Evaluation of a goalkeeper-specific adaptation to the Yo-Yo intermittent recovery test level 1: Reliability and variability

By: Alex M. Ehlert, John R. Cone, Laurie Wideman, and Allan H. Goldfarb

Ehlert AM, Cone JR, Wideman L & Goldfarb AH. 2019. Evaluation of a goalkeeper-specific adaptation to the Yo-Yo intermittent recovery test level 1: Reliability and variability. *Journal of Strength and Conditioning Research* 33(3): 819-824. PMID: 30363034

Made available courtesy of Lippincott, Williams, and Wilkins:
http://dx.doi.org/10.1519/JSC.0000000000002869

Abstract:

The Yo-Yo Intermittent Recovery Test Level 1 (YYIR1) has been shown to be a reliable test with strong correlations to physical match performance in field soccer players. However, the YYIR1 has less goalkeeper (GK) specificity. Therefore, the purpose of this study was to evaluate the test-retest reliability of a goalkeeper-specific adaptation of the YYIR1 (YYIR1-GK). Sixteen National Collegiate Athletic Association (NCAA) collegiate GKs (8 men and 8 women) performed the YYIR1-GK test twice (>4 and <7 days apart) to determine its reliability and variability. Subjects were tested at the same time of day and in a controlled indoor environment. Heart rate using polar monitors and rating of perceived exertion were obtained at the end of each stage of the YYIR1-GK. Test-retest reliability for each test was assessed by Pearson correlations, intraclass correlation coefficient (ICC), and coefficient of variation (CV). The YYIR1-GK was shown to have a strong test-retest reliability and low variability for male ($r = 0.981$, ICC = 0.980, CV = 5.82%) and female ($r = 0.969$, ICC = 0.956, CV = 9.60%) NCAA GKs, respectively. Male GKs performed significantly more stages and therein covered a greater distance in the YYIR1-GK than the female GKs ($p = 0.05$). This study suggests that the YYIR1-GK is a consistent assessment of intermittent fitness with high test-retest reliability and low variability in male and female NCAA GKs. It is suggested that larger numbers of GKs of various skill levels be evaluated in the future.

Keywords: fitness testing | soccer | goalkeeper

Article:

Introduction

Fitness is an important aspect for soccer success (2) and can be measured objectively using valid and reliable tests (6,16). Soccer matches are composed of intermittent and dynamic movements (runs, jumps, and kicks) with changes in speed, and direction at high to maximal intensity

interspersed with lower intensity periods (3). Thus, physical performance in soccer is highly reliant on intermittent exercise (8). The Yo-Yo Intermittent Recovery Test Level 1 (YYIR1) is a fitness test that can conveniently and accurately measure intermittent exercise fitness in team sport athletes (2). Previous studies have shown that the YYIR1 is a reliable and valid field test for assessing fitness in field soccer players (8,12). The ecological validity of the YYIR1 is documented through its association with variables of physical match performance such as high-intensity running, sprinting, and distance covered during match play in men (8,12) and women (9). YYIR1 results have also shown significant correlations with maximal oxygen uptake (VO_{2max}), a physiologic measure considered important for soccer players (8,15). In addition, YYIR1 performance has been shown to differentiate between low- and high-level players (2,8) as well as detecting seasonal and training-induced changes (9).

In contrast to field players in soccer, the physical demands and actions of the goalkeeper (GK) are vastly different and require wholly unique movement patterns (11) and are therefore typically excluded from studies on the YYIR1 (8,9,11). Goalkeepers cover less total distance than field players (6) but perform a variety of explosive actions in a smaller space (17). Although the GK rarely performs sprints of appreciable distance (1), high- to maximal-intensity actions are a considerable contributor to success, with professional GKs reported to average 92 high-intensity actions covering a mean distance of about 4 m per action during match play (11). These actions were typically in lateral, forward, and backward directions to obtain optimal positioning (11). Elite GKs have also been shown to perform a variety of defensive and technical actions throughout a match such as diving, catching, deflections (i.e., parrying and boxing), ball control (i.e., trapping), throwing, and kicking (14). In-between these explosive high- to maximal-intensity actions are extended periods of low-intensity or passive recovery. The result is that the GK position is unique in soccer because of its locomotion demand and highly intermittent nature with relatively exclusive focus on high- to maximal-intensity actions and rarely linear runs of appreciable distance. Despite the research and previous observations of practitioners, there has been limited development of a reliable test to evaluate goalkeeping fitness.

Considered holistically, GK performance is complex and relies on integrated tactical, technical, perceptual, and physical factors making it difficult to reliably evaluate in an objective manner (7,11). In efforts to address this, previous GK-specific tests have isolated aspects of the position such as technical ability and reactive agility (7,13). Despite recent reports that GKs perform considerable high-intensity activities during match play and emphases on the overall indefatigability of the GK during the course of the game, there has been little focus on GK-specific fitness. Because fitness capacity can be measured objectively, field tests targeting the ability to perform sport-specific activity patterns in a convenient and accurate manner are beneficial to coaches, strength and conditioning practitioners, and medical professionals looking to quantify readiness to train and compete (9). While the YYIR1 could be adapted to the GK position (9,11), to the best of our knowledge, an adaptation has yet to be developed and appropriately assessed. The development of any adaptation should involve actions common to match play to be relevant to GK-specific fitness. Therefore, a GK adaptation of the YYIR1 should focus on high-intensity lateral, forward, and backward displacements of 4 m (11). In addition, because of the logistical challenges of testing in the team environment, the ideal GK fitness test should have the ability to evaluate the positional demands conveniently alongside the rest of the team. The adaptation in this study addresses the differing demands of GKs by using

position-specific, high-intensity actions instead of the linear runs in the standard YYIR1 (6,11), providing overall longer rest intervals in comparison with the field player (e.g., normal) test and progressively challenging the GK's recoverability through decreasing rest interval duration to effectively assess recoverability from repeated high-intensity actions. As it is also synchronized to the same YYIR1 audio script used for testing outfield players in the normal team environment, the GK-specific test allows for testing during the same session as the complete team.

For these reasons, the primary aim of this study was to evaluate the reliability and variability of a GK-specific adaptation of the YYIR1 (YYIR1-GK) in both male and female National Collegiate Athletic Association (NCAA) GKs. The secondary aim was to compare performance on the YYIR1-GK between male and female GKs. It was hypothesized that the YYIR1-GK would exhibit strong reliability and low variability in both male and female GKs. It was also hypothesized that the men would outperform women as previously noted with the standard YYIR1.

Methods

Experimental Approach to the Problem

To investigate the hypotheses, this study used a repeated-measure, within-subject design. Subjects completed 2 sessions at the same time of day (≥5 and ≤7 days apart). Sessions were standardized in structure and performed in a controlled inside environment.

All GKs were asked to refrain from moderate- to high-intensity activity and alcohol consumption for 24 hours before each session. Although dietary and hydration information were not directly measured, participants were asked to prepare for each session as they would for a competitive soccer match. At the end of the first session, subjects were asked to replicate the nutritional and hydration preparation for the second session. Testing environment conditions, water intake, subject heart rate (HR), and rating of perceived exertion (RPE) were monitored during each session and analyzed for consistency between visits. All testing occurred during the offseason training period for NCAA GKs.

Table 1. Subject characteristics for male and female goalkeepers.*†

Characteristics	**Male (*n* = 8)**	**Female (*n* = 8)**
Age (yrs)	20.00 ± 1.51	19.50 ± 1.20
Height (cm)	185.34 ± 3.58	172.84 ± 6.21‡
Body mass (kg)	85.10 ± 8.25	79.95 ± 9.15
Body fat (%)	15.05 ± 6.04	27.24 ± 6.50‡

†Values are mean values ± *SD*s.
‡Significant group differences at $p < 0.05$.

Subjects

Sixteen highly trained NCAA GKs (8 men and 8 women, aged 18–23 years) participated in this study. The characteristics of the subjects are in Table 1. The University Institutional Review Board (IRB) for the protection of human subjects at the University of North Carolina at Greensboro approved all methods, and potential subjects signed consent forms after being

informed of all potential risks. Subjects were free of lower extremity injuries, had no history of cardiovascular disorders or events, were not currently hypertensive as measured at the beginning of their first visit, and female subjects were not currently pregnant.

Procedures

Each visit included (a) collection of baseline measures, (b) a standardized dynamic warm-up, and (c) YYIR1-GK in that order. Baseline measurements included resting blood pressure, height (stadiometer), body mass (Seca, Hamburg, Germany), body fat percentage through Bod Pod analysis (Cosmed, Rome, Italy), and resting HR using HR Polar V800 monitor (Polar, Kempele, Finland). The HR monitor was worn for the entire session. The procedures for the visit and the Borg RPE scale were explained (4) before beginning each session.

The YYIR1-GK used the standard YYIR1 audio file with distance and locomotion adapted to be GK-specific. Specifically, the YYIR1-GK consisted of 16-m bouts performed at progressive intensity synchronized to the YYIR1 audio. Each 16-m bout was split into four 4-m actions performed in sequence: (a) side shuffle, (b) side shuffle, (c) backward run, and (d) forward run (Figure 1). The direction of the first side shuffle was selected by the GK and kept consistent for each visit. Each 16-m action (i.e., complete sequence) needed to be completed before the first audible beep from the audio file, which indicates the turning point for the normal YYIR1 test, coinciding with a 20-m run. After each 16-m action, subjects rested passively on the start line through both the second beep (corresponding to the 20-m return run) and the 10-second rest interval (corresponding to the 5-m walk) dictated by the YYIR1 audio. The subjects started each 16-m sequence standing still, facing 90° from the direction of their movement, outside foot on the line ready to perform the initial side shuffle. Each action was monitored to ensure the foot of the subject touched the line before moving in the opposite direction and into the next 4-m action. If a subject missed 2 bouts (i.e., 16-m runs), the test was terminated, and cumulative distance covered was calculated. Intensity for each aspect of the session was monitored through beat-by-beat HR using HR monitor and RPE. The same researcher conducted all testing, and data were compared between each visit to measure variability.

Statistical Analyses

To assess the reliability and variability of the YYIR1-GK, Pearson product-moment correlation, intraclass correlation coefficient (ICC), and coefficient of variation (CV) compared cumulative distance covered on both visits for men and women, respectively. Mixed-factor analyses of variance (ANOVAs) were used to analyze within-group results between visits 1 and 2 and between-group comparisons by sex for YYIR1-GK distance covered, peak HR, and peak RPE. All other variables were compared for consistency between visits using paired t-tests. All statistical analyses were performed through SPSS (v25), and statistical significance was set at $p \leq 0.05$.

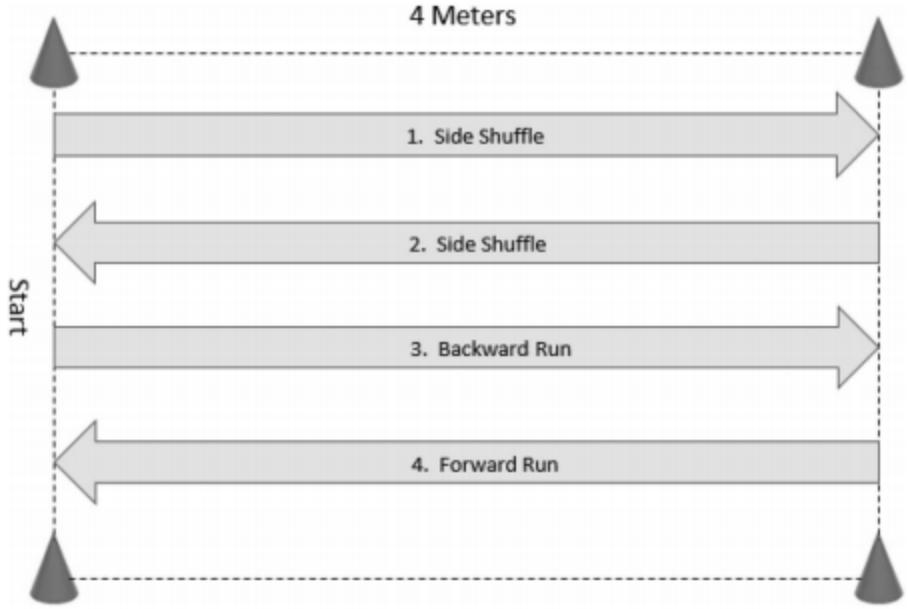

Figure 1. YYIR1-GK diagram of movement for 1 bout. Distance between cones is 4 m in each direction. Direction of movement (right or left) for the first side shuffle was self-selected by the goalkeeper and kept consistent.

Results

Paired t-tests found no significant differences in the testing environment between visits for laboratory temperature (21.33 ± 0.71 vs. 21.67 ± 0.50° C, $p = 0.282$) or humidity (41.11 ± 11.26 vs. 39.67 ± 9.91%, $p = 0.619$). There were significant differences between visits in the gym temperature (21.13 ± 0.89 vs. 21.63 ± 0.89° C, $p = 0.041$) and humidity (35.56 ± 14.32 vs. 43.19 ± 13.67%, $p = 0.004$).

Table 2 summarizes the results of reliability and variability analysis for YYIR1-GK cumulative distance covered for male and female GKs. Male GKs showed strong test-retest reliability as measured by Pearson product-moment correlation coefficient ($r = 0.981$, $n = 8$, $p < 0.001$), ICC ($ICC_{3,1} = 0.980$), and low variability (CV = 5.82%). Distance covered by female GKs also had strong Pearson correlations ($r = 0.969$, $n = 8$, $p < 0.001$) and ICC ($ICC_{3,1} = 0.956$) with slightly higher variability (CV = 9.60%) (Figure 2).

Table 2. Reliability and variability of YYIR1-GK performance.*

Group	Pearson correlation	ICC	CV (%)
Male	0.981†	0.980†	5.82
Female	0.969†	0.956†	9.60

*ICC = intraclass correlation coefficient; CV = coefficient of variation.
†$p < 0.05$.

Table 3 displays the mean distance covered, peak HR, and peak RPE for the YYIR1-GK by visit and group. Two-way mixed ANOVAs were used to analyze within-group effects by visit number and between-group comparisons between male and female groups in terms of YYIR1-GK distance covered, peak HR, and peak RPE.

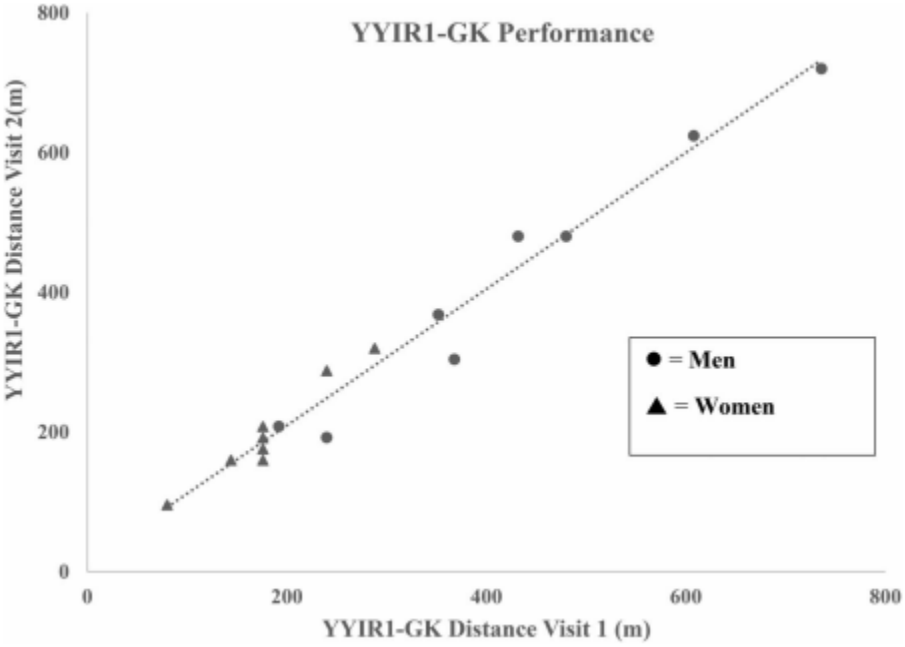

Figure 2. Test-retest correlations for YYIR1-GK distance covered.

Table 3. Mean distance covered successfully, peak HR, and peak RPE during the YYIR1-GK by group and visit number.*†

Variable	Group	Visit 1	Visit 2
YYIR1-GK Distance (m)	Men	426.00 ± 181.41	422.00 ± 189.69
	Women	182.00 ± 61.63‡	200.00 ± 72.57‡
Peak HR ($b \cdot min^{-1}$)	Men	190.00 ± 7.74	188.67 ± 8.31
	Women	188.75 ± 6.90	187.63 ± 8.43
Peak RPE	Men	16.38 ± 2.07	15.71 ± 1.98
	Women	15.50 ± 2.39	15.13 ± 2.95

*HR = heart rate; RPE = rating of perceived exertion.
†Values are mean ± *SD*. *SD* is displayed in parentheses.
‡Significantly different comparing groups at $p < 0.05$.

The 2-way mixed ANOVA found no significant main effect for distance covered by visit on the YYIR1-GK, $F(1,14) = 0.886$, $p = 0.362$, as well as no significant interaction effect with visit and sex, $F(1,14) = 2.189$, $p = 0.161$. However, a significant main effect for sex in regards to YYIR1-GK distance was revealed, $F(1,14) = 11.270$, $p = 0.005$, $\eta_p^2 = 0.446$, where men covered significantly more distance than women.

There was no significant main effect for peak HR by visit on the YYIR1-GK, $F(1,14) = 2.652$, $p = 0.129$, as well as no significant interaction effect with visit and sex, $F(1,14) = 0.382$, $p = 0.548$. Between-group analysis found no significant main effect for sex, $F(1,14) = 0.170$, $p = 0.688$, $\eta_p^2 = 0.014$.

Peak RPE showed no significant main effect for visit number $F(1,14) = 2.109$, $p = 0.170$, with no interaction effect for visit and sex, $F(1,14) = 0.205$, $p = 0.658$. Between-group analysis found no significant main effect for sex, $F(1,14) = 0.398$, $p = 0.539$, $\eta_p^2 = 0.030$.

Additional analyses indicated that height and distance covered by female GKs on the YYIR1-GK had large significant correlations by visit (visit 1: $r = 0.880$, $p = 0.004$; visit 2: $r = 0.892$, $p = 0.003$). However, height and distance covered on the YYIR1-GK in male GKs revealed low correlations that were not significant (visit 1: $r = 0.030$, $p = 0.944$; visit 2: $r = 0.112$, $p = 0.792$). None of the other variables measured had significant correlations with YYIR1-GK distance covered in male or female GKs.

Discussion

Fitness testing in sport serves multiple purposes, ranging from tracking physical development, to monitoring training program efficacy and evaluating physical training and match readiness (5). There has been a substantial amount of attention in soccer toward quantifying physical performance in field players, but little attempt to develop tests that evaluate the physical fitness of the GK position. The YYIR1 is frequently used to evaluate physical fitness in soccer players, but uses movement patterns vastly different than that of GKs. To address the need for a test more relevant to the GK position, an adaptation to the YYIR1 (YYIR1-GK) was designed using high-intensity dynamic movement patterns that have been observed in match play and longer rest periods to relate to GK-specific activity (11). In this study, NCAA GKs performed the YYIR1-GK twice within 5–7 days to evaluate the reliability and variability of performance outcomes.

The primary finding was that YYIR1-GK performance had strong reliability and low variability as measured by Pearson product-moment correlations, ICC, and CV, in both men and women. Pearson correlation and ICC from this study exceeded those previously reported for the standard YYIR1 (2,15). The CV for the YYIR1-GK was comparable in the male GKs with what has been reported for elite male field players on the standard YYIR1 (8) and lower than those reported with lower skill levels (15). Performance by female GKs also had strong reliability and slightly higher CV, although still comparable with that of nonelite field players on the YYIR1 (15). Overall, the reliability and variability of the YYIR1-GK in this study were comparable to previous findings with the standard YYIR1 (2,8,15), despite the additional complexity of multiple actions performed sequentially during each bout.

The male GKs performed significantly more distance on the YYIR1-GK than the female GKs. The gap between male and female performance was even greater on the YYIR1-GK than that reported in the standard YYIR1 (10). Interestingly, the substantially lower cumulative distance covered by female GKs was not related to higher peak HR or RPE.

There were several limitations to this study. First, the sample size was relatively small for group comparisons. This prevented our ability to compare performance by NCAA division level. In addition, because of the challenge of recruiting participants who met the inclusion criteria for this study for multiple visits, all testing was performed during the NCAA soccer offseason. There could be a difference in performance if GKs were tested during the competitive season.

There are a variety of future directions for study of the YYIR1-GK. There should be an effort to incorporate a larger sample size of GKs and different skill levels. As YYIR1 distance covered has been shown to differentiate between skill level and age (2,10), studies using the YYIR1-GK

should investigate these same comparisons. The standard YYIR1 has also shown sufficient sensitivity to track fitness changes throughout the course of a season (8). YYIR1-GK distance should therefore be tested during different times of the year to determine whether the test is sensitive enough to detect seasonal changes in fitness. There was a significant correlation between YYIR1-GK distance and height in the female cohort. It is possible that height or leg length were factors that influenced the 4-m distance covered in the women. This could be related to stride length or leg strength that were not measured and could be investigated in future studies. By contrast, YYIR1-GK distance covered and height in the male cohort demonstrated a poor relationship and thus was not a factor to influence YYIR1-GK distance covered.

Finally, the validity of the YYIR1-GK should be considered through examination of test performance relative to physical performance of the GK during competition. Validity between the standard YYIR1 and field player physical performance was established through relationships with distance covered at high intensity and distance sprinted (8,9). In addition, the standard YYIR1 differentiated between players of various skill levels (2). It is unlikely that validity for the YYIR1-GK can be established in the same manner that the standard YYIR1 was with field players. While overall distance covered and distance at high and sprint intensities are reasonable measures of physical performance for field players, the same cannot be said for goalkeeping play. The tactical role and subsequent physical locomotion of the GK is in response to the flow of the match, with the GK's role linked more strongly to the needs of the game than their physical ability to contribute to the game. Specifically, this is highlighted by research showing that the largest number of high-intensity actions and physical output for a GK occur when they are required to make a play on the ball or react to an opposing attack (6,7,11,14). Therefore, unlike field players, the physical requirements of a GK are dependent on the match itself rather than the physical capacity to perform the actions. The result is that inherent and unpredictable challenges exist in validating the YYIR1-GK to physical performance during match play. We would suggest instead that future research examine the association between YYIR1-GK and GK performance by examining GKs at various levels of competition to determine whether differences in physical performance (YYIR1-GK performance) are associated with varying competition levels (e.g., GK success).

Practical Applications

This is the first study to adapt the YYIR1 for GKs and showed it was reliable and had low variability in NCAA male and female GKs tested. The GK adaptation offers several advantages for coaches. First, it is one of only a few tests in the literature that targets GK-specific actions, and the only study that focused on evaluation of the ability to meet intermittent activity demands of the GK position. Second, it uses the same audio file as the standard YYIR1 with adjustments in setup and rest periods. This would allow GKs to be evaluated for intermittent fitness at the same time as the rest of the team, while still targeting the unique activity requirements of their position. Although needing more study to establish validity and reliability in GKs of various levels of play, the YYIR1-GK could offer a feasible and reliable option for evaluating GK-specific fitness in a convenient manner. Ideally, the test would allow for fitness evaluation as part of a complete GK evaluation process alongside tests targeting other aspects of GK play such as tactical ability and reactive agility (7,13).

References

1. Aziz A, Mukherjee S, Chia M, Teh K. Validity of the running repeated sprint ability test among playing positions and level of competitiveness in trained soccer players. Int J Sports Med 29: 833–838, 2008.

2. Bangsbo J, Iaia FM, Krustrup P. The Yo-Yo Intermittent Recovery Test. Sport Med 38: 37–51, 2008.

3. Bangsbo J, Michalsik L. Assessment of the physiological capacity of elite soccer players. In: Science and Football IV. London, United Kingdom: Routledge, 2002. pp. 53–62.

4. Borg E, Kaijser L. A comparison between three rating scales for perceived exertion and two different work tests. Scand J Med Sci Sport 16: 57–69, 2006.

5. Carling C, Williams AM, Reilly T. Introduction to Soccer Match Analysis. In: Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance. London, United Kingdom; New York, NY: Routledge, 2005. pp. 1–15.

6. Di Salvo V, Benito V. Activity profile of elite goalkeepers during football match-play. J Sports Med Phys Fitness 48: 443–446, 2008.

7. Knoop M, Fernandez-Fernandez J, Ferrauti A. Evaluation of a specific reaction and action speed test for the soccer goalkeeper. J Strength Cond Res 27: 2141–2148, 2013.

8. Krustrup P, Mohr M, Amstrup T, Rysgaard T, Johansen J, Steensberg A, et al. The Yo-Yo Intermittent Recovery Test: Physiological response, reliability, and validity. Med Sci Sport Exerc 35: 697–705, 2003.

9. Krustrup P, Mohr M, Ellingsgaard H, Bangsbo J. Physical demands during an elite female soccer game: Importance of training status. Med Sci Sport Exerc 37: 1242–1248, 2005.

10. Mujika I, Santisteban J, Impellizzeri FM, Castagna C. Fitness determinants of success in men's and women's football. J Sports Sci 27: 107–114, 2009.

11. Padulo J, Haddad M, Ardigò L, Chamari K, Pizzolato F. High frequency performance analysis of professional soccer goalkeepers: A pilot study. J Sport Med Phys Fit 55: 557–562, 2015.

12. Rampinini E, Bishop D, Marcora S, Ferrari Bravo D, Sassi R, Impellizzeri F. Validity of simple field tests as indicators of match-related physical performance in top-level professional soccer players. Int J Sports Med 28: 228–235, 2007.

13. Rebelo-Gonçalves R, Figueiredo AJ, Coelho-E-Silva MJ, Tessitore A. Assessment of technical skills in young soccer goalkeepers: Reliability and validity of two goalkeeper-specific tests. J Sports Sci Med 15: 516–523, 2016.

14. Sainz De Baranda P, Ortega E, Palao JM. Analysis of goalkeepers' defence in the World Cup in Korea and Japan in 2002. Eur J Sport Sci 8: 127–134, 2008.

15. Thomas A, Dawson B, Goodman C. The Yo-Yo Test: Reliability and association with a 20-m shuttle run and VO2max. Int J Sports Physiol Perform 1: 137–149, 2006.

16. Williams AM, Hodges NJ. Practice, instruction and skill acquisition in soccer: Challenging tradition. J Sports Sci 23: 637–650, 2005.

17. Ziv G, Lidor R. Physical characteristics, physiological attributes, and on-field performances of soccer goalkeepers. Int J Sports Physiol Perform 6: 509–524, 2011.