External Responsiveness and Intra-session Reliability of the Rope-Climbing Test

Wissem Dhahbi, Karim Chamari, Laurence Cheze, David G Behm, Anis Chaouachi

Journal of Strength and Conditioning Research Publish Ahead of Print
DOI: 10.1519/JSC.0000000000001367

External Responsiveness and Intra-session Reliability of the Rope-Climbing Test

Wissem Dhahbi[1], Karim Chamari[2], Laurence Chèze[3], David G Behm[4], Anis Chaouachi[1,5]

[1]Tunisian Research Laboratory "Sport Performance Optimisation", National Center of Medicine and Science in Sports (CNMSS), Tunis, Tunisia ; [2]Athlete Health and Performance Research Centre, ASPETAR, Qatar Orthopaedic and Sports Medicine Hospital, Doha, Qatar ; [3]Université de Lyon, Lyon, France Université Claude Bernard Lyon 1, Villeurbanne, France UMR_T9406, Laboratoire de Biomécanique et Mécanique des Chocs (LBMC), IFSTTAR, Bron, France; [4]School of Human Kinetics and Recreation, Memorial University of Newfoundland, St. John's, Newfoundland, Canada; [5]AUT University, Sports Performance Research Institute New Zealand, Auckland, New Zealand,

Corresponding author

Anis Chaouachi, PhD, HDR

Tunisian Research Laboratory "Sports Performance Optimisation"

National Center of Medicine and Science in Sports (CNMSS)

Bp263, Ave Med Ali Akid, 1004 El Menzah, Tunis, Tunisa

Tel: +216 21 17 69 40

Fax: +216 71 70 97 89

E-mail: anis.chaouachi@email.ati.tn

Preferred Running Head: Rope Climbing Test: intra-session reliability and discriminant ability

Abstract

Although the Rope Climbing Test (RCT) has been validated for upper body power assessment of Commando soldiers, the external responsiveness and intra-session reliability of the RCT have not been reported. In order to examine RCT external responsiveness and intra-session reliability, this study consisted of two separate phases. Forty male soldiers belonging to the special units of the National Guard, selected on the basis of their training and specialty operations levels, participated in the first phase of study to identify the discriminant ability of RCT. This group was then divided into anti-terrorism Commandos (21 soldiers) and Intervention-Brigade (19 soldiers). Only the anti-terrorism Commandos participated in the intra-session reliability study. Commandos were significantly better than Intervention-brigade soldiers on Execution Time (ET), Absolute Power Output (APO) and Relative Power Output (RPO-$p<0.001$). The areas under the receiver operator characteristics (ROC) curves were all higher than 0.70: 0.91, 0.85 and 0.90 for ET, APO and RPO, respectively. RCT provided good external responsiveness, thus RCT was considered to indicate "good" discriminative ability. No significant difference was found between groups in post-test rating of perceived exertion. The intra-session reliability coefficients were excellent for ET, APO and RPO (ICC[3,1]>0.90). The standard errors of measurement values for the ET, APO and RPO were all under 5% (range: 1.29-1.47%). The main findings of this study suggest that RCT is a tool with both high sensitivity and intra-session reliability, allowing the consistent detection of

differences in upper limbs' power performance between two military groups of different operational capacity levels.

Keywords: Military; Field testing; Intra-session error; Discriminant ability.

Introduction

The analysis of physical requirements of special forces soldiers (4) clearly shows that aerobic endurance, agility, muscle strength, power and endurance of the upper limbs are required (15). During the diverse operations performed by military, police and emergency (e.g. emergency medical technicians, fire fighters) personnel in their daily activities, these individuals must control their mass and their relatively heavy equipment with their upper limbs. This physical requirement is therefore of paramount importance for their overall physical performance, personal safety and safety of others. The inability to transport themselves and their equipment rapidly and reliably over or around obstacles can result in injury and possibly death. Similarly, many athletes can only perform optimally if they have sufficient relative strength and power to maneuver their body mass (e.g. gymnasts, rock climbers) as well as equipment (e.g. hockey goalies). Hence appropriate training, reliable testing and valid testing are necessary to prepare and identify those personnel that are ready to perform challenging operations and activities.

Historically, the typical methods for assessing upper limb power have been pull-ups (25), push-ups (7), bench press power test (6) and medicine ball put tests (24). To assess the power of the upper limbs, Execution Time (ET) and Relative Power Output (RPO) indices are widely used in specific tests in different sports (18), standard field tests (6) and standard laboratory tests (28). In the context of military, police, emergency medical personnel and athletes, a strong individual with lower body mass has the advantage in weight bearing tests

and activities (e.g. pull-up or rope climb tests) (3). Compared to individuals with greater muscle mass, the lighter individual is disadvantaged when required to pull, push, lift or carry an object with greater absolute mass (e.g. a victim, goalie equipment). Dhahbi et al. (11) reported that RPO was a more convenient parameter than ET in the specific Rope Climbing Test (RCT). The latter test has been recently validated for assessment of power of the upper limbs of Commando soldiers (11).

The concurrent validity, reliability and responsiveness are basic attributes used for evaluating the validity of any test in sport physiology (1, 16). The external responsiveness and intra-session reliability of RCT have not been reported. Dhahbi et al. (11) only considered the inter-session reliability and the criterion-related validity of RCT. External responsiveness determines the discriminative ability of a test and usually is assessed by testing differences between two groups of individuals with different performance profiles (16). One of the most important aims of the RCT test is to select soldiers. Thus, the external responsiveness of the RCT should discriminate between soldiers of different specialty operations levels (e.g. Commandos vs. Intervention-Brigade). The intra-session error is free of methodological errors, cannot be reduced, and thereby serves as an appropriate baseline for comparisons, remaining independent of other error sources (22). An unreliable or invalid test could allow for the placement of incapable professionals (or athletes), which could impact the safety of the individual and the dependent individuals (e.g. victims in a fire, injured victim in a car accident).

This theoretical background reveals the lack of knowledge on the assessment of the RCT to distinguish performance profiles and its intra-session reliability. Therefore, the aims of this study were to (1) investigate the discriminant ability of RCT (Commandos vs. Intervention-Brigade) and (2) to examine the absolute and relative intra-session reliabilities of RCT.

Methods

Experimental Approach to the Problem

The external responsiveness of the RCT was determined by comparing ET, Absolute Power Output (APO) and RPO between two groups of soldiers of different specialty operations levels (Commandos vs. Intervention-Brigade). During the second study phase, which aimed to establish the relative and absolute intra-session reliabilities of RCT, the experimental protocol consisted of performing 3 trials of RCT in a single session.

Subjects

Forty male soldiers belonging to the special units of the National Guard voluntarily participated (Table 1). Twenty-one Commandos soldiers were employed to investigate the discriminant ability of RCT. The inclusion criteria of Commandos soldiers was having regularly trained for at least 4 months in the National Guard School of Commandos, for ~32 h/week. Training was divided into ~14 h/week for fitness training and ~18 h/week dedicated to technical and tactical training. Another group of 19 soldiers participated from an Intervention-Brigade. The inclusion criteria of Intervention-Brigade was having trained for at least 8 weeks in the National Guard School of Intervention-Brigade/Commandos, for 4 sessions per week (1 session for strength and conditioning and 3 sessions per week for technical and tactical training), for approximately 2 hours in duration each. Both groups were used to establish external responsiveness, whereas only the anti-terrorism Commandos participated in the intra-session reliability study.

All participants were free from any injury or pain that would prevent maximal effort during performance testing. All the participants gave their written informed consent to the study after receiving a thorough explanation about the protocol. This protocol conformed to

internationally accepted policy statements regarding the use of human subjects and was approved by the University Ethics Committee in accordance with the Helsinki declaration.

Procedures

Participants were requested to follow their normal diet, eat a light meal at least 3 hours before each session, keep their usual sleep schedule, and stop any strenuous activity during the last 24 hours before the test. Seven days before baseline testing, one session was carried out to familiarize the participants with the measurement protocol. Before starting the tests, the participants achieved 15 min of standardized specific warm-up with 5 min of rest. Data were collected from participants at approximately the same time of day (between: 9:00 and 11:00 a.m.) in order to eliminate any influence of circadian variations on performance (12).

The session was performed outdoors in the following conditions (measurements monitored by a digital environmental station: VaisalaOyj, Helsinki, Finland; every 30 min during the experiment): temperature ranged from 15°C-17°C, humidity ranged from 55%-56% and the wind velocity was light (under 10 km/h). Participants performed the tests wearing the army combat uniform without a bulletproof vest and tactical foot wear (the mass of the equipment was ~5 kg). The protocol consisted of performing 3 trials of RCT, with 5 min rest between trials. The experimenter provided strong verbal encouragement during the tests so as to obtain maximum efforts. The Rate of Perceived Exertion (RPE) was recorded immediately after the RCT using the Borg scale (RPE, 1-10) (14).

5 m Rope Climbing Test (RCT)

The RCT test was performed using the criteria outlined in the investigation of Dhahbi et al. (11). The participant was instructed to climb the rope as fast as possible and hit the finish mark (see description below). The manual timer was triggered at the signal of the assessor and stopped when the participant touched the mark that was situated at a height of 5

m above the starting mark. Dhahbi et al. (10) showed excellent concurrent validity of hand timing with no significant difference between the stopwatch and video timer with a low systematic bias (0.18 sec) and very little difference in Standard Errors of Measurement (SEM) value (<5%). Moreover, Dhahbi et al. (10) found high agreement both within and between the two timing methods with the coefficient of correlation at r=0.99 ($p<0.001$) and the Intraclass Correlation Coefficient (ICC) at 0.98. The Rope Climbing Test began with the participant sitting on his buttocks with the rope between his legs, both hands placed on the rope without exceeding the starting mark situated at 1 m above the ground. The climbing was performed without skipping (without momentum), without the use of any gloves and without using lower limbs (i.e. the legs and feet were not allowed to touch the rope to help climbing) (see Figure 1).

The Execution Time (ET) was defined as the time between the starting signal and the noise of the slap of the hand hitting the finish mark. Both visual and auditory cues were used by the assessor to ensure that substantial and solid contact was made with the finish mark. The two best attempts out of the 3 trials were kept for analysis. The removal of the worst trial was an attempt to ensure that a single poor performance did not substantially affect the analysis. To provide greater reproducibility of measurement, only one assessor measured the ET (no inter-assessor differences in reaction and movement time). The measurement of ET allowed for the estimation of the Absolute (APO) and Relative Power Output (RPO), which were calculated using the following equations:

$$\mathbf{APO(W)} = \frac{\mathbf{Body\ mass\ (kg)} \times \mathbf{9.81} \times \mathbf{5\ m}}{\mathbf{ET(sec)}} = \frac{\mathbf{49.05 \times Body\ mass\ (kg)}}{\mathbf{ET(sec)}}$$

$$\mathbf{RPO(W \cdot kg^{-1})} = \frac{\mathbf{APO(W)}}{\mathbf{Bodymass(kg)}} = \mathbf{49.}\frac{\mathbf{05}}{\mathbf{ET(sec)}}$$

Statistical Analyses

Data analyses were performed using SPSS version 18.0 for Windows. Means and standard deviations (SD) were calculated after verifying the normality of distributions using Kolmogorov-Smirnov statistics. Estimates of effect size, mean differences, and 95% confidence intervals (CIs) protected against type 2 errors. Independent t-tests were used to evaluate the equality of means for Commandos and Intervention-Brigade soldiers' RCT ET, APO, RPO and RPE. The external responsiveness of the RCT was analyzed using the receiver operator characteristics (ROC) curve (16). The latter analysis determines the sensitivity and specificity of a tool to classify individuals according to a fixed criterion (9). The relative intra-session reliability (i.e. the degree to which individuals maintain their position in a sample over repeated measurements (2)) of the ET, APO and RPO were determined by calculating the ICC (ICC[3,1]), and the absolute intra-session reliability (i.e. the degree to which repeated measurements vary for individuals (2)) was expressed in terms of SEM and Coefficients of Variation (CV). Heteroscedasticity was examined. Significance for all the statistical tests was accepted at $p \leq 0.05$ a priori.

Results

Discriminant ability of RCT

Separate group (Commandos and Intervention-Brigade) anthropometric characteristics and RCT indices (ET, APO, RPO and RPE) are displayed in Tables 1 and 2, respectively. Residual data for anthropometric characteristics and RCT indices were normally distributed ($p = 0.052–0.200$). Independent sample t-test revealed no difference between groups for age (years) (t=-0.188, p=0.852, dz=0.06[trivial]); body mass (kg) (t=-1.018, p=0.315, dz=0.32[moderate]); height (cm) (t=-0.043, p=0.966, dz=0.01[trivial]); body mass index (BMI: $kg \cdot m^{-2}$) (t=-0.921, p=0.363, dz=0.29 [moderate]); or RPE (t=-0.269, p=0.789, dz=0.09 [trivial]). However, ET (t=-5.918, dz=1.87[large]), APO (t=4.255, dz=1.33[large]) and RPO

(t=5.122, dz=1.52[large]) were significantly higher for Commandos compared to Intervention-Brigade group ($p<0.001$). A ROC analysis was performed between Commandos and Intervention-Brigade soldiers: very good discriminant ability was found for RCT. The areas under the ROC curves of ET, APO and RPO were of 0.91, 0.85 and 0.90, respectively (95% confidence intervals [CI]: 0.77 to 0.98, 0.70 to 0.94 and 0.77 to 0.98, respectively; $p<0.001$) (Figure 2).

Absolute and relative intra-session reliability of RCT

Absolute and relative intra-session reliability indices are expressed in Table 3. Dependent t-tests evaluating the equality of means showed no significant test-retest bias for ET (sec) (t=-0.62, p=0.55, dz=0.13 [trivial]); APO (W) (t=0.78, p=0.44, dz=0.17 [trivial]); RPO ($W \cdot kg^{-1}$) (t=0.85, p=0.41, dz=0.21 [moderate]) and RPE (t=0.17, p=0.87, dz=0.05 [trivial]). The ET, APO and RPO showed a high degree of relative reliability between the test-retest trials (ICC[3,1] ranging from 0.96 to 0.97). The SEM of ET, APO and RPO were 0.23 sec, 3.25 W and 0.05 $W \cdot kg^{-1}$, respectively. The CVs of ET, APO and RPO were all under 10%. Heteroscedasticity coefficients for ET, APO, RPO and RPE were all small and statistically non-significant (r=0.01 [p=0.96], r=0.40 [p=0.08], r=0.43 [p=0.06] and r=-0.31 [p=0.16], respectively).

Discussion

The inability to provide reliable and valid strength and power testing to identify and progressively train athletes, military, police and emergency medical personnel could result in serious personal injury or injuries to individuals who are dependent upon them. Hence, this study assessed the discriminant ability of RCT to distinguish soldiers' specialty level as well as to establish the absolute and relative intra-session reliability. The main findings of this

study showed that RCT is a highly reliable intra-session and sensitive tool to differentiate upper limb power between two groups of soldiers of different operational capacity levels.

One of the main characteristics of the RCT is its discriminant ability. A significant difference was found between ET, APO and RPO performance of Commandos and Intervention-Brigade groups. Impellizzeri and Marcora (16) suggested that the ROC curve is an appropriate tool to validate the discriminant ability (and responsiveness) of physiological and performance tests and can determine test sensitivity and specificity to classify individuals according to a fixed criterion (5). The area under the ROC curve (AUC) was interpreted as the probability to correctly discriminate Commandos from Intervention-Brigade soldiers using the RCT protocol. An AUC value of 0.5 is interpreted as no discriminatory ability and 1.0 as complete discriminatory ability (9) with an AUC>0.70 considered to indicate good discriminative ability (10, 21). In the present study, the AUC values were: 0.91, 0.85 and 0.90 for ET, APO and RPO, respectively (10). The test scores (ET, APO and RPO) able to differentiate between groups of soldiers of different operational capacity levels were ≥20.14 sec, ≥185.64 W and ≥2.43 $W \cdot kg^{-1}$, respectively. ROC consists of a plot of "true positive rate" (sensitivity) vs. "false positive rate" (1-specificity) for each of several possible cut-off points in changing the score (10). These cut-off values give a true positive rate of 73.7%, for ET, APO and RPO; and a false positive rate of 95.2%, 85.7% and 95.2% for ET, APO and RPO, respectively (figure2). Therefore, RCT has excellent discriminant ability if its purpose is to differentiate between Commandos and other specialty soldiers. These results are complementary with those of Dhahbi et al. (11) who included the same group of Commandos that participated in this study. They assessed the internal responsiveness (i.e. the ability to detect longitudinal changes) of the RCT by calculating the likelihood that differences in RCT outcomes were substantial (i.e., the Smallest Worthwhile Change larger than the SEM) (19). This was the case for all ET, APO and RPO (11), indicating that such data have a good

potential to detect real changes in the power output of upper limbs. As well as, in the Dhabbi et al. (11) study, the Minimal Detectable Change (2) was used to find the score threshold corresponding to a true change in the performance. They showed that 1.62 sec, 31.45 W and 0.41 $W \cdot kg^{-1}$ or more of ET, APO and RPO, respectively were necessary to be 95% confident that a true change has occurred in Commandos soldiers.

Although the typical methods for assessing upper limb power have been pull-ups (25), push-ups (7), bench press power test (6) and medicine ball put tests (24), few studies provide data on their discriminant ability. For example, there was no data reported for the discriminant ability of 15 sec pull-ups (23), 15 sec push-ups (23), bench press (6, 26), medicine ball puts or throws (6, 27) or single arm seated shop puts (23). Using a laboratory Wingate test rather than a field test, Koutedakis et al. (17) had excellent discrimination as they could classify 91.8% of their subjects. A good level of discrimination was reported for bench press repeated power test (13) and a medicine ball throw test (8) with youth basketball players and children of 5-7 years respectively. A rock climbing specific test (arm jump board test) could discriminate between novice and experienced climbers (18). Hence, the excellent discriminant ability scores, that substantiated by a powerful statistical tool as the ROC curve, for a simple field test such as the RCT should be considered an important tool for professionals and practitioners in the field. Moreover, no significant difference was found between groups in RPE responses. This strongly suggests that both groups of participants did comparable efforts, most probably maximal efforts. The absence of a significant anthropometric and age differences between groups ensures these variables did affect performance.

The variability between trials may be considered as “intrinsic variation”, as it provides a basic indication of the variation independent from other sources of error. Intra-session reliability of RCT performance is critically important to ensure that observed differences between testing trials are not due to systematic bias, such as a learning effect, fatigue, or

241 random error due to possible biological or mechanical variations. This variability is usually
242 caused by the emotional state of the subject between the trials and his level of adaptation with
243 the measuring system (22). The results demonstrated a very high level of relative reliability of
244 RCT. Other upper limb field tests such as 15 sec pull-ups (0.99) (23), 15 sec push-ups (0.96)
245 (23), bench press (0.92-0.98) (6, 13), medicine ball throws (0.92-0.97) (6, 8) and rock
246 climbing specific test (0.98) (18) have also reported excellent ICC reliability scores. However,
247 one of the weaknesses of ICC as a measure of relative repeatability is that it is affected by
248 sample heterogeneity (29). Therefore an examination of the SEM, which provides an absolute
249 index of reliability in conjunction with the ICC is needed to confirm the ICC's results (20).
250 The SEM is not affected by inter-subject variability (29) and provides an estimate of
251 measurement error. In addition, if data are homoscedastic, which is the case in the current
252 study ($r=0.01$, $r=0.40$ and $r=0.43$; $p>0.05$ for ET, APO and RPO, respectively), SEM index is
253 more appropriate than CV to establish the absolute reliability (2, 29). In this study, SEMs
254 were low for all parameters, under 5%, thereby confirming the excellent absolute intra-session
255 reliability of RCT. Similarly, Dhahbi et al.(11) found an excellent inter-session reliability of
256 RCT: for ET, APO and RPO; ICC[3,1] values were all higher than 0.90, SEM% all under 5%
257 and CV% all under 10%. Thus, it can be concluded that the RCT has excellent intra- and
258 inter-session reliability.

259 In conclusion, the RCT has excellent relative and absolute intra-session reliability and
260 a good discriminant ability to detect difference in power performance of upper limbs between
261 two groups of soldiers of different operational capacity levels. A score of 20.14 sec, ≥185.64
262 W and ≥2.43 $W \cdot kg^{-1}$ for ET, APO and RPO respectively were the cut-off points
263 discriminating elite Commandos from less trained Intervention-Brigade soldiers. While these
264 scores were reliable and discriminant in the current study population, these cut-off points may
265 not be the same in other populations and that this should be examined in future studies.

Practical Applications:

The RCT is a fitness-specific field test designed to evaluate the power of the upper limbs performance of Commando soldiers. The results showed that this test has a good absolute and relative reliability and successfully discriminates soldiers by operational level. Considering that (i) reliability and (ii) discriminant ability of a test are two important aspects, RCT can therefore be recommended for similar professionals such as the military, police, fire fighters and emergency medical personnel.

Acknowledgments

The authors are grateful to all the participants for their enthusiasm and commitment to the completion of this study. They also wish to express their sincere gratitude to Colonel Imed Mekki, director of the Tunisian Commandos' School, for advice and cooperation in this study.

References

1. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, and Stein RE. Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 11: 193-205, 2002.
2. Atkinson G and Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports medicine* 26: 217-238, 1998.
3. Bishop PA, Crowder TA, Fielitz LR, Lindsay TR, and Woods AK. Impact of body weight on performance of a weight-supported motor fitness test in men. *Military medicine* 173: 1108-1114, 2008.

4. Carlson MJ and Jaenen SP. The development of a preselection physical fitness training program for Canadian Special Operations Regiment applicants. *Journal of strength and conditioning research / National Strength & Conditioning Association* 26 Suppl 2: S2-14, 2012.

5. Chaabene H, Hachana Y, Franchini E, Mkaouer B, Montassar M, and Chamari K. Reliability and construct validity of the karate-specific aerobic test. *Journal of strength and conditioning research / National Strength & Conditioning Association* 26: 3454-3460, 2012.

6. Clemons JM, Campbell B, and Jeansonne C. Validity and reliability of a new test of upper body power. *Journal of strength and conditioning research / National Strength & Conditioning Association* 24: 1559-1565, 2010.

7. Cuddy JS, Slivka DR, Hailes WS, and Ruby BC. Factors of trainability and predictability associated with military physical fitness test success. *Journal of strength and conditioning research / National Strength & Conditioning Association* 25: 3486-3494, 2011.

8. Davis KL, Kang M, Boswell BB, DuBose KD, Altman SR, and Binkley HM. Validity and reliability of the medicine ball throw for kindergarten children. *Journal of strength and conditioning research / National Strength & Conditioning Association* 22: 1958-1963, 2008.

9. de Vet HC, Bouter LM, Bezemer PD, and Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *International journal of technology assessment in health care* 17: 479-487, 2001.

10. Deyo RA and Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *Journal of chronic diseases* 39: 897-906, 1986.

11. Dhahbi W, Chaouachi A, Padulo J, Behm DG, and Chamari K. Five Meters Rope-climbing Test: Commando-specific Power Test of the Upper-Limbs. *International journal of sports physiology and performance* 10: 509-515, 2015.

12. Drust B, Waterhouse J, Atkinson G, Edwards B, and Reilly T. Circadian rhythms in sports performance--an update. *Chronobiology international* 22: 21-44, 2005.

13. Gonzalo-Skok O, Tous-Fajardo J, Arjol-Serrano JL, and Mendez-Villanueva A. Determinants, reliability, and usefulness of a bench press repeated power ability test in young basketball players. *Journal of strength and conditioning research / National Strength & Conditioning Association* 28: 126-133, 2014.

14. Haddad M, Chaouachi A, Castagna C, Hued O, Wong DP, Tabbena M, Behm DG, and Chamari K. Validity and psychometric evaluation of the French version of RPE scale in young fit males when monitoring training loads. *Sci Sports* 28: e29–e35, 2013.

15. Harman EA, Gutekunst DJ, Frykman PN, Sharp MA, Nindl BC, Alemany JA, and Mello RP. Prediction of simulated battlefield physical performance from field-expedient tests. *Military medicine* 173: 36-41, 2008.

16. Impellizzeri FM and Marcora SM. Test validation in sport physiology: lessons learned from clinimetrics. *International journal of sports physiology and performance* 4: 269-277, 2009.

17. Koutedakis Y and Sharp NC. A modified Wingate test for measuring anaerobic work of the upper body in junior rowers. *British journal of sports medicine* 20: 153-156, 1986.

18. Laffaye G, Collin JM, Levernier G, and Padulo J. Upper-limb power test in rock-climbing. *International journal of sports medicine* 35: 670-675, 2014.

19. Liow DK and Hopkins WG. Velocity specificity of weight training for kayak sprint performance. *Medicine and science in sports and exercise* 35: 1232-1237, 2003.

20. Looney MA. When is the intraclass correlation coefficient misleading? *Meas Phys Educ Exerc Sci* 4: 73–78, 2000.

21. Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Semmer NK, Jacobshagen N, Dvorak J, and Boos N. Outcome assessment in low back pain: how low can you go? *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 14: 1014-1026, 2005.

22. McGinley JL, Baker R, Wolfe R, and Morris ME. The reliability of three-dimensional kinematic gait measurements: a systematic review. *Gait & posture* 29: 360-369, 2009.

23. Negrete RJ, Hanney WJ, Kolber MJ, Davies GJ, Ansley MK, McBride AB, and Overstreet AL. Reliability, minimal detectable change, and normative values for tests of upper extremity function and power. *Journal of strength and conditioning research / National Strength & Conditioning Association* 24: 3318-3325, 2010.

24. Sharp MA, Knapik JJ, Walker LA, Burrell L, Frykman PN, Darakjy SS, Lester ME, and Marin RE. Physical fitness and body composition after a 9-month deployment to Afghanistan. *Medicine and science in sports and exercise* 40: 1687-1692, 2008.

25. Sporis G, Harasin D, Bok D, Matika D, and Vuleta D. Effects of a training program for special operations battalion on soldiers' fitness characteristics. *Journal of strength and conditioning research / National Strength & Conditioning Association* 26: 2872-2882, 2012.

26. Stock MS, Beck TW, DeFreitas JM, and Dillon MA. Test-retest reliability of barbell velocity during the free-weight bench-press exercise. *Journal of strength and conditioning research / National Strength & Conditioning Association* 25: 171-177, 2011.

27. Stockbrugger BA and Haennel RG. Validity and reliability of a medicine ball explosive power test. *Journal of strength and conditioning research / National Strength & Conditioning Association* 15: 431-438, 2001.

28. Vossen JF, Kramer, J. E., Burke, D. G., Vossen, D. P. Comparison of Dynamic Push-Up Training and Plyometric Push-Up Training on Upper-Body Power and Strength. *The Journal of Strength & Conditioning Research* 14: 248-253, 2000.

29. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of strength and conditioning research / National Strength & Conditioning Association* 19: 231-240, 2005.

Table 1: Descriptive data and comparison of the characteristics of Commandos and Intervention Brigade groups

Variables	Commandos (n=21)	Intervention Brigade (n=19)	*p*-values	Effect size
Age (years)	24.09±1.81	24.21±2.07	0. 852	0.06
BM (kg)	74.90±5.08	76.42±4.25	0. 315	0.32
Height (cm)	179.52±3.98	179.58±4.15	0. 966	0.01
BMI ($kg \cdot m^{-2}$)	23.26±1.65	23.72±1.51	0. 363	0.29

BM = Body Mass; BMI = Body Mass Index; *Significant differencebetween groups ($p < 0.001$); Values are given as mean ± SD.

Table 2: Descriptive data and comparison of the RCT indices of Commandos and Intervention Brigade groups

Variables		Commandos (n=21)	Intervention Brigade (n=19)	*p*-values	Effect size
ET	(sec)	15.55±3.48	22.11±3.53*	<0.001	1.87
APO	(W)	251.13±73.55	174.59±35.41*	<0.001	1.33
RPO	($W \cdot kg^{-1}$)	3.33±0.85	2.28±0.39*	<0.001	1.52
RPE		8.07±1.04	8.16±0.99	0. 789	0.09

RCT = 5 m Rope Climbing Test; ET = Execution Time; APO = Absolute Power Output; RPO = Relative Power Output; RPE= Rating of Perceived Exertion; *Significant difference between groups ($p < 0.001$); Values are given as mean ± SD.

Table 3: Relative and Absolute intra-session reliability indices of the RCT (n=21)

Variables		Mean±SD		*p*-values	$ICC_{3,1}$ (95%)¶	SEM (%)†	CV†
		Trial 1	Trial 2				
ET	(sec)	15.55±3.48	15.41±3.65	0. 545	0.96 (0.89-0.98)	0.23 (1.47%)	6.92
APO	(W)	251.13±73.55	254.28±74.40	0.443	0.97 (0.93-0.99)	3.25 (1.29%)	7.30
RPO	($W \cdot kg^{-1}$)	3.33±0.85	3.38±0.88	0.407	0.96 (0.91-0.98)	0.05 (1.46%)	7.30
RPE		8.07±1.04	8.10±0.94	0.871	0.78 (0.53-0.90)	0.31 (3.84%)	8.18

RCT = 5 m Rope Climbing Test; ET = Execution Time; APO = Absolute Power Output; RPO = Relative Power Output; RPE = Rating of Perceived Exertion; $ICC_{3,1}$ = Intra-class Correlation Coefficient model 3,1; SEM = Standard Error of Measurement; CV = Coefficient of Variation; *Significant difference between trials ($p<0.05$).

†Absolute intra-session reliability index.

¶Relative intra-session l reliability index.

Figure 1: The 5 m Rope Climbing Test from the starting to the finishing position. A = starting position, B = execution and C = finishing position.

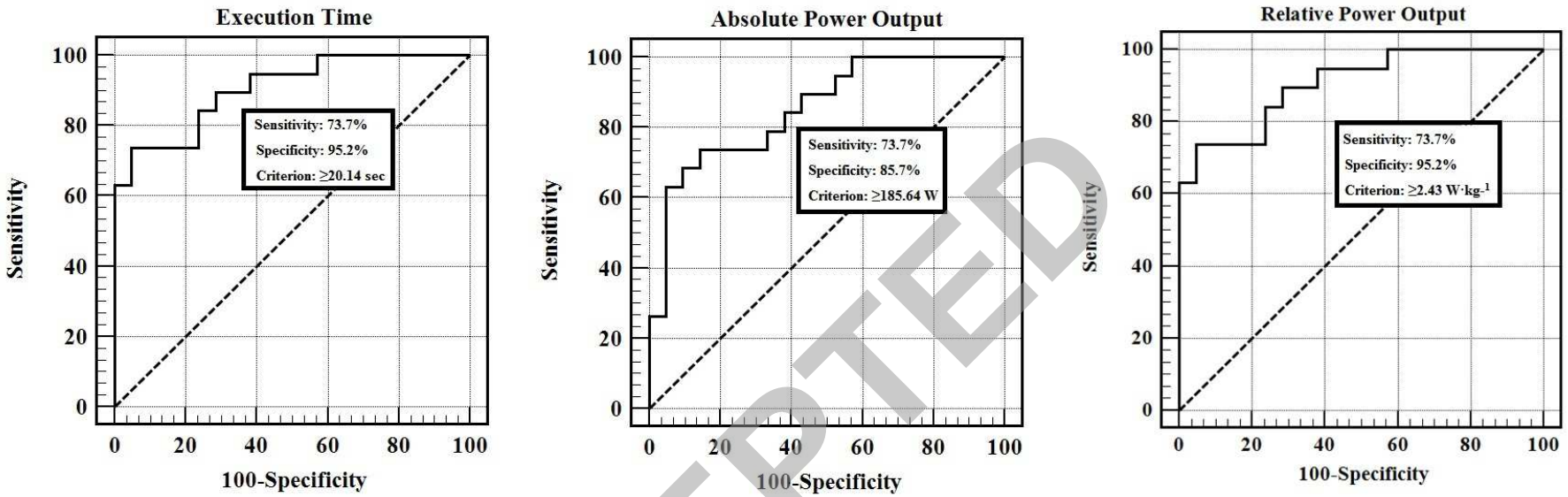

Figure 2: Receiver operating characteristics (ROC) curves for the Execution Time, Absolute Power Output and Relative Power Output between Commandos and Intervention Brigade soldiers.