



Original research

Reliability of FORCE COMBAT™: A Canadian army fitness objective

Tara Reilly*, Evan Walsh, Barry Stockbrugger

Human Performance Research and Development, Canadian Forces Morale and Welfare Services, Department of National Defence, Canada



ARTICLE INFO

Article history:

Received 15 August 2018
 Received in revised form 25 October 2018
 Accepted 12 November 2018
 Available online 22 November 2018

Keywords:

Military
 Physical fitness testing
 Physical employment standard
 Marching

ABSTRACT

Objectives: The Canadian Army fitness objective is FORCE COMBAT™. This consists of a 5 km march (35 kg) between 50–60 min, and the annual FORCE physical employment standard as a circuit without the usual 5 min rest intervals (FORCE circuit) in full fighting order (25 kg). The objective of this research was to determine the number of practice attempts required to establish reliability of FORCE COMBAT™. Additionally, this study aims to identify the minimal detectable change once reliability is established.

Design: The study used a within participant design.

Methods: 33 Canadian Army members were divided into two groups. Group 1 performed four maximum effort attempts of FORCE COMBAT™ completely. Group 2 performed only the FORCE circuit component, not the loaded march. On the fifth trial, the tasks of the groups were switched. Variability within and between trials of the FORCE circuit of FORCE COMBAT™ were analysed based on six statistical factors of reliability.

Results: Four statistical factors indicated that one maximal effort practice attempt of FORCE COMBAT™ is required, as reflected in the variability of performance between trials 1 and 2. The minimum detectable change, highlighting inherent measurement error of the test was 89.18 s with 95% confidence.

Conclusions: Results suggest that Canadian Army members should complete one FORCE COMBAT™ in completion at maximal effort, before formal assessment. Further, members performing within 89 s of the recommended completion time, on FORCE COMBAT™ should be offered a re-attempt (in its entirety).

Crown Copyright © 2018 Published by Elsevier Ltd on behalf of Sports Medicine Australia. All rights reserved.

1. Introduction

In April 2014, the Fitness for Operational Requirements of CAF Employment (FORCE) evaluation was implemented to assess for the minimum physical fitness required for Canadian Armed Forces (CAF) employment.¹ The FORCE evaluation, and standard applies to all members of the CAF, regardless of age, gender, rank, occupation, or trade. The FORCE evaluation consists of four physical test elements administered with 5 min rest periods in between: (1) 20 m Rushes (20mR); (2) Sandbag Lift (SBL); (3) Intermittent Loaded Shuttle (ILS); and (4) Sandbag Drag (SBD). The Canadian Army (CA) expressed concern that the minimum standards required by the FORCE evaluation may not account for the demands of certain CA occupations and roles, particularly those operating in urban environments.

The directly measured aerobic demands of FORCE at the minimum standard are 18–25 mL/kg/min of O₂ for 9 min, exclusive of the 5 min rest intervals between test items.¹ Physiological

data collected on CA members during urban military exercises, specifically rescuing casualties from 3 story buildings, resulted in oxygen demands measured at 25–32 mL/kg/min of O₂ for a 5 min scenario.^{2,3} These durations are reflective of those measured in the development of a validated dismounted assault task simulation for the Australian Defence Force.⁴ The FORCE evaluation, when performed with loads ranging from 15–25 kg and as a circuit without the mandated 5 min rest periods between test items, was found to elicit metabolic demands similar to those experienced during combat or working in urban environments.⁵ Following an extensive process to define and address the needs of the CA, a modification to the FORCE evaluation, FORCE COMBAT™, was developed and trialed as a potential fitness objective to guide members in preparing for the physical demands of urban operations. FORCE COMBAT™ is not considered a physical employment standard (PES), as employment is not determined on the basis of success on FORCE COMBAT™. However to complete Individual Battle Task Standards (IBTS) in preparation for deployment, a CA member must successfully complete FORCE COMBAT™. Currently, success is determined by completing the Load Bearing March (LBM) component in between 50–60 min and, following a 5 min rest, completing FORCE as a circuit in less than 15 min (Fig. 1).

* Corresponding author.

E-mail address: Tara.reilly@forces.gc.ca (T. Reilly).



Fig. 1. FORCE COMBAT™: The march and the FORCE Circuit.

The FORCE COMBAT™ physical fitness battle task standard, begins with a 5 km LBM in Battle Order (BO), consisting of Fighting Order (FO), as defined by the CA Combat Training Centre Infantry School to include: 25 kg combat fatigues (boots, helmet, weapon, full tactical vest, fragmentation vest with ballistic plates) and a 10 kg day bag. The LBM is self-paced but must be completed between 50–60 min. Following the LBM is a 5 min rest period, in which the 10 kg bag is removed and FO is maintained. The timed FORCE circuit is then completed in FO without the rest periods. Success on FORCE COMBAT™ is determined as completion of the LBM in the prescribed time, and completion of the FORCE circuit in less than 15 min. FORCE COMBAT™ is age and gender free, and the details of its application and policies regarding FORCE COMBAT™ directives can be found on www.forcecombat.ca, where published military guidance from the CA is available.

Typically for physical fitness evaluations, especially those with as many degrees of freedom as FORCE COMBAT™'s five components, it is important to perform reliability trials.⁶ Reliability refers to a measure of consistency (reproducibility) within the data. To ensure minimal measurement error, systematic (learning effects and fatigue) and random error (mechanical variation of participants/facilitator) must be taken into consideration.^{7,8}

Previous research on skill based, loaded, circuit-type fitness tests indicate that maximum performance is achieved by the fourth trial.^{8,9} Additionally, increased task complexity has shown to increase degrees of freedom and potential for improvement through learning or strategy development.¹⁰ This is further supported by Pandorf et al.¹¹, who evaluated the reliability of a six element, indoor, obstacle course designed to simulate impediments to movement of a soldier during a conflict. Performance improvements of 4% from trials 1–2 and 3% from trials 2–3 were captured, along with an intra-class correlation coefficient (ICC) of 0.92 on trial 4 and mean coefficient of variation (CV) of 4.1% across all trials¹¹. Factors such as strategy development and skill improvement are known to affect reliability of performing unfamiliar tasks without actual physiological adaptation.¹¹

As FORCE COMBAT™ is included in the IBTS for CA, the purpose of this study was to investigate the number of times CA members should complete FORCE COMBAT™ to achieve reliable FORCE circuit performance times. Once reliability is achieved, the remaining error can be attributed to biological variation and could result in the implementation of an inclusion zone around the cut-score.^{8,10,12}

It is hypothesized that:

H1. Reliability of the FORCE circuit performance can be established as a measure of proficiency of participants to achieve consistent performances, such that proficiency is achieved under the following conditions within the six statistical factors: (1) the change in the mean must not be significantly different between subsequent trials¹⁰; (2) the 95% minimum detectable change must be larger than the change in the mean¹³; (3) the typical error must be less than the smallest worthwhile change¹⁴; (4) the coefficient

of variation must be less than 5% and preferably approaching 2.2%⁸; (5) the intra-class correlation coefficient must exceed 0.75¹⁵; and (6) the effect size should be below 0.20 or considered small.¹⁶

H2. The inclusion of the LBM before the FORCE circuit will result in slower circuit times, and that when practiced FORCE COMBAT™ should be performed in completion (LBM + FORCE circuit).

2. Methods

Thirty-three military personnel participants were recruited from the CA. Of those recruited, 24 male CA members completed a VO₂max and all FORCE COMBAT™ trials and therefore were included in the analysis. Females were included in the original sample of 33, however no females completed all trials. Demographics for these participants separated by experimental group are shown in Table 1. Twelve of these participants were designated as the “Marchers” who completed the first four repeated trials of FORCE COMBAT™ in its entirety with the LBM. The remaining twelve participants were designated as the “Non-Marchers” who did not complete the LBM before the FORCE circuit. Participants were assigned to either group to match aerobic power, as determined with absolute VO₂ max, and lean body mass. On the fifth trial, the groups switched tasks such that Marchers only performed the FORCE circuit and the Non-Marchers performed all of FORCE COMBAT™. This experimental design was included to assess the influence of the LBM on FORCE circuit performance after performance has stabilised. If the LBM did not influence performance on the FORCE circuit, the recommendation would be to practice the FORCE circuit only, given that to include the LBM requires an additional hour.

All participants were pre-screened using the Canadian Society for Exercise Physiology's Physical Activity Readiness Questionnaire (PAR-Q+).¹⁷ Ethical approval was obtained from the Defence Research and Development Canada Human Research Ethics Committee, and all participants provided informed consent.

Waist Circumference was obtained as the participant stood erect with feet shoulder width apart and arms crossed over the chest.¹⁸ Height was obtained using a Seca 213 Portable Stadiometer (Seca Industries, Hanover Maryland). Bioelectric Impedance was used to obtain Lean Body Mass, Percent Body Fat and mass using the InBody 520 (BioSpace Technologies, Los Angeles, California).

Once participants had completed the intake procedures, they were provided with an orientation on a treadmill VO₂max test and an explanation of the testing over 6 sessions and spanning 3 weeks. Participants were instructed to maintain their normal activity level during the participation period (3 weeks) of the study and to refrain from strenuous physical activity during the 6 h prior to all test sessions. As these participants were in Garrison, this required them to forfeit morning physical training sessions. There was a minimum 72 h between testing sessions for rest and recovery, specific to loaded maximal effort activity.

Table 1
FORCE COMBAT™ participant demographics by group.

	Marchers N = 12 Mean ± SD (Range)	Non-Marchers N = 12 Mean ± SD (Range)	Sig. Diff.
Age, years	29 ± 5 (21 – 38)	29 ± 7 (19 – 46)	
Height, cm	180 ± 6 (171 – 188)	178 ± 9 (162 – 190)	
Mass, kg	92 ± 12 (77 – 114)	83 ± 14 (54 – 101)	
Waist Circ., cm	94 ± 8 (85 – 110)	92 ± 11 (72 – 110)	
Percent body fat, %	22 ± 5 (15 – 32)	19 ± 8 (7 – 36)	
Lean body mass, kg	71 ± 7 (63 – 85)	67 ± 9 (50 – 82)	
VO ₂ max, mL/kg/min	43.4 ± 4.8 (32.6 – 49.2)	51.0 ± 9.5 (36.8 – 71.3)	p < 0.05
VO ₂ max, L/min	3.93 ± 0.37 (3.45 – 4.41)	4.18 ± 0.72 (3.28 – 5.51)	

All participants wore heart rate monitors (Polar RS800CX, Polar Technologies Inc., Jyväskylä, Finland) for all testing sessions, to verify that a best effort performance was being elicited. Maximum heart rate for each participant was identified during their VO₂max testing.

In Session 1 all participants completed a VO₂max test using a modified Balke graded exercise treadmill procedure.^{19,20} The modified Balke protocol uses a constant speed and grade increases of 2% every 2 min beginning at the end of minute 2.

Session 2 through session 6: On Day 2 participants were divided into their two groups, Marchers and Non-Marchers. On day 6, the group tasks were switched. Time of day was maintained for the participants throughout the repeated trials in an attempt to control for diurnal effects. Testing was completed on the fieldhouse rubberized 200 m indoor track for the LBM and polished hardwood gym floor for the FORCE circuit. Each test was conducted in the same manner, at maximal effort, after a standardised warm up, and timings were taken by trained research or fitness personnel employed for the purposes of fitness testing of the FORCE evaluation, the PES. Environmental conditions were consistent at approximately 20 °C. All participants were familiar with the four elements of the circuit as they are annually required to perform the FORCE evaluation. There are no circumstances where a CA member would be exposed to FORCE COMBAT™ before having passed the FORCE evaluation. None of the participants had been previously exposed to FORCE COMBAT™ as this was its first year of implementation.

Data were analyzed using Microsoft Excel and IBM SPSS. Changes in the mean performance times between consecutive trials was evaluated using a paired sample t-test. The Minimum Detectable Change with a 95% confidence interval (MDC₉₅) is a component in the analysis of determining if changes in performance are due to true change or random measurement error. It was calculated as $MDC_{95} = SD_{diff} \times 1.96$, where SD_{diff} is the standard deviation of the difference between consecutive trials.

Typical Error (TE) or systematic bias is defined as the random variation in a measure when a participant is tested multiple times on a given test item. It is a type of standard error of measurement and can be interpreted as the amount of inherent test error that could be reasonably expected when a participant completes a “best effort” trial. Typical Error was calculated as $TE = SD_{diff}/\sqrt{2}$, where again SD_{diff} is the standard deviation of the difference between consecutive trials.

In contrast to the TE, the Smallest Worthwhile Change (SWC) in FORCE circuit time performance considers the pooled standard deviation between consecutive trials multiplied by a weighting factor of 0.2 (based on Cohen's effect size). It has been suggested that one indication of reliable performance would be if TE is less than SWC, then change in performance on the test is related to a real change and not simply error.¹⁴

Coefficient of variation (CV) is expressed as a percent value of the standard deviation of consecutive trials divided by the mean of the trials and represents the variability between successive trials. It was calculated as $CV = SD/\bar{x} \times 100$.

Table 2
FORCE COMBAT™ circuit performance times for the Marcher and Non-marcher groups.

Research group	Trial	FORCE circuit time, mm:ss			
		Minimum	Maximum	Mean	Standard deviation
Marchers	1	06:22	14:55	09:07	02:20
	2	06:05	13:59	08:17	02:01
	3	06:00	13:00	08:01	01:48
	4	06:09	13:33	08:00	01:58
	Circuit only	06:06	12:07	07:23	01:41
Non-Marchers	1	05:56	10:36	07:48	01:32
	2	05:34	09:39	07:30	01:15
	3	05:37	09:02	07:15	01:05
	4	05:30	09:07	07:10	01:11
	LBM and circuit	05:44	10:46	07:32	01:35

The Intra-class Correlation Coefficient (ICC) describes the similarity of performances of individual participants within the same group between two trials and is rated as poor (<0.40), fair (0.40 – 0.59), good (0.60 – 0.74), and excellent (>0.74).¹⁵

Effect Size (ES) is determined by the Cohen's d statistic calculated as $d = \bar{x}_1 - \bar{x}_2 / SD_{pooled}$, where \bar{x}_1 and \bar{x}_2 are the respective means for consecutive trials and SD_{pooled} is the pooled standard deviation for the two trials. If performance is stable or reliable between consecutive trials, ES should be less than 0.2 and considered small.

3. Results

Effort was determined to be maximal, consistent, and acceptable as demonstrated by the 96–98% maximum heart rate achieved by participants, on average, in response to each FORCE COMBAT™ trial.

Table 2 contains the results for the FORCE circuit performance times for the Marcher and the Non-Marcher groups. Not including the fifth trial in which the groups' tasks were switched, mean performances for the both groups improved with each of the four trials. Despite attempts to match the two groups by absolute VO₂max (L/min) and lean body mass (kg), the loss of participants from both groups resulted in significantly different relative VO₂max values between the Marchers and Non-Marchers. No other significant differences in demographic data were recorded between the two groups, however the circuit results are consistently faster for the Non-Marcher group.

The inter-trial reliability statistics for the Marchers are found in Table 3. As stated in H1, six factors are being assessed to determine the required number of practice trials to achieve a consistent or reliable FORCE circuit performance time.

The significant changes in mean values between trials 1-2 and 2-3 and non-significant changes in trials 3-4 suggests that two practice trials are required before stable performance is achieved.

Table 3
Summary of reliability statistics for FORCE COMBAT™ performance times for the marchers (N = 12) and Non-marchers (N = 12) over four trials.

Research group	Reliability statistic	Trials 1-2	Trials 2-3	Trials 3-4
Marchers	Change in mean, s	-50.0*	-16.6*	-0.7
	MDC ₉₅ , s	89.18	42.07	58.74
	TE, s	32.17	15.18	21.19
	SWC, s	26.18	22.91	22.58
	CV, %	6.56	2.69	3.70
	ICC	0.95	0.99	0.97
	ES	0.38	0.14	0.01
Non-Marchers	Change in mean, s	-17.3*	-12.9	-2.2
	MDC ₉₅ , s	44.85	43.07	49.49
	TE, s	16.18	15.54	17.86
	SWC, s	15.28	12.65	12.90
	CV, %	2.68	2.78	3.48
	ICC	0.96	0.95	0.94
	ES	0.23	0.20	0.03

MDC₉₅ = Minimum detectable change, TE = Typical error, SWC = Smallest worthwhile change,

CV = Coefficient of variation, ICC = Intra-class correlation coefficient, ES = Effect size.

* Significant difference ($p < 0.05$) using paired sample t-test.

MDC₉₅, which indicates the biological variability with 95% confidence, was found to range between 89.18 and 42.07 s. As MDC₉₅ values are greater than the associated changes in the means, it does not delimit the number of practice trials needed before assessing FORCE COMBAT™. Change between consecutive trials would need to be greater the MDC to be able to be 95% confident that the change is real. TE for the FORCE circuit ranges from 32.17 s for trials 1-2 to 15.18 s for trials 2-3. As such, a change in performance of up to 32 s between trials may be attributable to typical or systematic error. SWC between trials were found to range between 22.58 and 26.18 s. TE (32.17) was found to be greater than SWC (26.18) for trials 1-2, and less than SWC for trials 2-3 and 3-4. If the SWC change is greater than TE we can attribute the change to biological variability and not systematic error. This supports the need for one practice trial before a reliable test can be administered.

The decrease in CV values between trials 1-2 (6.56%) and subsequent trials (<5%), demonstrate less variability in later trials, further supporting the role of a single practice trial suggested previously by the TE to SWC ratio.

The ICC improved from an already strong 0.95 between trials 1-2-0.99 and 0.97 in the subsequent trials, demonstrating within-group agreement between consecutive trials. This also suggest that ICC was not a limiting factor in determining the number of required trials before reliable results are achieved.

Effect Size of FORCE COMBAT™ performance times for the Marchers decreased from 0.38 for trials 1-2-0.07 for trials 2-3, which meets the threshold of 0.2. This finding further supports that only a single practice trial is needed before reliable FORCE COMBAT™ results are achieved.

4. Discussion

This research examined the reliability of the time to complete the FORCE circuit component of FORCE COMBAT™, the fitness test in IBTS for the Canadian Army. The results demonstrated maximal effort from the participants as evidenced by their heart rate responses, which is necessary to assess inter-trial variability in determining test reliability.²¹ Similarly, consistent heart rate responses have shown to reduce the influence of learning effects and motivational differences across trials.¹⁴

The results from this study suggest that the largest improvements on the FORCE circuit within the marcher group (who were all familiar with the evaluation) were found between trials 1-2 with an average 9.1% decrease in circuit time. Subsequent improvements

between trials 2-3, and 3-4 were 3.3% and 0.1% respectively. When reviewing all of the statistical analysis of the various reliability criteria the conclusion is that one practice trial of FORCE COMBAT™ is required to elicit consistent and reliable test results. Statistically significant changes in mean circuit times between trials 2-3 would offer that two practice trials be required for reliable results; however, MDC₉₅ of the change in mean and ICC values indicate that no additional practice trials (above 1) would be required.

The practice periods demonstrated within this study were shorter than the four trials suggested by Boyd⁸ and Rogers.⁹ This may be attributable to the CA members' practice to the unloaded test elements of the circuit, having annually performed the normal FORCE evaluation.

As hypothesized, the difference in performance of the two groups is likely associated to an absence of the pre-fatiguing LBM for the Non-Marchers, but also possibly influenced by drop out of eight of the participants (5 original march and 3 original non-march). This may have unbalanced the aerobic power between groups, which can be demonstrated by the fact that mean FORCE circuit time for the Non-Marchers' fifth trial (with LBM), is nearly a half-minute faster than that of any trials of the Marcher group.

Further analysis included an examination of Relative Recovery Status (RRS) as indicated by the Excess Post-exercise Oxygen Consumption (EPOC)²² in mls/kg for each participant at the end of the 5 min rest, just before starting the FORCE circuit. To calculate a RRS for each participant EPOC was divided by their VO_{2max} values. RRS demonstrated a relationship with FORCE COMBAT™ with a R² (95% confidence) of 0.74, confirming the theory that the participants (non-marchers) with higher aerobic power had recovered to a greater extent than those with lower aerobic power, and therefore were faster on the FORCE circuit.

As guided by current research,^{8,10,12} once a pass/fail cut-score and a number of practice trials are established for FORCE COMBAT™, the associated MDC₉₅ can be used to establish a borderline category based on the inherent error within the test. For example, if one practice trial is implemented, a CA member within 89s (Table 3) of a cut-off score on their final attempt should be given a second opportunity to achieve the 15:00. Likewise, if two practice trials are used, Performance within 42 s of the 15:00 would warrant a second attempt. The recommendation to those CA members who do not achieve even the lower limit of the borderline category would be to undergo targeted physical training before attempting the test again. For the purpose of developing a feasible recommendation for the CA, one practice trial is recommended.

Limitations of this research include the access to CA in Garrison over repeat trials, which resulted in 9/33 participants unable to perform all sessions due to occupational and operational commitments. This included the females originally participating as well as an unbalanced design with regard to aerobic power, which our investigations indicate is influential to performance on the FORCE circuit component of FORCE COMBAT™.

5. Conclusions

The present study identified the necessary number of practice trials on FORCE COMBAT™, a CA member should complete before attempting a timed trial of the physical fitness check to ensure reliable results. Based on the findings, it is the authors' suggestion that CA members who are already familiar with the FORCE evaluation should complete one FORCE COMBAT™ practice trial before being assessed. Further, if a member does not meet established cut-off times on the FORCE circuit but is within 89 s of the cut-off score, they should be allowed to re-attempt FORCE COMBAT™ in its entirety before formally failing the test.

Practical implications

- One FORCE COMBAT™ practice trial should be provided before being timed on the second attempt.
- The FORCE COMBAT™ practice trial should include the 5 km LBM and not only the FORCE circuit component as the march has a pre-fatiguing effect and the potential to modify one's pacing strategy
- Should the first formal attempt result in an inability to achieve the cut-off time, a borderline category has been established for which a second attempt should be permitted

Acknowledgements

This work was supported by Canadian Forces Morale and Welfare Services, an agency of DND, Government of Canada.

We would like to thank the members of the CA for their participation in the research as well as the additional team members of the Human Performance Research Team.

References

1. Gagnon P, Spivock M, Reilly T et al. The FORCE fitness profile – adding a measure of health-related fitness to the Canadian armed forces operational fitness evaluation. *J Strength Cond Res* 2015; 29(Suppl. 11):S192–S198. <http://dx.doi.org/10.1519/JSC.0000000000001163>.
2. Reilly T, Olinek S, McGinn M. A current performance evaluation for the casualty evacuation component of the Canadian land forces command, P. Gagnon, *2nd International Congress on Soldiers' Physical Performance*, 2011.
3. Reilly T, Olinek S, Driscoll C. The physical demands of conducting operations in the urban environment, *CMVHR 4th Annual Canadian Military and Veterans Health Research Forum*, 2013.
4. Silk AJ, Billing DC. Development of a valid simulation assessment for a military dismounted assault task. *Mil Med* 2013; 178(3):315–320. <http://dx.doi.org/10.7205/MILMED-D-12-00294>.
5. Reilly T, Stockbrugger B, Saucier S et al. Physiological demands of FORCEcombat: An Fitness Objective for the Canadian Army, *Canadian Institute Military Veterans Health Research Forum*, 2016.
6. Milligan G, Reilly T, Zumbo B et al. Validity and reliability of physical employment standards. *Appl Physiol Nutr Metab* 2016; 41:S83–S91.
7. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998; 26(4):217–238.
8. Boyd L, Rogers T, Docherty D et al. Variability in performance on a work simulation test of physical fitness for firefighters. *Appl Physiol Nutr Metab* 2015; 40:364–370. <http://dx.doi.org/10.1139/apnm-2014-0281>.
9. Rogers TW, Docherty D, Petersen S. Establishment of performance standards and a cut-score for the Canadian Forces Firefighter Physical Fitness Maintenance Evaluation (FF PFME). *Ergonomics* 2014; 57(11):1750–1759. <http://dx.doi.org/10.1080/00140139.2014.943680>. PMID:25102916.
10. Stockbrugger BG, Reilly TJ, Blacklock RE. Reliability of the individual components of the Canadian Armed Forces Physical Employment Standard. *Appl Physiol Nutr Metab* 2018; 999:1–6.
11. Pandorf CE, Nindl BC, Montain SJ et al. Reliability assessment of two military relevant occupational physical performance tests. *Can J Appl Physiol* 2013; 28(1):27–37.
12. Tipton MJ, Milligan GS, Reilly TJ. Physiological employment standards I. Occupational fitness standards: objectively subjective? *EJAP* 2013; 113(10):2435–2446.
13. Hachana Y, Attia A, Nassib S et al. Test–retest reliability: criterion-related validity, and minimal detectable change of score on an abbreviated Wingate test for field sport participants. *J Strength Cond Res* 2010; 26(5):1324–1330.
14. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000; 30:1–15.
15. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994; 6(4):284–290. <http://dx.doi.org/10.1037/1040-3590.6.4.284>.
16. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed Hillsdale, NJ, Lawrence Erlbaum Associates, 1988. p. 109–144.
17. The Physical Activity Readiness Questionnaire for Everyone (PAR-Q+). 2017. Available at: http://eparmedx.com/?page_id=79. Accessed 8 April 2016.
18. Canadian Society for Exercise Physiology. *The Revision of the Measurement of Waist Circumference in the CPAFLA*, 2008. Available at: <http://www.csep.ca/KnowledgeTranslations.asp>. Accessed 08 04 2016.
19. American College of Sports Medicine. Health-related physical fitness testing and interpretation, Chapter 4, In: *ACSM's Guidelines for Exercise Testing and Prescription*. 3rd edn Philadelphia, Lippincott Williams & Wilkins, 2010. p. 71–85.
20. Balke B, Ware RW. An experimental study of physical fitness of Air Force personnel. *US Armed Forces Med J* 1959; 10(6):675–688.
21. Glaister M, Howatson G, Lockey RA et al. Familiarization and reliability of multiple sprint running performance indices. *J Strength Cond Res* 2007; 21(3):857–859.
22. Short K, Sedlock D. Excess postexercise oxygen consumption and recovery rate in trained and untrained subjects. *J Appl Phys* 1997; 83(1):153–159. <http://dx.doi.org/10.1152/jappl.1997.83.1.153>.