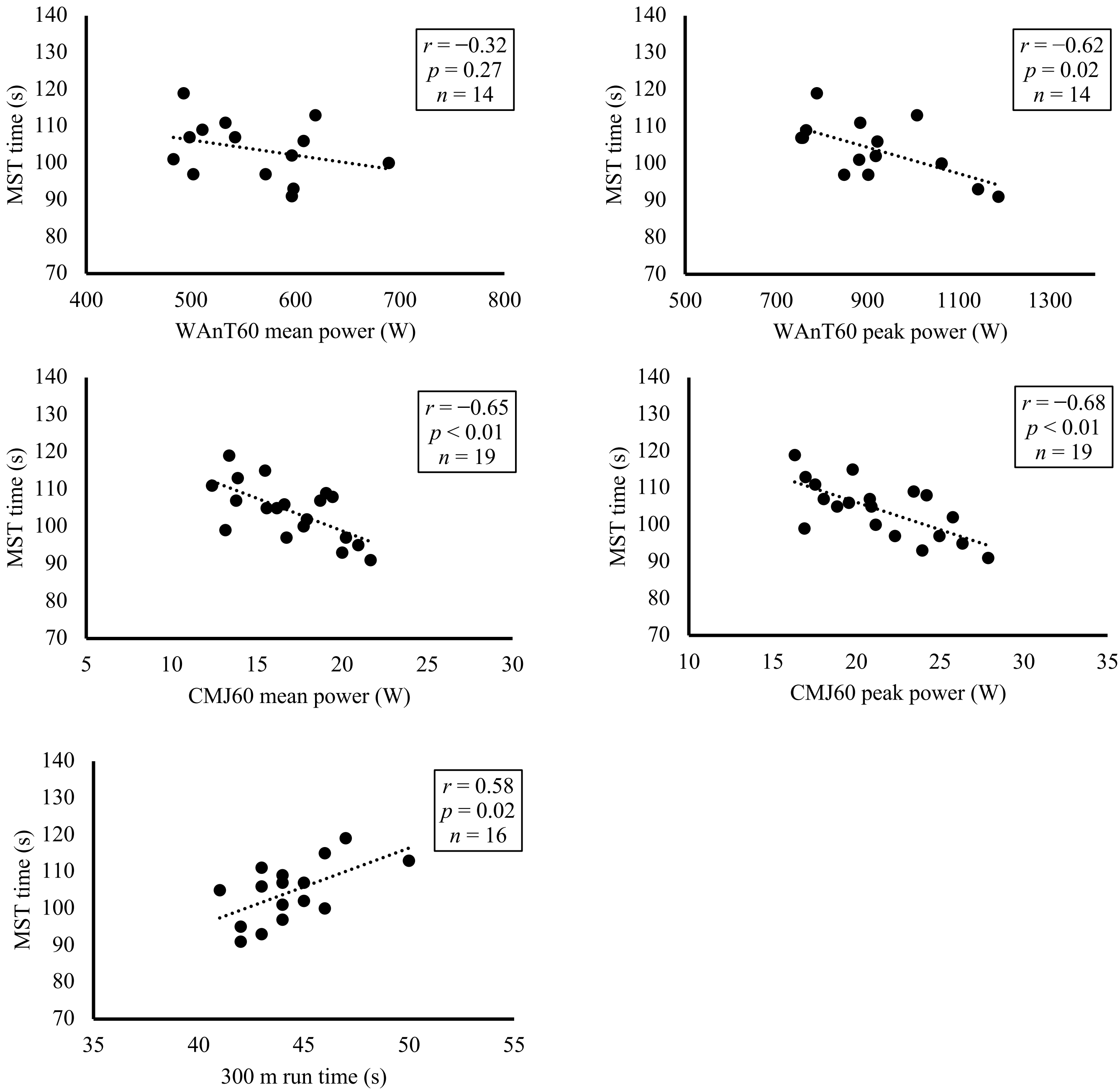

FIGURE 8. The relationship between the MST (trial 2) and the anaerobic performance variables.

8 DISCUSSION

The purpose of the present study was to assess reliability and criterion validity of the MST and to evaluate the associations of MST performance with physical fitness and body composition characteristics. The primary findings showed that the MST time improved across all three trials, which could be explained, at least in part, by a significant learning effect. In addition, the MST performance was associated with several measured variables, suggesting that multiple physical characteristics contribute to such a high-intensity task.

8.1 Reliability of the military simulation test

The improvement in the MST time in every trial suggested a significant learning effect, especially between the first and second trials. ICC demonstrated, however, moderate reliability between the first and second, and second and third trial. Due to the complex nature of the MST, several familiarization trials are required to reach statistically stable results.

It seems that the complexity and the relatively long duration of the MST is associated with the observed learning effect. Different phases of the MST are simple tasks as such, but when combined into one test, the number of ways to improve performance increases after the test becomes familiar. For example, Angeltveit et al. (2016) observed that if a subject tried to switch hand when dragging a doll in a simulated casualty evacuation test, the doll often stopped completely, and it required extra effort to get it moving again. This type of strategy development could have been used in the evacuation task of the MST. The evacuation technique in the MST was not standardized, and the subjects could choose to drag the sled using both hands with back facing forward, or using one hand with frontside of the body facing forwards. In these variations, the joint angles and the contribution of working muscles are different, resulting in different performance times before the best technique for the individual is found. Despite the initial submaximal MST performance for warm-up and familiarization before each trial, this type of change in evacuation technique was observed in many subjects.

Other possible factors that explain the learning effect are pacing strategies, skill improvement, motor learning, and motivation. Peak blood lactate after the MST decreased between the second and third trial, suggesting improved neuromuscular efficiency and/or pacing. Correspondingly, Pihlainen et al. (2020) found that the relative EMG activity of hamstring and quadriceps muscles increased during the MST when follow-up testing was completed after a 6-month military operation. Increased muscle activity suggests that improved neuromuscular efficiency of the thigh muscles at least partly explains improvement in the MST time. However, it must be taken into account that the study included a training intervention, which could explain some part of the changes in EMG activity of the thigh muscles. Hence, the results are not directly applicable into the discussion of the present study.

Similar findings about the learning effect have been reported in other studies that have investigated the reliability of military-specific fitness tests. Pandorf et al. (2003) concluded that the observed longer familiarization time for an indoor 6-station obstacle course was perhaps due to its complex nature, in which many skills contributed to optimal performance. The obstacle course required two test sessions before reliable results were obtained, while a simpler repetitive box lifting test required only one. Correspondingly, Spiering et al. (2012) found that the most skill-related test, grenade throw, required three familiarization trials to reach statistically stable results, whereas more simple tests, such as repetitive box lift and carry, required only one. Foulis et al. (2017b) discussed the role of previous experience in observed learning effect in simulated military tasks. When compared to common soldiering tasks, such as foot march and sandbag carry, the least familiar tests (stowing ammo, loading the main gun of a tank) continued to show improvements over all four test sessions.

These previous findings support the present results, suggesting that the learning effect is associated with improved performance in complex simulation tests, which contain unfamiliar skill-related tasks. This factor needs to be taken into account, especially when using task simulation tests for the assessment of changes in physical fitness.

8.2 Validity and performance determinants of the military simulation test

All variables measuring anaerobic performance, except $WAnT60_{mean}$, were associated with the MST (r = 0.58–0.68). Nevertheless, direct conclusions cannot be made that the MST is a valid test for the measurement of anaerobic performance, as the MST showed similar correlations with MVC_{upper}, SLJ, sit-ups, and VO_2max (r = 0.63–0.69). Indeed, Pihlainen et al. (2018) also found that the MST was associated with measures of explosive force of the lower extremities, and aerobic fitness which was measured using a 3 000 m running test. The main explanation for the role of aerobic endurance might be the duration of the MST, because it exceeds the point (> 75 s) at which the aerobic metabolism is the dominant energy source (Gastin 2001).

The present results also highlight the role of explosive strength and anaerobic alactic performance of the lower body on the MST performance, as significant correlations were found with $WAnT60_{peak}$, $CMJ60_{peak}$, and SLJ. This is also supported by previous studies that have investigated high-intensity military task simulation tests (Harman et al. 2008; Mala et al. 2015; Pihlainen et al. 2018).

At the moment, there is no test which enables the accurate measurement of anaerobic metabolism, as the contribution of anaerobic alactic, and anaerobic lactic energy production are difficult to distinguish from total energy production. For this reason, when developing valid anaerobic tests, the associations with different physical performance characteristics must be taken into consideration. In many cases, however, the validity of anaerobic test is established by evaluating the associations solely with tests that are considered to be valid anaerobic tests, such as WAnT and MAOD (Zagatto et al. 2009; Dal Pupo et al. 2014; Kimura et al. 2014; Andrade et al. 2015). With these methods, the role of other fitness characteristics, especially aerobic endurance, remains unclear.

It is obvious that the anaerobic metabolism has a significant contribution in the MST, as the peak blood lactate values were high after each test (> 14 $mmol \cdot L^{-1}$). Although the accuracy and applicability of this method is controversial, the post-exercise blood lactate level reflects the quantity of energy production through anaerobic lactic metabolism (Green & Dawson 1993; Gastin 2001).

As all three energy pathways have a contribution in energy production during the MST, definition of the purpose of the test must be done with caution. If the MST were described as a valid anaerobic test, the contribution of other physical characteristics cannot be in a dominant role. For example, Angeltveit et al. (2016) concluded the evacuation (EVAC) test to be a valid measurement of anaerobic performance. The average time in the test was less than 60 seconds, which seems to be suitable for the assessment of anaerobic capacity (Gastin 2001). The EVAC test correlated well with the WAnT ($r = -0.62$), 300 m sprint ($r = 0.51$), and 300 m mean power ($r = -0.67$). However, the EVAC test was also found to be associated with VO_2max ($r = -0.72$). As the authors pointed out, aerobic energy production in a maximal work lasting as long as the EVAC test is approximately 40 % (Gastin 2001). It was concluded that the aerobic energy system is important for the performance in the EVAC test, although less important than the anaerobic energy system. While this conclusion is certainly correct and is supported by the fundamental basis of exercise physiology, aerobic endurance capacity still plays a significant role in such relatively short-duration effort.

Correspondingly, because of the significantly longer duration of the MST, it can be assumed that aerobic metabolism is the dominant energy source, while the anaerobic metabolism obviously has a significant role in energy production. In addition, the MST was associated with VO_2max to the same extent as with the measures of anaerobic performance. It seems that attempts to establish a valid military-specific anaerobic test should focus on the development of a short-duration (≤ 60 seconds) test, which does not require high levels of skill. (Vandewalle et al. 1987; Gastin 2001.) Translating a simulation test into a task-related predictive test, which includes vertical or horizontal jumping, could be a viable option (Harman et al. 2008).

The MST performance was associated with several measured variables, suggesting that such high-intensity combat-related tasks require a great variety of physical characteristics, and it seems that the MST may not be considered as a valid test to measure anaerobic performance of soldiers. The MST relies on the simultaneous participation of all three energy pathways, and a combination of neuromuscular, anaerobic and anaerobic performance characteristics. Therefore, using terms ‘aerobic’ or ‘anaerobic’ to describe the purpose of the MST is most likely incorrect. In exercise with a duration exceeding 1 min, the use of the term ‘endurance intensive effort’ is suggested (Chamari & Padulo 2015).

8.3 Limitations of the study

Based on the recommendation by Atkinson and Nevill (1998), repeated-measures ANOVA was used to study changes in the MST time between trials. In the case of significant improvement in performance (potential learning effect), assessment of reliability (ICC) is not recommended. In the present study, the MST time improved across all three trials. Because of the limited schedule, it was not possible to run additional trials to investigate whether the MST performance had improved even further.

In addition, the subjects were young, trained (VO_2max 55.8 $mL \cdot kg^{-1} \cdot min^{-1}$) male cadets, so the results of this study may not be applicable for conscripts and reservists, who represent the majority of soldiers in the Finnish Defence Forces. In less fit soldiers, the time to complete the MST is longer, and the contribution of aerobic metabolism is likely higher, which further emphasizes the benefits of short-duration (≤ 60 seconds) test if the purpose is to evaluate anaerobic performance.

9 CONCLUSIONS

In conclusion, the MST does not appear to be a reliable test to measure changes in military-specific high-intensity performance of soldiers, as the performance in subsequent tests is probably affected by learning. In the present study, the observed improvement in the MST time was likely due to pacing strategies, skill improvement, motor learning, and motivation. Different phases of the MST are simple tasks as such, but when combined into one test, the number of ways to improve performance increases after the test becomes familiar. In addition, the MST may not be considered as a valid anaerobic test, as the performance in MST relies on the simultaneous participation of all three energy pathways, and a combination of neuromuscular, anaerobic, and anaerobic performance characteristics.

In order to improve the physical performance of soldiers in high-intensity combat-related military tasks, units should be encouraged to train all components of physical fitness. The MST may have a logical relationship with soldiers' occupational physical demands, and it could be translated into a task-related predictive test when special attention is paid to familiarization procedures. Military task simulation tests provide high face validity when compared to generic fitness tests and can be used to obtain more information about absolute job-related military physical performance and operational readiness. For the development of valid and reliable military-specific anaerobic tests, future studies should focus on simple and short-duration (≤ 60 seconds) tests.

REFERENCES

Alele, F. O., Malau-Aduli, B. S., Malau-Aduli, A. E., & Crowe, M. J. 2020. Epidemiology of exertional heat illness in the military: a systematic review of observational studies. International Journal of Environmental Research and Public Health 17 (19), 7037.

Allison, K. F., Keenan, K. A., Wohleber, M. F., Perlsweig, K. A., Pletcher, E. R., Lovalekar, M., Beals, K., Coleman, L. C., & Nindl, B. C. 2017. Greater ankle strength, anaerobic and aerobic capacity, and agility predict Ground Combat Military Occupational School graduation in female Marines. Journal of Science and Medicine in Sport 20, 85–90.

Andrade, V. L., Zagatto, A. M., Kalva-Filho, C. A., Mendes, O. C., Gobatto, C. A., Campos, E. Z., & Papoti, M. 2015. Running-based anaerobic sprint test as a procedure to evaluate anaerobic power. International Journal of Sports Medicine 36, 1156–1162.

Angeltveit, A., Paulsen, G., Solberg, P. A., & Raastad, T. 2016. Validity, reliability, and performance determinants of a new job-specific anaerobic work capacity test for the Norwegian Navy Special Operations Command. The Journal of Strength & Conditioning Research 30 (2), 487–496.

Armstrong, L. E., Casa, D. J., Millard-Stafford, M., Moran, D. S., Pyne, S. W., & Roberts, W. O. 2007. Exertional heat illness during training and competition. Medicine & Science in Sports & Exercise 39 (3), 556–572.

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

Balke B., & Ware R. W. 1959. An experimental study of physical fitness of Air Force personnel. United States Armed Forces medical journal 10 (6), 675–688.

Bangsbo, J., Iaia, F. M., & Krustrup, P. 2008. The yo-yo intermittent recovery test. Sports Medicine 38 (1), 37–51.

Bar-Or, O. 1987. The Wingate anaerobic test. An update on methodology, reliability and validity. Sports Medicine 4(6), 381–394.

Beck, B., Billing, D. C., & Carr, A. J. 2016a. Developing physical and physiological employment standards: Translation of job analysis findings to assessments and performance standards–A systematic review. International Journal of Industrial Ergonomics 56, 9–16.

Beck, B., Carstairs, G. L., Caldwell Odgers, J. N., Doyle, T. L., & Middleton, K. J. 2016b. Jerry can carriage is an effective predictor of stretcher carry performance. Ergonomics 59 (6), 813–820.

Beck, B., Middleton, K. J., Carstairs, G. L., Billing, D. C., & Caldwell, J. N. 2016c. Predicting stretcher carriage: investigating variations in bilateral carry tests. Applied Ergonomics, 55, 124–132.

Beck, B., Carstairs, G. L., Billing, D. C., Caldwell, J. N., & Middleton, K. J. 2017. Modifiable anthropometric characteristics are associated with unilateral and bilateral carry performance. The Journal of Strength & Conditioning Research 31 (2), 489–494.

Bedno, S. A., li, Y., Han, W., Cowan, D. N., Scott, C. T., Cavicchia, M. A., & Niebuhr, D. W. 2010. Exertional heat illness among overweight U.S. Army recruits in basic training. Aviation, Space, and Environmental Medicine 81, 107–111.

Bell, N. S., Mangione, T. W., Hemenway, D., Amoroso, P. J., & Jones, B. H. 2000. High injury rates among female Army trainees: a function of gender? American Journal of Preventive Medicine 18 (3), 141–146.

Beneke, R., Pollmann, C. H., Bleif, I., Leithäuser, R., & Hütler, M. 2002. How anaerobic is the Wingate Anaerobic Test for humans? European Journal of Applied Physiology 87 (4–5), 388–392.

Bishop, P. A., Fielitz, L. R., Crowder, T. A., Anderson, C. L., Smith, J. H., & Derrick, K. R. 1999. Physiological determinants of performance on an indoor military obstacle course test. Military Medicine 164 (12), 891–896.

Billing, D. C., Silk, A. J., Tofari, P. J., & Hunt, A. P. 2015. Effects of military load carriage on susceptibility to enemy fire during tactical combat movements. The Journal of Strength & Conditioning Research 29, 134–138.

Borg, G. A. 1982. Psychophysical bases of perceived exertion. Medicine & Science in Sports & Exercise 14, 377–381.

Bosco, C., Luhtanen, P., & Komi, P. V. 1983. A simple method for measurement of mechanical power in jumping. European Journal of Applied Physiology and Occupational Physiology 50 (2), 273–282.

Boye, M. W., Cohen, B. S., Sharp, M. A., Canino, M. C., Foulis, S. A., Larcom, K., & Smith, L. 2017. U.S. Army physical demands study: Prevalence and frequency of performing physically demanding tasks in deployed and non-deployed settings. Journal of Science and Medicine in Sport 20, 57–61.

Bullock, S. H., Jones, B. H., Gilchrist, J., & Marshall, S. W. 2010. Prevention of physical training–related injuries: recommendations for the military and other active populations based on expedited systematic reviews. American Journal of Preventive Medicine 38 (1), 156–181.

Caiozzo, V. J., & Kyle, C. R. 1980. The effect of external loading upon power output in stair climbing. European Journal of Applied Physiology and Occupational Physiology 44 (3), 217–222.

Carstairs, G. L., Ham, D. J., Savage, R. J., Best, S. A., Beck, B., & Doyle, T. L. 2016. A box lift and place assessment is related to performance of several military manual handling tasks. Military Medicine 181 (3), 258–264.

Caldwell, J. N., Patterson, M. J., & Taylor, N. A. 2012. Exertional thermal strain, protective clothing and auxiliary cooling in dry heat: evidence for physiological but not cognitive impairment. European Journal of Applied Physiology 112 (10), 3597–3606.

Chamari, K., & Padulo, J. 2015. ‘Aerobic’ and ‘anaerobic’ terms used in exercise physiology: a critical terminology reflection. Sports Medicine – Open 1 (1), 9.

Cheung, S. S., Lee, J. K., & Oksa, J. 2016. Thermal stress, human performance, and physical employment standards. Applied Physiology, Nutrition, and Metabolism 41 (6), 148–164.

Coffman, K. E., Luippold, A. J., Salgado, R. M., Heavens, K. R., Caruso, E. M., Fulco, C. S., & Kenefick, R. W. 2020. Aerobic Exercise Performance During Load Carriage and Acute Altitude Exposure. Journal of Strength and Conditioning Research 34 (4), 946–951.

Corbett, J., Wright, J., Tipton, M. J. 2020. Sex differences in response to exercise heat stress in the context of the military environment. BMJ Military Health. Published online first: 23 February 2020. doi: 10.1136/jramc-2019-001253

Dada, E. O., Anderson, M. K., Grier, T., Alemany, J. A., & Jones, B. H. 2017. Sex and age differences in physical performance: A comparison of Army basic training and operational populations. Journal of Science and Medicine in Sport 20, 68–73.

Dal Pupo, J., Gheller, R. G., Dias, J. A., Rodacki, A. L., Moro, A. R., & Santos, S. G. 2014. Reliability and validity of the 30-s continuous jump test for anaerobic fitness evaluation. Journal of Science and Medicine in Sport 17 (6), 650–655.

Dardouri, W., Gharbi, Z., Selmi, M. A., Sassi, R. H., Moalla, W., Chamari, K., & Souissi, N. 2014. Reliability and validity of a new maximal anaerobic shuttle running test. International Journal of Sports Medicine 35 (4), 310–315.

di Prampero, P. E., & Ferretti, G. 1999. The energetics of anaerobic muscle metabolism: a reappraisal of older and recent concepts. Respiration Physiology 118 (2–3), 103–115.

Drain, J., Billing, D., Neesham-Smith, D., & Aisbett, B. 2016. Predicting physiological capacity of human load carriage – a review. Applied Ergonomics 52, 85–94.

Eddy, M. D., Hasselquist, L., Giles, G., Hayes, J. F., Howe, J., Rourke, J., Coyne, M., O'Donovan, M., Batty, J., Brunyé, T. T., & Mahoney, C. R. 2015. The effects of load carriage and physical fatigue on cognitive performance. PloS One 10 (7), e0130817.

Epstein, Y., Rosenblum, J., Burstein, R., & Sawka, M. N. 1988. External load can alter the energy cost of prolonged exercise. European Journal of Applied Physiology and Occupational Physiology 57 (2), 243–247.

Epstein, Y., Yanovich, R., Moran, D. S., & Heled, Y. 2013. Physiological employment standards IV: integration of women in combat units physiological and medical considerations. European Journal of Applied Physiology 113 (11), 2673–2690.

Epstein, Y., & Yanovich, R. 2019. Heatstroke. New England Journal of Medicine 380 (25), 2449–2459.

Fallowfield, J. L., Delves, S. K., Hill, N. E., Cobley, R., Brown, P., Lanham-New, S. A., Frost, G., Brett, S. J., Murphy, K. G., Montain, S. J., Nicholson, C., Stacey, M., Ardley, C., Shaw, A., Bentley, C., Wilson, D. R., & Allsopp, A. J. 2014. Energy expenditure, nutritional status, body composition and physical fitness of Royal Marines during a 6-month operational deployment in Afghanistan. British Journal of Nutrition 112 (5), 821–829.

Foulis, S. A., Sharp, M. A., Redmond, J. E., Frykman, P. N., Warr, B. J., Gebhardt, D. L., Baker, T. A., Canino, M. C., & Zambraski, E. J. 2017a. U.S. Army physical demands study: Development of the occupational physical assessment test for combat arms soldiers. Journal of Science and Medicine in Sport 20, 74–78.

Foulis, S. A., Redmond, J. E., Frykman, P. N., Warr, B. J., Zambraski, E. J., & Sharp, M. A. 2017b. U.S. Army physical demands study: Reliability of simulations of physically demanding tasks performed by combat arms soldiers. The Journal of Strength & Conditioning Research 31 (12), 3245–3252.

Friedl, K. E. 2004. Can you be large and not obese? The distinction between body weight, body fat, and abdominal fat in occupational standards. Diabetes Technology & Therapeutics 6 (5), 732–749.

Friedl, K. E. 2012. Body composition and military performance – many things to many people. The Journal of Strength & Conditioning Research 26, 87–100.

Friedl, K. E., Knapik, J. J., Häkkinen, K., Baumgartner, N., Groeller, H., Taylor, N. A., Duarte, A. F. A., Kyröläinen, H., Jones, B. H., Kraemer, W. J., & Nindl, B. C. 2015. Perspectives on aerobic and strength influences on military physical readiness: report of an international military physiology roundtable. The Journal of Strength & Conditioning Research 29, 10–23.

Fulco, C. S., Rock, P. B., & Cymerman, A. 1998. Maximal and submaximal exercise performance at altitude. Aviation, Space, and Environmental Medicine 69 (8), 793–801.

Gagnon, P., Spivock, M., Reilly, T., Mattie, P., & Stockbrugger, B. 2015. The FORCE fitness profile – adding a measure of health-related fitness to the Canadian Armed Forces operational fitness evaluation. The Journal of Strength & Conditioning Research 29, 192–198.

Gastin, P. B. 2001. Energy system interaction and relative contribution during maximal exercise. Sports Medicine 31 (10), 725–741.

Green, S., & Dawson, B. 1993. Measurement of anaerobic capacities in humans. Sports Medicine 15 (5), 312–327.

Green, S. 1994. A definition and systems view of anaerobic capacity. European Journal of Applied Physiology and Occupational Physiology 69 (2), 168–173.

Greeves, J. P. 2015. Physiological implications, performance assessment and risk mitigation strategies of women in combat-centric occupations. The Journal of Strength & Conditioning Research 29, 94–100.

Groeller, H., Burley, S., Orchard, P., Sampson, J. A., Billing, D. C., & Linnane, D. 2015. How effective is initial military-specific training in the development of physical performance of soldiers? The Journal of Strength & Conditioning Research 29, 158–162.

Harman, E. A., Gutekunst, D. J., Frykman, P. N., Sharp, M. A., Nindl, B. C., Alemany, J. A., & Mello, R. P. 2008. Prediction of simulated battlefield physical performance from field-expedient tests. Military Medicine 173 (1), 36–41.

Hendrickson, N. R., Sharp, M. A., Alemany, J. A., Walker, L. A., Harman, E. A., Spiering, B. A., Hatfield, D. L., Yamamoto, L. M., Maresh, C. M., Kraemer, W. J., & Nindl, B. C. 2010. Combined resistance and endurance training improves physical capacity and performance on tactical occupational tasks. European Journal of Applied Physiology 109 (6), 1197–1208.

Hetzler, R. K., Vogelpohl, R. E., Stickley, C. D., Kuramoto, A. N., DeLaura, M. R., & Kimura, I. F. 2010. Development of a modified Margaria-Kalamen anaerobic power test for American football athletes. The Journal of Strength & Conditioning Research 24 (4), 978–984.

Hickson, R. C. 1980. Interference of strength development by simultaneously training for strength and endurance. European Journal of Applied Physiology and Occupational Physiology 45 (2–3), 255–263.

Hill, A. V. 1938. The heat of shortening and the dynamic constants of muscle. Proceedings of the Royal Society of London 126 (843), 136–195.

Hill, A. V. 1950. The dimensions of animals and their muscular dynamics. Science Progress 38 (150), 209–230.

Huang, H. C., Nagai, T., Lovalekar, M., Connaboy, C., & Nindl, B. C. 2018. Physical fitness predictors of a warrior task simulation test. The Journal of Strength & Conditioning Research 32 (9), 2562–2568.

Huskey, T., Mayhew, J. L., Ball, T. E., & Arnold, M. D. 1989. Factors affecting anaerobic power output in the Margaria-Kalamen test. Ergonomics 32 (8), 959–965.

Häkkinen, K., Alen, M. Kallinen, M. Izquierdo, M. Jokelainen, K. Lassila, H., Mälkiä, E., Kraemer, W. J., Newton, R. U. 1998. Muscle CSA, Force Production, and Activation of Leg Extensors during Isometric and Dynamic Actions in Middle-Aged and Elderly Men and Women. Journal of Aging and Physical Activity 6, 232–247.

Inbar, O., Bar-Or, O., & Skinner, J. S. 1996. The Wingate Anaerobic Test. Human Kinetics: Champaign IL.

Jaafar, H., Rouis, M., Coudrat, L., Attiogbé, E., Vandewalle, H., & Driss, T. 2014. Effects of load on Wingate test performances and reliability. The Journal of Strength & Conditioning Research 28 (12), 3462–3468.

Jacobs, I. R. A., & Kaiser, P. 1982. Lactate in blood, mixed skeletal muscle, and FT or ST fibres during cycle exercise in man. Acta Physiologica 114 (3), 461–466.

Jaric, S., & Markovic, G. 2013. Body mass maximizes power output in human jumping: a strength-independent optimum loading behavior. European Journal of Applied Physiology 113 (12), 2913–2923.

Jones, B. H., Hauret, K. G., Dye, S. K., Hauschild, V. D., Rossi, S. P., Richardson, M. D., & Friedl, K. E. 2017. Impact of physical fitness and body composition on injury risk among active young adults: a study of Army trainees. Journal of Science and Medicine in Sport 20, 17–22.

Joseph, A., Wiley, A., Orr, R., Schram, B., & Dawes, J. J. 2018. The impact of load carriage on measures of power and agility in tactical occupations: a critical review. International Journal of Environmental Research and Public Health 15 (1), 88.

Kalamen, J. L. 1968. Measurement of maximum muscular power in man. Doctoral dissertation. Ohio State University.

Kimura, I. F., Stickley, C. D., Lentz, M. A., Wages, J. J., Yanagi, K., & Hetzler, R. K. 2014. Validity and reliability of the Hawaii anaerobic run test. The Journal of Strength & Conditioning Research 28 (5), 1386–1393.

Knapik, J. J., Sharp, M. A., Canham-Chervak, M., Hauret, K., Patton, J. F., & Jones, B. H. 2001. Risk factors for training-related injuries among men and women in basic combat training. Medicine and Science in Sports and Exercise 33 (6), 946–954.

Knapik, J. J., Reynolds, K. L., & Harman, E. 2004. Soldier load carriage: historical, physiological, biomechanical, and medical aspects. Military Medicine 169 (1), 45–56.

Knapik, J. J., Harman, E. A., Steelman, R. A., & Graham, B. S. 2012. A systematic review of the effects of physical training on load carriage performance. The Journal of Strength & Conditioning Research 26 (2), 585–597.

Knudson, D. V. 2009. Correcting the use of the term "power" in the strength and conditioning literature. The Journal of Strength & Conditioning Research 23 (6), 1902–1908.

Knuttgen, H. G., & Kraemer, W. J. 1987. Terminology and measurement in exercise performance. Journal of Applied Sport Science Research 1 (1), 1–10.

Kraemer, W. J., Vogel, J. A., Patton, J. F., Dziados, J. E., & Reynolds, K. L. 1987. The effects of various physical training programs on short duration, high intensity load bearing performance and the Army physical fitness test. Natick, MA. U.S. Army Research Institute of Environmental Medicine. Technical report no. T30–87.

Kraemer, W. J., Mazzetti, S. A., Nindl, B. C., Gotshalk, L. A., Volek, J. S., Bush, J. A., Marx, J. O., Dohi, K., Gómez, A. L., Miles, M., Fleck, S. J., Newton, R. U., & Häkkinen, K. 2001. Effect of resistance training on women's strength/power and occupational performances. Medicine & Science in Sports & Exercise 33 (6), 1011–1025.

Kraemer, W. J., Vescovi, J. D., Volek, J. S., Nindl, B. C., Newton, R. U., Patton, J. F., Dziados, J. E., French, D. N., & Häkkinen, K. 2004. Effects of concurrent resistance and aerobic training on load-bearing performance and the Army physical fitness test. Military Medicine 169 (12), 994–999.

Kraemer, W. J. & Szivak, T. K. 2012. Strength training for the warfighter. The Journal of Strength & Conditioning Research 26, 107–118.

Kraemer, W. J., Feltwell, D., & Szivak, T. K. 2017. Physiological issues related to military personnel. In Alvar, B. A., Sell, K. & Deuster, P. A. Essentials of tactical strength and conditioning. Champaign, IL: Human Kinetics.

Krustrup, P., Mohr, M., Amstrup, T., Rysgaard, T., Johansen, J., Steensberg, A., Pedersen, P. K., & Bangsbo, J. 2003. The yo-yo intermittent recovery test: physiological response, reliability, and validity. Medicine & Science in Sports & Exercise 35 (4), 697–705.

Lechner, R., Küpper, T., & Tannheimer, M. 2018. Challenges of military health service support in mountain warfare. Wilderness & Environmental Medicine 29 (2), 266–274.

Lee, J. K., Kenefick, R. W., & Cheuvront, S. N. 2015. Novel cooling strategies for military training and operations. The Journal of Strength & Conditioning Research 29, 77–81.

Legaz-Arrese, A., Munguía-Izquierdo, D., Carranza-García, L. E., & Torres-Dávila, C. G. 2011. Validity of the Wingate anaerobic test for the evaluation of elite runners. The Journal of Strength & Conditioning Research 25 (3), 819–824.

Lericollais, R., Gauthier, A., Bessot, N., & Davenne, D. 2011. Diurnal evolution of cycling biomechanical parameters during a 60-s Wingate test. Scandinavian Journal of Medicine & Science in Sports 21 (6), 106–114.

Lester, M. E., Knapik, J. J., Catrambone, D., Antczak, A., Sharp, M. A., Burrell, L., & Darakjy, S. 2010. Effect of a 13-month deployment to Iraq on physical fitness and body composition. Military Medicine 175 (6), 417–423.

Lima, M. C. S., Ribeiro, L. F. P., Papoti, M., Santiago, P. R. P., Cunha, S. A., Martins, L. E. B., & Gobatto, C. A. 2011. A semi-tethered test for power assessment in running. International Journal of Sports Medicine 32 (7), 529–534.

Lovalekar, M., Sharp, M. A., Billing, D. C., Drain, J. R., Nindl, B. C., & Zambraski, E. J. 2018. International consensus on military research priorities and gaps – Survey results from the 4th International Congress on Soldiers' Physical Performance. Journal of Science and Medicine in Sport 21 (11), 1125–1130.

Malmberg, J. 2011. Physical fitness tests in the nordic armed forces – a description of basic test protocols. The Norwegian Defence University College, Norwegian School of Sport Sciences/Defence Institute, Oslo.

Margaria, R., Aghemo, P., & Rovelli, E. 1966. Measurement of muscular power (anaerobic) in man. Journal of Applied Physiology 21 (5), 1662–1664.

Margolis, L. M., Murphy, N. E., Martini, S., Spitz, M. G., Thrane, I., McGraw, S. M., Blatny, J. M., Castellani, J. W., Rood, J. C., Young, A. J., Montain, S. J., Gundersen, Y., & Pasiakos, S. M. 2014. Effects of winter military training on energy balance, whole-body protein balance, muscle damage, soreness, and physical performance. Applied Physiology, Nutrition, and Metabolism 39 (12), 1395–1401.

Markovic, G., Dizdar, D., Jukic, I., & Cardinale, M. 2004. Reliability and factorial validity of squat and countermovement jump tests. Journal of Strength and Conditioning Research 18 (3), 551–555.

Markovic, G., & Jaric, S. 2005. Scaling of muscle power to body size: the effect of stretch-shortening cycle. European Journal of Applied Physiology 95 (1), 11–19.

Markovic, G., & Jaric, S. 2007. Is vertical jump height a body size-independent measure of muscle power? Journal of Sports Sciences 25 (12), 1355–1363.

Matic, M. S., Pazin, N. R., Mrdakovic, V. D., Jankovic, N. N., Ilic, D. B., & Stefanovic, D. L. 2015. Optimum drop height for maximizing power output in drop jump: the effect of maximal muscle strength. The Journal of Strength & Conditioning Research 29 (12), 3300–3310.

McArdle, W. D., Katch, F. I., & Katch, V. L. 2014. Exercise physiology – nutrition, energy, and human performance. 8th edition. Lippincott Williams & Wilkins, USA.

Medbø, J. I., Mohn, A. C., Tabata, I., Bahr, R., Vaage, O., & Sejersted, O. M. 1988. Anaerobic capacity determined by maximal accumulated O_2 deficit. Journal of Applied Physiology 64 (1), 50–60.

Milligan, G. S., Reilly, T. J., Zumbo, B. D., & Tipton, M. J. 2016. Validity and reliability of physical employment standards. Applied Physiology, Nutrition, and Metabolism 41 (6), 83–91.

Minahan, C., Chia, M., & Inbar, O. 2007. Does power indicate capacity? 30-s Wingate anaerobic test vs. maximal accumulated O_2 deficit. International Journal of Sports Medicine 28 (10), 836–843.

Mäkinen, T. M. 2007. Human cold exposure, adaptation, and performance in high latitude environments. American Journal of Human Biology 19 (2), 155–164.

Nagai, T., Abt, J. P., Sell, T. C., Keenan, K. A., McGrail, M. A., Smalley, B. W., & Lephart, S. M. 2016. Effects of deployment on musculoskeletal and physiological characteristics and balance. Military Medicine 181 (9), 1050–1057.

Nagai, T., Lovalekar, M., Wohleber, M. F., Perlsweig, K. A., Wirt, M. D., & Beals, K. 2017. Poor anaerobic power/capability and static balance predicted prospective musculoskeletal injuries among Soldiers of the 101st Airborne (Air Assault) Division. Journal of Science and Medicine in Sport 20, 11–16.

Nedeljkovic, A., Mirkov, D. M., Pazin, N., & Jaric, S. 2007. Evaluation of Margaria staircase test: the effect of body size. European Journal of Applied Physiology 100 (1), 115–120.

Nindl, B. C., Barnes, B. R., Alemany, J. A., Frykman, P. N., Shippee, R. L., & Friedl, K. E. 2007. Physiological consequences of US Army Ranger training. Medicine & Science in Sports & Exercise 39 (8), 1380–1387.

Nindl, B. C., Castellani, J. W., Warr, B. J., Sharp, M. A., Henning, P. C., Spiering, B. A., & Scofield, D. E. 2013. Physiological Employment Standards III: physiological challenges and consequences encountered during international military deployments. European Journal of Applied Physiology 113 (11), 2655–2672.

Nindl, B. C., Alvar, B. A., Dudley, J. R., Favre, M. W., Martin, G. J., Sharp, M. A., Warr, B. J., Stephenson, M. D., & Kraemer, W. J. 2015. Executive summary from the National Strength and Conditioning Association's second Blue Ribbon Panel on military physical readiness: Military physical performance testing. The Journal of Strength & Conditioning Research 29, 216–220.

Nindl, B. C., Jones, B. H., Van Arsdale, S. J., Kelly, K., & Kraemer, W. J. 2016. Operational physical performance and fitness in military women: physiological, musculoskeletal injury, and optimized physical training considerations for successfully integrating women into combat-centric military occupations. Military Medicine 181 (Suppl.1), 50–62.

Nindl, B. C., Eagle, S. R., Frykman, P. N., Palmer, C., Lammi, E., Reynolds, K., Allison, K., & Harman, E. 2017. Functional physical training improves women's military occupational performance. Journal of Science and Medicine in Sport 20, 91–97.

Noordhof, D. A., De Koning, J. J., & Foster, C. 2010. The maximal accumulated oxygen deficit method: a valid and reliable measure of anaerobic capacity? Sports Medicine 40 (4), 285–302.

Nummela, A., Alberts, M., Rijntjes, R. P., Luhtanen, P., & Rusko, H. 1996. Reliability and validity of the maximal anaerobic running test. International Journal of Sports Medicine 17 (2), 97–102.

Ode, J. J., Pivarnik, J. M., Reeves, M. J., & Knous, J. L. 2007. Body mass index as a predictor of percent fat in college athletes and nonathletes. Medicine and Science in Sports and Exercise 39 (3), 403–409.

Orr, R. M. & Pope, R. R. 2015. Load carriage: An integrated risk management approach. The Journal of Strength & Conditioning Research 29, 119–128.

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

Parsons, I. T., Stacey, M. J., & Woods, D. R. 2019. Heat adaptation in military personnel: mitigating risk, maximizing performance. Frontiers in Physiology 10, 1485.

Pierce, J. R., DeGroot, D. W., Grier, T. L., Hauret, K. G., Nindl, B. C., East, W. B., McGurk, M. S., & Jones, B. H. 2017. Body mass index predicts selected physical fitness attributes but is not associated with performance on military relevant tasks in US Army soldiers. Journal of Science and Medicine in Sport 20, 79–84.

Pihlainen, K., Santtila, M., Häkkinen, K., Lindholm, H., & Kyröläinen, H. 2014. Cardiorespiratory responses induced by various military field tasks. Military Medicine 179 (2), 218–224.

Pihlainen, K., Santtila, M., Häkkinen, K., & Kyröläinen, H. 2018. Associations of physical fitness and body composition characteristics with simulated military task performance. The Journal of Strength & Conditioning Research 32 (4), 1089–1098.

Pihlainen, K., Pesola, A. J., Helén, J., Ojanen, T., Vaara, J., Santtila, M., Häkkinen, K., Finni, T., & Kyröläinen, H. 2020. Changes in quadriceps and hamstring EMG activity and performance in a military simulation test after a 6-month military operation including combined strength and endurance training. Manuscript submitted for publication.

Rappole, C., Grier, T., Anderson, M. K., Hauschild, V., & Jones, B. H. 2017. Associations of age, aerobic fitness, and body mass index with injury in an operational Army brigade. Journal of Science and Medicine in Sport 20, 45–50.

Reilly, T. J., Gebhardt, D. L., Billing, D. C., Greeves, J. P., & Sharp, M. A. 2015. Development and implementation of evidence-based physical employment standards: key challenges in the military context. The Journal of Strength & Conditioning Research 29, 28–33.

Rintamäki, H., Kyröläinen, H., Santtila, M., Mäntysaari, M., Simonen, R., Torpo, H., Mäkinen, T., Rissanen, S., & Lindholm, H. 2012. From the subarctic to the tropics: effects of 4-month deployment on soldiers' heat stress, heat strain, and physical performance. The Journal of Strength & Conditioning Research 26, 45–52.

Roberts, D., Gebhardt, D. L., Gaskill, S. E., Roy, T. C., & Sharp, M. A. 2016. Current considerations related to physiological differences between the sexes and physical employment standards. Applied Physiology, Nutrition, and Metabolism 41 (6), 108–120.

Roy, T. C., Ritland, B. M., Knapik, J. J., & Sharp, M. A. 2012a. Lifting tasks are associated with injuries during the early portion of a deployment to Afghanistan. Military Medicine 177 (6), 716–722.

Roy, T. C., Knapik, J. J., Ritland, B. M., Murphy, N., & Sharp, M. A. 2012b. Risk factors for musculoskeletal injuries for soldiers deployed to Afghanistan. Aviation, Space, and Environmental Medicine 83 (11), 1060–1066.

Petersen, S. R., Anderson, G. S., Tipton, M. J., Docherty, D., Graham, T. E., Sharkey, B. J., & Taylor, N. A. 2016. Towards best practice in physical and physiological employment standards. Applied Physiology, Nutrition, and Metabolism 41 (6), 47–62.

Pihlainen, K., Pesola, A. J., Helén, J., Ojanen, T., Vaara, J., Santtila, M., Häkkinen, K., Finni, T., & Kyröläinen, H. 2020. Quadriceps and hamstring EMG activity patterns in a military simulation test before and after a 6-month military operation. Manuscript submitted for publication.

Santtila, M., Kyröläinen, H., & Häkkinen, K. 2009. Changes in maximal and explosive strength, electromyography, and muscle thickness of lower and upper extremities induced by combined strength and endurance training in soldiers. The Journal of Strength & Conditioning Research 23 (4), 1300–1308.

Sharp, M. A., Knapik, J. J., Walker, L. A., Burrell, L., Frykman, P. N., Darakjy, S. S., Lester, M. E., & Marin, R. E. 2008. Physical fitness and body composition after a 9-month deployment to Afghanistan. Medicine & Science in Sports & Exercise 40, 1687–1692.

Sharp, M. A., Cohen, B. S., Boye, M. W., Foulis, S. A., Redmond, J. E., Larcom, K., Hydren, J. R., Gebhardt, D. L., Canino, M. C., Warr, B. J. & Zambraski, E. J. 2017. U.S. Army physical demands study: identification and validation of the physically demanding tasks of combat arms occupations. Journal of Sciearpnce and Medicine in Sport 20, 62–67.

Spiering, B. A., Walker, L. A., Hendrickson, N. R., Simpson, K., Harman, E. A., Allison, S. C., & Sharp, M. A. 2012. Reliability of military-relevant tests designed to assess soldier readiness for occupational and combat-related duties. Military Medicine 177 (6), 663–668.

Taipale, R. S., Heinaru, S., Nindl, B. C., Vaara, J. P., Santtila, M., Häkkinen, K., & Kyröläinen, H. 2015. Hormonal responses to active and passive recovery after load carriage. The Journal of Strength & Conditioning Research 29, 149–153.

Taylor, N. A. 2015. Overwhelming physiological regulation through personal protection. The Journal of Strength & Conditioning Research 29, 111–118.

Taylor, N. A., Peoples, G. E., & Petersen, S. R. 2016. Load carriage, human performance, and employment standards. Applied Physiology, Nutrition, and Metabolism 41 (6), 131–147.

Tesch, P. A., Daniels, W. L., & Sharp, D. S. 1982. Lactate accumulation in muscle and blood during submaximal exercise. Acta Physiologica 114 (3), 441–446.

Tipton, M. J., Milligan, G. S., & Reilly, T. J. 2013. Physiological employment standards I. Occupational fitness standards: objectively subjective? European Journal of Applied Physiology 113 (10), 2435–2446.

van den Heuvel, A. M., Kerry, P., van der Velde, J. H., Patterson, M. J., & Taylor, N. A. 2010. Torso undergarments: Their merit for clothed and armored individuals in hot-dry conditions. Aviation, Space, and Environmental Medicine 81 (12), 1107–1113.

Vanderburgh, P. M. 2007. Correction factors for body mass bias in military physical fitness tests. Military Medicine 172 (7), 738–742.

Vandewalle, H., Pérès, G., & Monod, H. 1987. Standard anaerobic exercise tests. Sports Medicine 4 (4), 268–289.

Viljanen, T., Viitasalo, J. T., Kujala, U. M. 1991. Strength characteristics of a healthy urban adult population. European Journal of Applied Physiology 63, 43–47.

Wardle, S. L. & Greeves, J. P. 2017. Mitigating the risk of musculoskeletal injury: a systematic review of the most effective injury prevention strategies for military personnel. Journal of Science and Medicine in Sport 20, 3–10.

Warr, B. J., Heumann, K. J., Dodd, D. J., Swan, P. D., & Alvar, B. A. 2012. Injuries, changes in fitness, and medical demands in deployed National Guard soldiers. Military Medicine 177 (10), 1136–1142.

Zagatto, A. M., Beck, W. R., & Gobatto, C. A. 2009. Validity of the running anaerobic sprint test for assessing anaerobic power and predicting short-distance performances. The Journal of Strength & Conditioning Research 23 (6), 1820–1827.

Zagatto, A. M., Bertuzzi, R., Miyagi, W. E., Padulo, J., & Papoti, M. 2016. MAOD determined in a single supramaximal test: a study on the reliability and effects of supramaximal intensities. International Journal of Sports Medicine 37 (9), 700–707.