Research note

Applicability of a Change of Direction Ability Field-Test in Soccer Assistant Referees

Carlo Castagna[1,2], Franco Impellizzeri[3], Mario Bizzini[3,4] , Matthew Weston[5] and Vincenzo Manzi[1]

[1]School of Sport and Exercise Sciences, Team-Sports Department, Faculty of Medicine and Surgery, University of Rome Tor Vergata, Rome Italy;

[2]Italian Soccer Referees Association, Rome, Italy;

[3]Neuromuscular Research Laboratory, Schulthess Clinic, Zurich, Switzerland;

[4]FIFA-Medical Assessment and Research Centre (F-MARC), Zurich, Switzerland.

[5]Department of Sport and Exercise Science, Teesside University, UK;

Running Title: Change of direction ability in Soccer Assistant Referees

Corresponding Author:

Carlo Castagna Dr.

Via Sparapani 30, 60131, Ancona, Italy
e-mail: castagnac@libero.it
phone +39 071 28 66 532
fax +39 071 28 66 478

Research note

Applicability of a Change of Direction Ability Field-Test in Soccer Assistant Referees

Running Title: Change of Direction Ability in Soccer Assistant Referees

ABSTRACT

The aim of this study was to examine the applicability of a test for change of direction ability (10-8-8-10 test, involving line and sideward sprinting, 36m) in elite level Soccer Assistant Referees (AR). One hundred AR of the first-second and third Italian Championships (AR_{A-B} and AR_C, n=50, respectively) performed the 10-8-8-10 on three separate occasions. Twenty AR authorities scored test relevance (1 to 5 scale, from trivial to very large) for logical validity using a questionnaire. Construct validity was examined comparing AR_{A-B} and AR_C for 10-8-8-10 performance. Short-term reliability was assessed testing a random selection of ARs (n=64) over three separate occasions every other day. Performance in the 10-8-8-10 test was assumed as total coverage time using telemetric photo-cells. Results showed that the 10-8-8-10 test was perceived as possessing from large (n=4/20) to very-large (n=16/20) relevance to AR physical match-performance. No significant performance difference was found between competitive levels (p=0.57). Area under the curve (AUC=0.49; p=0.87) showed no significant sensitivity of 10-8-8-10 in detecting competitive-level difference. The ICC (n=64) and TEM (test 2 vs 3) values were 0.90 (p<0.0001) and 0.18 s, respectively. This study showed that the 10-8-8-10 test posses logical validity, good reliability and it is independent to competitive level. As such, this original investigation represents the first step in the identification and assessment of a valid and reliable AR COD test. Given the strength of our findings, governing bodies should look to integrate the 10-8-8-10 test into the fitness test protocols devised for AR's, with scores equal or higher than 9.67 being considered as a starting point for the empirical validation of minimum selection criteria for elite-level AR's.

Keywords: Refereeing; Fitness Test; Sprint; Agility; Association Football

INTRODUCTION

Each and every match is controlled by a field referee (FR) who has full authority to enforce the Laws of the Game in connection with the match to which he/she has been appointed (13). The FR controls and regulates players and coaches behaviour during the match in cooperation with two Assistant Referees (AR) whose main duty is to take a position to check the offside line, which is a priority in any case (13). Differently from the FR, who is free to move inside the soccer pitch, the ARs control the game moving parallel to the side-line of the half pitch.. As a result the assistant referee is expected to face the pitch whilst running. Preferentially sideways movement should be used when judging offside as it gives the assistant referee a better line of vision. Like the FRs also the ARs must keep up with play whatever the tempo is (10). However compared to FRs the ARs experience lower physiological and physical demands during the game (14, 21, 22, 26). Nevertheless fatigue has been reported to be experienced by top-class AR and documented as decrements in ability to repeat sprint after the match (21).

During the game (i.e. ~90 min) the AR covers approximately 6-8 km of which 1.2 with sideways running (i.e. 16%) (21). In top-class ARs 110 high-intensity running activities were reported with a mean duration of 2 s (21). Interestingly moving along the 50 m portion of the side-line of competence a top-level AR changes activity every 5 s totalling 1053 activities (21). Given the high frequency rate and variety in activity changes experienced by an elite-level AR during the game, the ability to perform sudden changes of direction, shifting from forward sprinting to sideways movement, assume face validity.

The possession of a well developed physical fitness is considered as a necessary prerequisite for optimal positioning and work-rate output in AR (21). Indeed distance from the offside line as determined by the second last defender position, has been related to the sprint ability of the AR (21). Furthermore, association between individual aerobic fitness and the ability to repeat sprint with distance covered at high intensity during the game was reported (21).

The refereeing international governing body (UEFA, European Union of Association Football, and FIFA (International Union of Association Football) acknowledged the need for physical fitness assessment of the elite level AR. This was accomplished requiring a number of field tests aiming to evaluate AR repeated sprint ability (6x40m sprints) and intermittent high intensity endurance (30-40") before top-competitions (36). These tests although significantly stressing the physiological characteristics of the AR possess low face validity as per distance considered and exercise mode used (36). Indeed they consider only line running, neglecting sideways running considered as a specific of AR performance (21). Furthermore they do not evaluate the ability to perform sudden changes of direction, repeated changes and also the specific distances experienced by AR during the most demanding phases of the game (21). Unfortunately to the best of this study authors knowledge no study has been published in scientific journals that proposed and examined the validity (i.e. face and construct validity) and reliability of a test for change of direction ability (CODA) for ARs.

Therefore the aim of this study was to examine the applicability of a new field test (i.e. 10-8-8-10) aiming to assess CODA of elite level AR involving forward and sideways shuttle-running. In the pursue of this aim construct validity and reliability of the 10-8-8-10 test were assessed. Prior to the experimental procedures face validity was established using time motion analysis and expert judgment procedures.

METHODS

Experimental Approach to the problem

The elite level AR during his/her match activity perform approximately 1053 activity changes, most of them in response to match development (i.e. off-side line changes) (21). As such the AR must react to the visual cues (i.e. next last defender position) associating a proper COD that usually involves shuffling and shuttle running in combination at maximal effort (21, 22, 26). Research has reported that the closer the position of the AR with respect to the off-side line the better the visual perspective will be to make the correct decision (i.e. avoid flag or no flag errors) (14, 15). Therefore, the CODA should be considered as prerequisite to proper positioning and assessed and trained according to game demands in ARs (14, 15, 27). The CODA should be considered as a functional prerequisite of Agility that involves with COD also the aptitude to react to developing action (27). In this the 10-8-8-10 test is the first systematic proposal for a simple field test to assess the CODA in elite level AR. As no gold standard has been established for Agility or CODA the 10-8-8-10 validity was assessed as compliance to expert judgement of match relevance and qualitative match analysis as constructs at this stage (20).

The 10-8-8-10 test was developed qualitatively examining the activity pattern of Italian elite level ARs during official championships. The development of the activity pattern considered in the 10-8-8-10 test was undertaken according to the distance and time reported for the mean sprint coverage in scientific reports published and adjusted for test feasibility to provide consistent data collection (21, 22, 26). Specifically forward running was assumed as 10 m according to the average time scored by AR in a preliminary pilot study and usual discrete sprint time revealed by time motion analysis (i.e. 2s) (21). The sideward shuttle-running pattern used in the 10-8-8-10 test was considered as consequence of the AR's most frequent action of choice aiming to remain in-line with the second last defender. The shuttle running (i.e. involving start and end test on the same line) nature of 10-8-8-10 test was chosen for relevance to match demands and for test feasibility (i.e. to ease timing).

The 10-8-8-10 test involves the AR sprinting forward for 10m and rapidly performing two 8 m sideways shuttle-runs having as reference the 10m end line (i.e. change of direction line). Once completed the two 8-m sideway shuttle runs the AR sprints back to the start-line to complete the 10-8-8-10 test (10+8+8+10m=36m). During the 10-8-8-10 test the AR must have his/her forward foot stepping on the reference line (i.e. at 10m at the 8m lines) at each direction change. Only trials that were performed in compliance with the explained test procedures were considered for analysis. Prior to the 10-8-8-10 test all participants were familiarised with the test rules and allowed for practise trials.

Construct validity was assessed comparing 50 AR randomly chosen from the relative competitive cohorts (Serie A-B and Can-Pro, n=90 and 200, respectively). Construct validity has been used to assess the sensitivity of a test to discriminate between players of different competitive levels (18, 25). With this criterion-based validity differences, once detected, are difficult to be explained unless an accurate control over possible intervening variables is exerted (29). In order to operate control over the possible training effect on test performance in this population of AR, the experimental procedures were intentionally undertaken after the preparation phase of the competitive season. This with the aim to account for possible difference in training background affecting physical performance (i.e. 10-8-8-10 test) across the considered competitive levels. During this preparation phase all the ARs involved undertook a common supervised training program aiming to develop the fitness determinant of match performance (21, 26). Compliance with the training procedures was assured with the help of certified assistant referees strength and conditioning coaches appointed by the Italian Soccer Referees Association (AIA). Training load was assumed as time devoted for the development of the following fitness categories: endurance, sprint/agility, repeated sprint ability and flexibility (11, 12).

Before the commencement of the procedures used in this study face validity was gained by a questionnaire in which experts (i.e. current and former top class AR, n=20) rated the 10-8-8-10 using a 1 to 5 scale for specificity (i.e. test relevance for AR performance: 1=trivial, 2=small, 3=medium, 4=large, 5=very large). The overall ratings (n=20) for the 10-8-8-10 was "large" to "very large" (n= 4 and 16 out of 20, respectively).

Subjects

Participants were 100 AR belonging to the Italian Soccer Referees Association (AIA) who were appointed for officiating in the 2009-2010 Serie A-B (n=50, AR_{A-B}, age 37±2.9 years, Height 1.78±6.9 cm, body mass 74±6.9 kg) and Lega Pro (n=50, $AR_{LP,}$ age 34±2.0 years, Height 1.77±4.9 cm, body mass 75±4.9 kg) Italian professional championships. The ARs had at least 2 years (range 2-8 years) of experience at their respective competitive level (i.e. Serie A-B and Lega Pro, respectively) and performed at least three training sessions a week for the development of specific fitness. All ARs were tested during the training camp held at the end of the precompetitive season (i.e. July-August). To avoid undue stress on the AR in the days preceding the implementation of the 10-8-8-10 test, training loads were intentionally reduced and familiarisation sessions were considered. The ARs were advised to maintain a regular diet during the day before testing (i.e. 60%, 25% and 15% of carbohydrates, fat and protein, respectively) and to refrain from smoking and caffeinated drinks during the two hours preceding testing. To avoid hypo-hydration ARs were allowed to drink fluids "ad libitum". Written informed consent was obtained from all the participants after familiarization and explanation of the benefit and risks involved in this study procedures. All participants were informed that they were free to withdraw from the study at any time without penalty. The Institutional Research Board (Settore Tecnico AIA, Modulo per la preparazione Atletica) provided clearance for the procedures before the commencement of this study. All procedures were carried out in accordance to the Declaration of Helsinki of the World Medical Association as regards the conduct of clinical research.

Procedures

All the test procedures were performed at the same hours of the day (i.e. 9-11 am) with wind absence and similar environmental conditions (i.e. 23-26C°, 50-60% humidity). The 10-8-8-10 test was performed at the end of a standard warm-up consisting each time in 15 min slow jogging (i.e. 2-3 of Börg's CR 10 scale) followed by static stretching (5 min) and agility and sprint practise (8 min) (1, 8). The AR performed three trials of the 10-8-8-10 test interspersed by 2 min of passive recovery in between. All tests were performed on the same synthetic turf usually used for fitness-training sessions. The 10-8-8-10 test performance was assumed as total time and assessed using a telemetric photocells system (Polifemo Kit Racetime2, Microgate, Bolzano, Italy). To avoid undue switch-on of the timing system ARs had to position the front foot immediately before a line set 0.50 m from the photocell beam. The photocell beam was positioned at 0.5 m height and 1.5m apart. All the AR performed the 10-8-8-10 test with a self administered start and maximum performance was induced through strong verbal encouragements by the test leader (i.e. first author) during all the test duration.

The 10-8-8-10 test reliability was established having ARs (n=64) performing the test under the above declared conditions on three different occasions separated by a recovery (a low training load session) day. Best and average 10-8-8-10 performances were used for calculation.

Statistical Analysis

The results are expressed as means ± standard deviations (SD) and 95% confidence intervals (95% CI). Normality assumption was verified using the Shapiro-Wilk W-test. Student's t-tests (unpaired design) were used to determine any significant difference between the two competitive levels (i.e. AR_{A-B} and AR_{LP}). Homogeneity of variance was tested with the Bartlett test. The effect size (ES) was calculated to assess meaningfulness of differences (9). Effect sizes of above 0.8, between 0.8 and 0.5, between 0.5 and 0.2 and lower than 0.2 were considered as large, moderate, small, and trivial respectively. The ICC (Intra Class Correlation Coefficient) was used to assess relative reliability of the 10-8-8-10 test. Absolute reliability was assessed calculating the Typical Error of Measurement (TEM) according to Hopkins (16). Sensitivity of 10-8-8-10 test was evaluated using Receiver Operating Characteristic curve (ROC) statistics. To allow ROC calculations ARs were dichotomized according to their competitive level (i.e. AR_{AB} and AR_{LP}). Normative data were reported as inter-quartile range (32). The the smallest worthwhile change was assumed as 0.2xSD according to Hopkins at al. (17). Significance was set at 5% ($p \leq 0.05$).

RESULTS

The AR_{A-B} were significantly older than AR_{LP} ($p<0.0001$; 95CI% -4.10 ÷ -2.10; ES= 1.25).There were no significant differences between the AR_{A-B} and AR_{LP} groups for the percentage of time spent training for endurance, sprint and agility and repeated sprint ability during the pre-test preparation phase (i.e. 6 weeks). The percentage of time spent training for endurance, sprint and agility and repeated sprint ability was 68±5.1 and 67±6.8% (p=0.34, 95%CI -0.78 ÷1.78**;** ES=0.17), 15±4.6 and 14±5.8% (p=0.22, 95%CI -0.81 ÷1.81; ES=0.19) and 10±4.3 and 9±6.8% (p=0.44, 95%CI -0.6 ÷1.6; ES=0.18), of the total training time for the AR_{A-B} and AR_{LP} groups, respectively. Percentage of time devoted to flexibility training was significantly higher in AR_{LP} than in AR_{A-B} (9±1.8 and 7±1.8 %, p=0.04, 95%CI 1.3÷2.7, ES=0.32).

The mean best 10-8-8-10 test time for the AR_{A-B} and AR_{LP} was 9.61±0.45 and 9.66±0.41s respectively (p=0.57; 95%CI -0.13 ÷0.23; ES= 0.11; n=100). Performance in the 10-8-8-10 test assumed as mean of all trials was 9.81±0.41 and 9.78±0.41s for AR_{A-B} and $AR_{LP,}$ respectively (p=0.72; 95%CI -0.21 ÷ 0.14; ES=0.07; n=100).

Mean, median and mode of 10-8-8-10 test pooled data (i.e. AR_{A-B} plus AR_{LP}) were 9.74±0.34s (95%CI 9.67 ÷ 9.81), 9.79 (95%CI 9.67 ÷9.84) and 9.64s for the mean of the three trials, respectively. The corresponding values for the 10-8-8-10 test best performance pooled data were 9.60±0.36 (95%CI 9.52 ÷ 9.67), 9.63 (95%CI 9.54 ÷9.70), and 9.48s.

Receiver Operating Characteristic analysis showed that 10-8-8-10 performance was insensitive in detecting competitive level differences in this population of ARs as revealed by area under the curve size (AUC=0.49, p=0.87; 95%CI 0.38 ÷ 0.60; fig. 1).

The ICC for the 10-8-8-10 test was 0.90 (n=64, p<0.0001, 95%CI 0.84÷0.93). The across trials TEM was of 0.21 (95%CI 0.18÷0.24) and 0.18 s (95%CI 0.16÷0.21) for trial 1 vs. 2 and trial 2 vs. 3 respectively.

The minimum worthwhile change for the 10-8-8-10 test (pooled data, n=100) was of 0.07s using either the mean and the best of the pooled (i.e. AR_{A-B} and AR_C) values.

The inter-quartile range values for the 10-8-8-10 scores were 9.57s (95%CI 9.39 ÷9.64) and 9.94s (95%CI 9.85÷10.09) for the 25 and 75% percentiles respectively (fig.2).

----Insert figure 1 and 2 about here----

DISCUSSION

This is the first study to examine the applicability (i.e. construct validity and reliability) of a CODA test in elite level ARs. The results of this study showed that the 10-8-8-10 test did not discriminate between AR of different competitive level and to possess good short-term reliability. Furthermore the 10-8-8-10 test showed to have high face validity considered as perceived test relevance (i.e. from large to very large relevance) by a panel of experienced AR authorities. Consequently, this original investigation represents the first step in the identification and assessment of a valid and reliable AR COD test.

Construct validity is usually assumed as a prerequisite of test applicability in sport science (29). Indeed with this aspect of criterion-based validity authors usually examine the sensitivity of a test in discriminating between athletic populations belonging to different competitive levels assumed as construct (19). This criterion-based validity feature is assumed as a viable strategy to performance prediction and talent selection and identification in sport science (18, 20, 24). Furthermore, it is suggested to provide valuable information for training prescription when dealing with youth and elite level athletes (30, 31). Despite the interest of this aspect of criterion-based validity, detection of construct validity does not provide evidence of cause and effect relationships "*per se*" (20, 29). Indeed construct validity when assuming qualitative criteria for conceptualization (i.e. elite vs. non-elite level) may not provide a clear reflection of the supporting cause of status difference. Specifically, the competitive status may be the consequence of difference in skill level, training background and or genetic factors (31). Consequently, to gain meaningful information from this particular aspect of criterion-based validity the items determining the construct definition must be carefully set (20, 29). In this research design we operated control over construct (i.e. competitive level) supporting variables testing two population of AR (i.e. $AR_{A\text{-}B}$ and AR_{LP}) after the completion of a similar training program. This was undertaken with the very intention to experimentally account for difference in training background due to competitive level membership. Indeed in Italy AR reach the semi-professional status only when affiliation with the higher national competitive level is awarded (Serie A and B) and this impacts considerably on the training load in term of frequency of weekly training sessions, volume and intensity (34, 35). As a result differences in test performance may be the consequence of superior training effort and reasonably considered as test bias in talent identification conduct.

To account for training bias for construct validity of 10-8-8-10 test we assumed control over short-term (i.e. 6 weeks) pre-experimentation training background for the AR groups considered in this study. Analysis of training loads assumed as percentage of time devoted to physiological abilities considered as determinants of elite level AR performance (i.e. Endurance, Sprint and Agility, Repeated Sprint Ability and Flexibility), showed no significant differences between all the considered physical abilities but flexibility (11, 12, 21, 26). This provided support to the experimental purpose of implementing similar training stimuli to the two competitive level different AR populations.

Results showed that no significant difference between means was detected for 10-8-8-10 test performance across competitive levels (i.e. AR_{A-B} vs AR_{LP}). Absence of competitive level sensitivity of the 10-8-8-10 test was further supported by ROC analysis that showed balance between sensitivity and specificity as represented by the area under the curve value (AUC=0.49). In soccer refereeing competitive level progression parallels age progression as experience is considered as an attribute of skill development in elite level soccer refereeing (7, 28). Consequently impairment in physical performance as a result of ageing may be expected in soccer refereeing (3, 6). A number of studies have showed that age related impairment in physical performance are expected in the neuromuscular performance domain as soccer officials get older in the progression of their competitive careers (3, 7, 28). In this study despite a significant older mean age of AR_{A-B} no significant difference were detected in CODA mainly considered as a neuromuscular physical ability (2, 27). Furthermore no significant association between age and 10-8-8-10 test performance was detected when examining pooled data (n=100, r=0.03, p=0.74, 95%CI -0.16 ÷ 0.23). This contrasts with previous studies that addressed the effect of age on speed and explosive strength in FR (3, 5). Although comparison between studies are difficult to be performed as different designs were used, it could be speculated that control over training loads may have had an effect over the lack of difference in neuromuscular performance (i.e. 10-8-8-10 test) between competitive levels in this study (3, 5).

This evidence suggests that the 10-8-8-10 test may be successfully used to detect CODA of AR across the competitive levels and considered as a reflection of individual abilities. This is of particular interest in soccer refereeing as fitness tests have shown to be affected by competitive-level associated variables such as training background and age (3, 4, 6, 33). This study's findings suggest that the 10-8-8-10 test performance may be considered as a trainable physical ability and that competitive-level independent normative may be considered in AR assessment. However, training studies examining the effect of CODA training load nature, volume and intensity on 10-8-8-10 performance are warranted.

The 10-8-8-10 showed good absolute and relative reliability as reported by ICC and TEM calculations respectively. Specifically 10-8-8-10 performance showed short-term variation (i.e. every other day testing) in the order of approximately 0.19 s (n=66). This absolute variation in performance was higher (i.e. ~0.19 vs 0.07s) in size than the minimum worthwhile change estimated as fraction of pooled data standard deviation (i.e. 0.20xSD). It could be suggested that to detect meaningful information as per size of the effect of the intervention change, higher proportion of standard deviation should be used. Indeed considering for the estimation of the minimum worthwhile change instead of small effect (i.e. ES=0.20) medium to large effect size (i.e. ES from 0.5 on) the test noise can be exceeded providing likelihood for more plausible changes.

Despite the encouraging results of this study the implementation of our findings as an evaluation criteria of CODA of elite level ARs should be corroborated by further studies providing evidence for sensitivity as consequence of intervention (i.e. training studied) and seasonal variations of the 10-8-8-10 test (18, 20). Normative values should be extended by examining 10-8-8-10 test performance in AR of different age (i.e. younger), competitive level (i.e. junior level) and gender (i.e. male vs female ARs). Studies addressing a possible relationship between 10-8-8-10 test performance and the AR ability to keep last defender off-side line are warranted. Nonetheless, despite the relevance of COD for AR match performance, its assessment is overlooked in the current FIFA AR fitness test protocol. With this in mind, this original investigation represents the first step in the identification and assessment of a valid and reliable AR COD test. Given the strength of our findings, governing bodies should look to integrate the 10-8-8-10 test into the fitness test protocols devised for AR's, with scores equal or higher than 9.67 being considered as a starting point for the empirical validation of minimum selection criteria for elite-level AR's

PRACTICAL APPLICATIONS

The CODA is considered as an important prerequisite to promote specific agility in ARs (14, 15, 21, 26). In this regard the 10-8-8-10 possessing logical validity and good reliability may be successfully used to track changes in CODA performance in trained ARs at the beginning the competitive season. Furthermore the 10-8-8-10 test considering sideways shuttle running may be used to train AR in developing the physical prerequisite useful to track offside situation during the match.

Due to short-term variation in 10-8-8-10 performance (~0.18s) difference in performance equal or higher than 0.18 s can be considered as meaningful. Performance changes equal or higher than 0.07 s and lower than 0.18 s should be considered with caution. The same consideration should be used when considering inter-individual differences for ranking difference. Given the population addressed in this study the 10-8-8-10 test values here reported may be considered as reference normative values for highly competitive level ARs before the beginning of the competitive season (23). For AR populations similar to this study 10-8-8-10 test scores equal or lower than 9.67-s should be considered as of interest. This information may be considered of importance for national and international soccer referees governing bodies aiming to implement and assess agility training programs for elite ARs.

Acknowledgements

Authors wish to acknowledge the devotion of all the ARs that volunteered to this study. This study did not benefit of any financial support. The authors has no conflict of interest with the study.

REFERENCES

1. BÖRG, G. *Borg's perceived exertion and pain scales*. Champaign IL: Human Kinetics, 1998.
2. BRUGHELLI, M., J. CRONIN, G. LEVIN, and A. CHAOUACHI. Understanding change of direction ability in sport: a review of resistance training studies. *Sports Med*. 38(12):1045-63. 2008.
3. CASAJUS, J.A. and C. CASTAGNA. Aerobic fitness and field test performance in elite Spanish soccer referees of different ages. *J Sci Med Sport*. 10(6):382-9. 2007.
4. CASTAGNA, C., G. ABT, and S. D'OTTAVIO. Competitive-level differences in Yo-Yo intermittent recovery and twelve minute run test performance in soccer referees. *J Strength Cond Res.* 19(4):805-9. 2005.
5. CASTAGNA, C., G. ABT, S. D'OTTAVIO, and M. WESTON. Age-related effects on fitness performance in elite-level soccer referees. *J Strength Cond Res.* 19(4):785-90. 2005.
6. CASTAGNA, C., G. ABT, S. D'OTTAVIO, and M. WESTON. Age-related effects on fitness performance in elite-level soccer referees. *J Strength Cond Res*. 19(4):785-90. 2005.
7. CASTAGNA, C., G. ABT, and S. D'OTTAVIO. Physiological aspects of soccer refereeing performance and training. *Sports Med*. 37(7):625-46. 2007.
8. CHAOUACHI, A., C. CASTAGNA, M. CHTARA, M. BRUGHELLI, O. TURKI, O. GALY, K. CHAMARI, and D.G. BEHM. Effect of Warm-Ups Involving Static or Dynamic Stretching on Agility, Sprinting, and Jumping Performance in Trained Individuals. *J Strength Cond Res*. 2009.

9. COHEN, J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

10. D'OTTAVIO, S. and C. CASTAGNA. Analysis of match activities in elite soccer referees during actual match play. *J Strength Cond Res.* 15(2):167-171. 2001.

11. ESTEVE-LANAO, J., A.F. SAN JUAN, C.P. EARNEST, C. FOSTER, and A. LUCIA. How do endurance runners actually train? Relationship with competition performance. *Med Sci Sports Exerc.* 37(3):496-504. 2005.

12. ESTEVE-LANAO, J., C. FOSTER, S. SEILER, and A. LUCIA. Impact of training intensity distribution on performance in endurance athletes. *J Strength Cond Res*. 21(3):943-949. 2007.

13. FÉDERATION INTERNATIONALE DE FOOTBALL ASSOCIATION. *Laws of the Game 2005*. Zurich, Switzerland: Féderation Internationale de Football Association, 2005.

14. HELSEN, W. and J.B. BULTYNCK. Physical and perceptual-cognitive demands of top-class refereeing in association football. *Journal of Sports Sciences*. 22:179–189. 2004.

15. HELSEN, W., B. GILIS, and M. WESTON. Errors in judging "offside" in association football: test of the optical error versus the perceptual flash-lag hypothesis. *J Sports Sci.* 24(5):521-528. 2006.

16. HOPKINS, W.G. Measures of reliability in sports medicine and science. *Sports Med.* 30:1-15. 2000.

17. HOPKINS, W.G., S.W. MARSHALL, A.M. BATTERHAM, and J. HANIN. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*. 41(1):3-13. 2009.

18. IMPELLIZZERI, F.M., E. RAMPININI, C. CASTAGNA, D. BISHOP, D. FERRARI BRAVO, A. TIBAUDI, and U. WISLOFF. Validity of a repeated-sprint test for football. *Int J Sports Med*. 29(11):899-905. 2008.

19. IMPELLIZZERI, F.M., E. RAMPININI, C. CASTAGNA, D. BISHOP, D. FERRARI BRAVO, A. TIBAUDI, and U. WISLØFF. Validity of a repeated-sprint test for football. *Int J Sports Med*. 29(11):899-905. 2008.

20. IMPELLIZZERI, F.M. and S.M. MARCORA. Test validation in sport physiology: lessons learned from clinimetrics. *Int J Sports Physiol Perform*. 4(2):269-77. 2009.

21. KRUSTRUP, P., M. MOHR, and J. BANGSBO. Activity profile and physiological demands of top-class soccer assistant refereeing in relation to training status. *J Sports Sci*. 20:861-871. 2002.

22. MALLO, J., E. NAVARRO, J.M. GARCIA-ARANDA, B. GILIS, and W. HELSEN. Analysis of the kinematical demands imposed on top-class assistant referees during competitive soccer matches. *J Strength Cond Res*. 22(1):235-42. 2008.

23. MONTGOMERY, P.G., D.B. PYNE, W.G. HOPKINS, and C.L. MINAHAN. Seasonal progression and variability of repeat-effort line-drill performance in elite junior basketball players. *J Sports Sci*. 26(5):543-50. 2008.

24. MUJIKA, I., J. SANTISTEBAN, F.M. IMPELLIZZERI, and C. CASTAGNA. Fitness determinants of success in men's and women's football. *J Sports Sci*. 27(2):107-14. 2009.

25. RAMPININI, E., A. SASSI, A. AZZALIN, C. CASTAGNA, P. MENASPA, D. CARLOMAGNO, and F.M. IMPELLIZZERI. Physiological determinants of Yo-Yo intermittent recovery tests in male soccer players. *Eur J Appl Physiol*. 2009.

26. REILLY, T. and W. GREGSON. Special populations: the referee and assistant referee. *J Sports Sci*. 24(7):795-801. 2006.

27. SHEPPARD, J.M. and W.B. YOUNG. Agility literature review : classifications, training and testing. *J. Sports Sci.* 24(9):919-932. 2006

28. STØLEN, T., K. CHAMARI, C. CASTAGNA, and U. WISLØFF. Physiology of Soccer: An Update. *Sports Med.* 35(6):501-536. 2005.

29. THOMAS, J.R., J.K. NELSON, and J. SILVERMAN. *Research methods in physical activity*. 5th ed. Champaign, IL.: Human Kinetics, 2005.

30. VAEYENS, R., R.M. MALINA, M. JANSSENS, B. VAN RENTERGHEM, J. BOURGOIS, J. VRIJENS, and R.M. PHILIPPAERTS. A multidisciplinary selection model for youth soccer: the Ghent Youth Soccer Project. *Br J Sports Med*. 40(11):928-934. 2006.

31. VAEYENS, R., M. LENOIR, A.M. WILLIAMS, and R.M. PHILIPPAERTS. Talent identification and development programmes in sport : current models and future directions. *Sports Med*. 38(9):703-14. 2008.

32. VINCENT, W.J. *Statistics in Kinesiology*. Champaign, IL: Human Kinetics, 1995.

33. WESTON, M., W. HELSEN, C. MACMAHON, and D. KIRKENDALL. The Impact of Specific High-Intensity Training Sessions on Football Referees' Fitness Levels. *American Journal of Sport Medicine*. 32(1 suppl.):54-61s. 2004.

34. WESTON, M., S. BIRD, W. HELSEN, A. NEVILL, and C. CASTAGNA. The effect of match standard and referee experience on the objective and subjective match workload of English Premier League referees. *J Sci Med Sport*. 9(3):256-62. 2006.

35. WESTON, M., C. CASTAGNA, F.M. IMPELLIZZERI, E. RAMPININI, and S. BREIVIK. Ageing and physical match performance in English Premier League soccer referees. *J Sci Med Sport*. 2008.

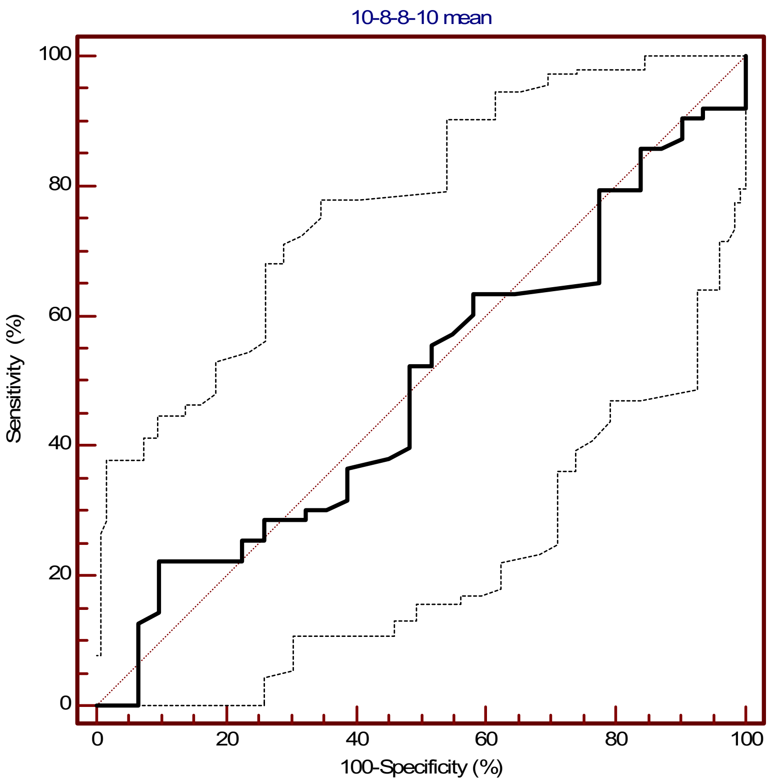
36. WESTON, M., C. CASTAGNA, W. HELSEN, and F. IMPELLIZZERI. Relationships among field-test measures and physical match performance in elite-standard soccer referees. *J Sports Sci*:1-8. 2009.

FIGURE LEGENDS

Fig.1. Receiver operating characteristics (ROC) plot for the 10-8-8-10 test performance assuming competitive level as dichotomous variable (i.e. Serie A-B vs Lega Pro; n=50 respectively). Area under the curve (AUC=0.49, p=0.87; 95%CI 0.38 ÷ 0.60). Black line= ROC curve.

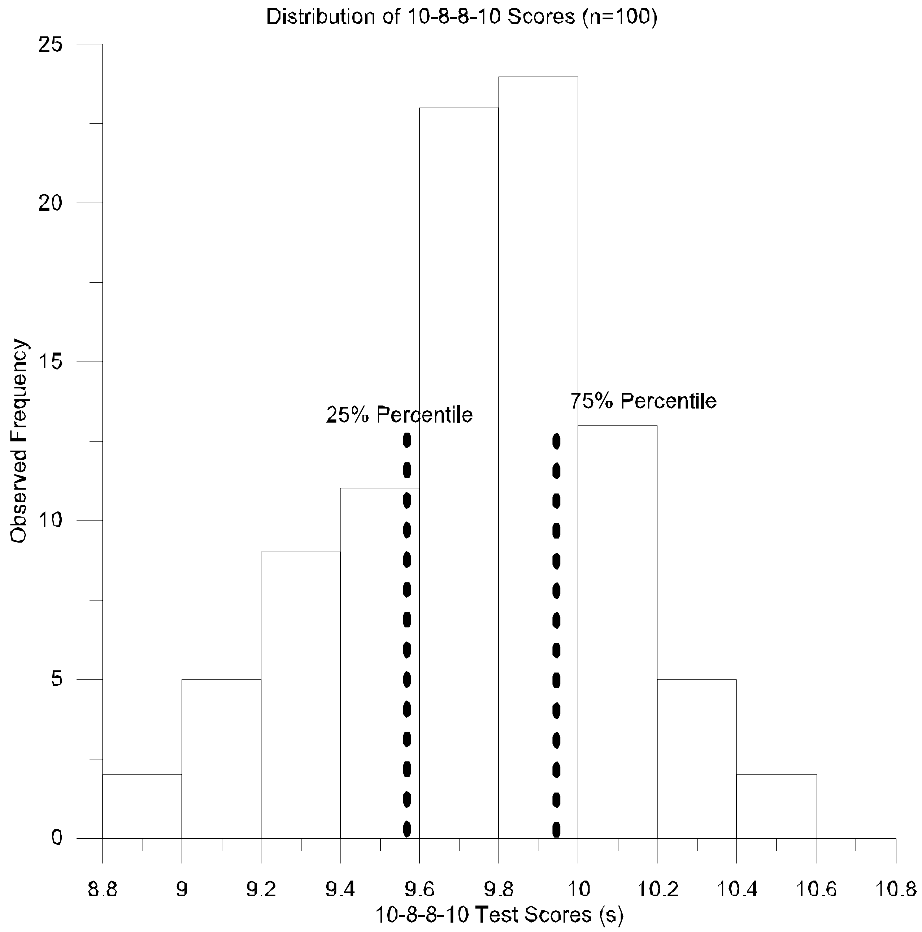
Fig.2. Distribution of 10-8-8-10 scores. Vertical black dotted lines represents the limit of the 25 and 75% percentiles**.**

Fig.1



10-8-8-10 mean
100
80
60
40
20
0
Sensitivity (%)
0
20
40
60
80
100
100-Specificity (%)

Fig.2



Distribution of 10-8-8-10 Scores (n=100)
25
20
15
10
5
0
Observed Frequency
25% Percentile
75% Percentile
8.8
9
9.2
9.4
9.6
9.8
10
10.2
10.4
10.6
10.8
10-8-8-10 Test Scores (s)