Test–Retest Reliability of the Yo-Yo Test: A Systematic Review

Test-retest reliability of the Yo-Yo test: a systematic review

Jozo Grgic,[1] Luca Oppici,[1] Pavle Mikulic,[2] Jens Bangsbo,[3] Peter Krustrup,[4,5] Zeljko Pedisic,[1]

[1]Institute for Health and Sport (IHES), Victoria University, Melbourne, Australia.

[2]Faculty of Kinesiology, University of Zagreb, Zagreb, Croatia.

[3]Department of Nutrition, Exercise and Sports (NEXS), University of Copenhagen, Copenhagen N, Denmark.

[4]Department of Sports Science and Clinical Biomechanics, Sport and Health Sciences Cluster (SHSC), University of Southern Denmark, Odense, Denmark.

[5]Faculty of Life and Environmental Sciences, Sport and Health Sciences, University of Exeter, Exeter, UK.

Short title: Test-retest reliability of the Yo-Yo test

Grgic, J., Oppici, L., Mikulic, P., Bangsbo, J., Krustrup, P., & Pedisic, Z. (2019). Test–retest reliability of the Yo-Yo test: A systematic review. *Sports Medicine*, *49*(10), 1547-1557.

Corresponding author:

Jozo Grgic

Institute for Health and Sport (IHES), Victoria University, Melbourne, Australia

jozo.grgic@vu.edu.au

Abstract

Background

The Yo-Yo test is widely used both in the practical and research contexts; however, its true test-retest reliability remains unclear.

Objective

The present systematic review aims to identify studies that have examined the test-retest reliability of the Yo-Yo test and summarize their results.

Methods

A search of ten databases was performed to find studies that have investigated test-retest reliability of any variant of the Yo-Yo test. The COSMIN checklist was employed to assess the methodological quality of the included studies.

Results

Nineteen studies of excellent or moderate methodological quality were included. When considering all variants of the Yo-Yo test, the included studies reported intra-class correlation coefficients (ICCs) for test-retest reliability ranging from 0.78 to 0.98 where 62% of all ICCs were higher than 0.90, while 97% of ICCs were higher than 0.80. The coefficients of variation (CVs) ranged from 3.7% to 19.0%. Regardless of the variant of the test, the participants' familiarization with the test, and previous sport experience, the ICCs generally seem high (≥ 0.90) and CVs low (<10%).

Conclusion

The results of this review indicate that the Yo-Yo test (in all its variants) generally has good to excellent test-retest reliability. The evidence concerning reliability arises from 19 included

studies that were of moderate or high methodological quality. Considering that most of the included studies examined the Yo-Yo intermittent recovery level 1 test while including Association Football players, more reliability studies examining Yo-Yo intermittent recovery level 2 test, Yo-Yo intermittent endurance level 1 and level 2 tests, and in the context of sports other than Association Football as well as in non-athletic populations, are required. Finally, future studies should explicitly state the type of ICC used for the reliability data analysis to allow for better between-study comparisons.

Key points:

- The Yo-Yo test has good to excellent test-retest reliability.
- The reliability seems to be similarly high for different variants of the Yo-Yo test.
- Reliability also seems to be high regardless of the participants' prior familiarization with the test.

1 Introduction

The Yo-Yo test variants are widely used due to their simplicity (i.e. these tests are of low cost as they require only cones and speakers), high validity for estimating the capacity of an individual to perform repeated exercise, and because they allow assessment of up to 15 athletes at once [1]. These tests allows coaches to readily examine their athlete's level of fitness and long-term changes in performance, providing information that shapes the design of a training intervention [1]. The Yo-Yo tests are divided into two commonly used forms, the Yo-Yo intermittent recovery test, and the Yo-Yo intermittent endurance test.

In the Yo-Yo intermittent recovery test, the individual is required to run distances of 2 × 20 m at progressively increasing speeds, interspersed with a 10-s period (controlled by audio signals) of jogging around a marker placed 5 m behind the finish line after each 40 m [1]. The test ends when the participant chooses to terminate it, or when he/she cannot finish the shuttle run in time on two consecutive occasions due to exhaustion. The outcome of this test is the total distance successfully covered. This test results in a very high aerobic loading and low-to-moderate anaerobic loading. The Yo-Yo intermittent recovery test has two levels (level 1 and level 2), with the level 2 version starting at a higher initial speed. Due to the higher initial speed, level 2 (as compared to the level 1 version) has a greater contribution from the anaerobic energy system and is generally employed for those with a higher level of physical fitness [1]. Due to the greater contribution from the anaerobic energy system, the level 2 variant likely has greater practical application in predominately anaerobic based sports [1].

The Yo-Yo intermittent endurance test is different from the intermittent recovery version, as the speeds are lower, there are more repetitions at each speed, the progressive increases in

speed are lower and it includes a 5-s recovery of jogging around a marker placed 2.5 m behind the finish line after each interval [2]. Like the intermittent recovery test, this version also has two different levels (i.e., level 1 and level 2) that differ in their initial speed. The Yo-Yo intermittent endurance test is useful in determining an individual's ability to recover from repeated exercise that has a high contribution from the anaerobic system [1, 2].

In 2008, Bangsbo et al. [1] published a review of the Yo-Yo intermittent recovery test, and summarized several important aspects, including: (a) physiological responses during the test; (b) performance in the test across different groups of athletes; (c) use of the test for tracking seasonal changes in performance; (d) the relationship between the test results and maximal oxygen uptake (VO_{2max}); and (e) the test-retest reliability of the test. In the review, Bangsbo and colleagues [1] suggested that the Yo-Yo test is a highly reliable tool for determining an athlete's ability to perform intermittent exercise. However, at the time of publication of the review [1], only three studies [3-5], with a pooled sample of 75 male participants (62 of whom were not competitive athletes), had examined the test-retest reliability of the Yo-Yo test. In addition, no review to date has compiled data on the test-retest reliability of Yo-Yo intermittent endurance test. Correct information about test reliability can help practitioners determine whether a change in the performance of their client in the test is greater than random variation [6, 7]. For researchers, correct information about test reliability is essential for the interpretation of statistical results, because measurement error may attenuate effect sizes and increase the probability of type II error [6, 7].

In recent years, several studies have been published that examined the test-retest reliability of the Yo-Yo test in diverse population groups, including women, trained and untrained children,

Association Football assistant referees, Association Football athletes, rugby athletes, and basketball players [8–22]. Given the increasing amount of empirical evidence on this topic, and the broad application of this test in both practical and research contexts [1], a comprehensive review of the test-retest reliability of the Yo-Yo test appears to be warranted. The present systematic review, therefore, aims to identify studies that have examined the test-retest reliability of the Yo-Yo test and summarize their results. The findings obtained may help elucidate the true reliability of this test, and thus provide a valuable resource for practitioners and researchers interested in using the Yo-Yo test or interpreting its results.

2 Methods

2.1 Search strategy

The guidelines proposed by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were followed for this review [23]. Two authors of the review (the first and second author) conducted the literature searches independently to minimize potential selection bias. The searches were conducted through the following databases: Academic Search Premier, CINAHL, ERIC, Health Source: Nursing/Academic Edition, MasterFILE Premier, PsycINFO, PubMed/MEDLINE, Scopus, SPORTDiscus, and Web of Science. In all these databases, the following search syntax was employed: (“yo yo” OR “yo-yo” OR “yoyo”) AND (reliability OR repeatability OR reproducibility). The search was performed on November 14th, 2018.

2.2 Inclusion criteria

Studies meeting the following criteria were included in the present review: (1) published as a full-text article in English and a peer-reviewed journal; (2) investigated test-retest reliability of any variant of the Yo-Yo test (including any modified versions); (3) presented intra-class correlation coefficients (ICCs) and/or coefficient of variation (CV); and (4) included human participants. The ICCs were classified as poor reliability (less than 0.50), moderate reliability (between 0.50 and 0.75), good reliability (between 0.75 and 0.90), and excellent reliability (greater than 0.90), according to Koo and Li [24].

2.3 Data extraction

Two of the authors (the first and second author) extracted the following data from the included studies: (1) sample size, participants' age, and sports experience; (2) variant of the Yo-Yo test employed; (3) number of days between the first (test) and second (re-test) test assessment; and (4) ICC and/or CV values for test-retest reliability with the outcome being distance covered. Following data extraction, the authors cross-checked the files, and any differences in the data extraction files between the authors were resolved through discussion and consensus.

2.4 Methodological quality

The methodological quality of the included studies was evaluated using the COSMIN checklist [25]. Specifically, we used Form B of the COSMIN checklist, as this form is designed for reliability studies. Form B has 11 items that refer to measurement administration, number of measurements, the time interval between assessments, reporting of missing items, the similarity of conditions for both measurements, adequacy of the sample size, important flaws in the study design, and reporting of ICCs. The response "yes" to all but one checklist

item corresponds to a point. For item 10, the answer "no" corresponds to a point, as this question asks whether there were any important flaws in the design or methods of the study. The appraised studies were classified according to their summary score on the COSMIN checklist as follows, a summary score of 10–11 points was considered "excellent" methodological quality; 7–9 points was considered "moderate" methodological quality; less than 7 points was considered "poor" methodological quality. Studies were rated independently by two reviewers (the first and second author); discrepancies in the assessments between the authors were resolved through discussion before reaching a consensus.

3 Results

3.1 Study selection

The total number of search results from the ten databases was 374 (Fig. 1). Of those, 33 papers were read in full; the remainder of the search results were excluded based on their titles and/or abstracts. Fourteen of the 33 papers were excluded from the review either because they did not examine the test-retest reliability of the Yo-Yo test or because they presented duplicate data (i.e. previously presented in another paper). Nineteen studies [2–5, 8–22] were included in the present review.

*****Insert Figure 1 about here*****

3.2 Study characteristics

The total number of participants across all the included studies was 802, while the individual sample sizes ranged from 13–142. The median number of participants per study was 28.

Sixteen studies were conducted in team-sport athletes, including Association Football (the most common study population), rugby, basketball players, and wheelchair basketball players. Six studies were conducted in individuals without any individual or team-sports experience (note that three studies included both athletes and untrained participants). The most commonly used period between first testing and repeated testing was 7 days (range: 2–8 days). Of the 19 included studies, seven did not present ICC values while one did not present CV values. The summary of the included studies, along with the reliability data, is presented in Table 1.

*****Insert Table 1 about here*****

3.3 Methodological quality

Three studies were classified as being of excellent methodological quality, while all the remaining studies were classified as being of moderate methodological quality. No studies were classified as being of poor methodological quality. The average score on the checklist was 9 points (range: 8–11 points). Detailed results of the assessment against individual COSMIN checklist items can be found in Table 2.

*****Insert Table 2 about here*****

3.4 Results for all variants of the Yo-Yo test

The pooled sample size across all included studies was 802 participants. When considering all variants of the Yo-Yo test, the ICCs in the included studies ranged from 0.78 to 0.98 where 62% of all ICCs were higher than 0.90, while 97% of ICCs were higher than 0.80. The CVs ranged from 3.7% to 19.0%.

3.4.1 Yo-Yo intermittent recovery level 1 test

Twelve studies (*n* = 420) explored the reliability of the Yo-Yo intermittent recovery level 1 test. The ICCs ranged from 0.78 to 0.98 where 71% of ICCs were higher than 0.90 and 93% of ICCs were higher than 0.80 (when the study that included wheelchair basketball players as participants was excluded, 77% of ICCs were above 0.90, and 92% above 0.80). The CVs ranged from 4.1% to 19.0%.

3.4.2 Yo-Yo intermittent recovery level 2 test

Five studies (*n* = 119) explored the reliability of the Yo-Yo intermittent recovery level 2 test. The ICCs ranged from 0.86 to 0.96 where 66% of ICCs were higher than 0.90. The CVs ranged from 4.2% to 12.7%.

3.4.3 Yo-Yo intermittent endurance level 1 test

Two studies (with 2 study samples in both studies; *n* = 88) explored the reliability of the Yo-Yo intermittent endurance level 1 test. The ICCs ranged from 0.85 to 0.95 where half of the ICCs were below and half above 0.90. The CVs ranged from 9.0% to 11.0%.

3.4.4 Yo-Yo intermittent endurance level 2 test

Six studies (*n* = 231) explored the reliability of the Yo-Yo intermittent endurance level 2 test. The ICCs ranged from 0.80 to 0.98 where 57% of ICCs were higher than 0.90. The CVs ranged from 3.9% to 15.0%.

3.4.5 Effect of familiarization sessions on Yo-Yo test reliability

Eleven studies (*n* = 361) had a familiarization session as part of their testing protocol or included participants who had already been familiarized with the Yo-Yo test. The ICCs ranged from 0.78 to 0.98 where 69% of ICCs were higher than 0.90 and 92% of ICCs were higher than 0.80. The CVs ranged from 4.1% to 17.3%.

Nine studies (*n* = 441) were conducted in participants who had not previously been familiarized with the Yo-Yo test and that did not have a familiarization session as part of their testing protocol. Across the included studies, the ICCs ranged from 0.80 to 0.98 where 60% of ICCs were higher than 0.90. The CVs ranged from 3.9% to 19.0%.

3.4.6 Association Football and other sports

Eleven studies (*n* = 414) conducted their reliability studies among Association Football players as participants. Across the included studies, the ICCs ranged from 0.78 to 0.98 where 80% of ICCs were higher than 0.90, while 93% of ICCs were higher than 0.80. The CVs ranged from 3.9% to 17.3%.

Three studies (*n* = 102) were conducted among athletes from sports other than Association Football. Across the included studies, the ICCs ranged from 0.83 to 0.97 where 75% of ICCs were higher than 0.90. The CVs ranged from 9.7% to 15.0%.

4 Discussion

The present review examined the test-retest reliability of the Yo-Yo test. Our findings indicate that the Yo-Yo test, in most cases, has good to excellent test-retest reliability. These findings seem consistent regardless of the variant of the test and of the participants' familiarization with the test. This evidence is based on a relatively large number of studies that were all of moderate or excellent methodological quality, which adds to the credibility of these findings. As we discuss further in the text, several important implications arise from this review.

Of the total number of included studies, 12 studies examined the reliability of the Yo-Yo intermittent recovery level 1 test. Other variants of the Yo-Yo test, namely, the Yo-Yo intermittent recovery level 2 (5 studies) and the Yo-Yo intermittent endurance level 1 and level 2 (2 and 6 studies, respectively), have been much less explored in the context of test-retest reliability. The findings for the three latter variants of the Yo-Yo test are based on a handful of studies, which does indicate that a certain degree of caution must be exercised when interpreting the respective results and drawing conclusions. Nevertheless, the available ICCs are suggestive of good to excellent test-retest reliability; however, future studies may be needed to confirm these findings.

It might be hypothesized that the reliability of the Yo-Yo test would be lower if the participants were not provided with a practice session or if they had not already been familiarized with the test during their usual training routines, and *vice versa*. However, studies that were conducted in participants with no prior familiarization with the Yo-Yo test, and studies that included participants who had already been familiar with this test, reported very similar reliability values. In both cases, ICCs were suggestive of good to excellent reliability and these were coupled with mostly low CVs. Based on these results, we may conclude that a familiarization session with the Yo-Yo test might not be needed to achieve high test-retest reliability. If practitioners estimate that some familiarization with the test is warranted, familiarization could be incorporated as part of the warm-up. As an example, Dobbin et al. [14] allowed two practice shuttles prior to starting the test. Similarly, Ahler et al. [8] provided a trial run for the first 2 min of the test.

Historically, the Yo-Yo test has been used most often in Association Football [1]. Thus, it is not surprising that most of the studies that included athletes as participants were conducted with Association Football players. Due to the high prevalence of studies with Association Football players as participants, it might be argued that the generalizability of reliability values is predominantly in the context of Association Football. With that being said, ICCs for the studies that included athletes from other sports were in the range from 0.83 to 0.97 which is suggestive of good to excellent reliability. In total, however, only three studies established the test-retest reliability of the Yo-Yo test in athletes from sports other than Association Football. The overall low number of studies conducted in athletes competing in sports other than Association Football opens up an avenue for future high-quality studies to explore the reliability of the Yo-Yo test in specific populations of athletes.

4.1 Methodological quality

Of the 19 included studies, 16 were classified as being of moderate quality, while three studies [16, 19, 20] were classified as being of excellent methodological quality. It is also relevant to note that seven included studies did not present the test-retest ICCs and did not score a point on item 11. Hopkins [26] highlighted that the test-retest ICC is a good measure of reliability. However, he has also suggested that CV may provide greater applicability in practical settings. In that context, we would recommend that future reliability studies present ICC and CV, as both may provide relevant and valuable information regarding test-retest reliability. A detailed description of these measures, their limitations, and their application for assessing reliability in the context of sports medicine and related disciplines can be found elsewhere [6].

4.2 Limitations

There are several types of ICC coefficients, and their values may differ [27]. The ICC of choice for presenting the test-retest reliability of the Yo-Yo test is an ICC for absolute agreement and single measurements and based on a two-way mixed effect model, sometimes denoted as ICC (3,1). The ICC for agreement should be preferred to reflect the possible disagreements between test and retest scores in absolute value (e.g. because of the learning effect). The ICC for single measurements (rather than for the average score of the test and the retest) should be preferred given that when using this test, researchers and practitioners will most likely rely on a single measurement. The use of the average score of two or more repeated measurements is unlikely because this would require testing over multiple days, given this is an endurance test and participants need to be allowed enough time to recuperate between the measurements completely. A two-way model should be preferred because the

order of measurements is relevant in the context of the reliability of this test, that is, the test and the retest may differ systematically. The variation in results across the included studies may be attributed to the fact that the type of ICC was not reported in most of the included papers. To allow for better between-study comparisons future studies should specify the type of ICC used in the analysis. Guidance on this topic can be found in a recent paper by Koo and Li [24], and in a review by Trevethan [28].

4.3 Factors that may affect reliability of the Yo-Yo test

If not controlled, several factors may affect Yo-Yo test performance. For example, caffeine (a well-established performance-enhancing supplement [29]) has been shown to acutely increase total distance covered in the Yo-Yo test [30]. Additionally, fasting may also impact performance in this test and thus affect reliability data [31]. Ten included studies [4, 5, 8, 14-16, 18-21] noted that they advised the participants to maintain their usual nutritional habits during the study with three [4, 14, 18] specifically restricting any caffeine intake. Additionally, exercise performance varies according to the time of day at which the testing is conducted [32] and therefore may also impact performance in the Yo-Yo test. In one study [33] that explored the time of day effects on performance in this test, the total distance covered was higher at 17:00 h, as compared to 07:00 h. While most studies standardized the testing time of day, three studies [2, 10, 17] did not specify this information. Some form of encouragement during the Yo-Yo test was provided to the participants in four studies [5, 12, 18, 21]. Encouragement may introduce a source of variation given that the reaction of the participants to encouragement may differ between individuals [7]; therefore, Currell and Jeukendrup [7] suggested that encouragement should be avoided during the test. Other factors such as sleep duration the night before the testing session may also impact performance on the Yo-Yo test [34]. Future studies that explore test-retest reliability of the Yo-Yo test should

control for these factors that may affect reliability and specify these data in their study methods section.

One additional caveat that needs to be taken into account is that some of the included studies conducted the testing sessions indoors, on an artificial grass surface, whereas others held the testing sessions outdoors (e.g. on a basketball court; Table 1). It currently remains unclear if the use of different testing conditions across the studies affected the reliability values. The reliability data from studies that conducted the testing sessions indoors and outdoors seem to be similar. To further explore this area, future studies may consider comparing test-retest reliability of the Yo-Yo test in different environmental conditions (e.g. testing indoors vs. outdoors; testing on artificial vs. natural turf).

5 Conclusion

The results of this review indicate that the Yo-Yo test has good to excellent test-retest reliability. These findings seem consistent regardless of the variant of the test and of the participants' prior familiarization with the test. Reliability seems high both in Association Football players and in athletes competing in other sports such as rugby and basketball; however, this requires further study. The evidence concerning reliability arises from 19 included studies that were of moderate or high methodological quality. More reliability studies are needed for Yo-Yo intermittent recovery level 2 and Yo-Yo intermittent endurance level 1 and level 2 in the context of sports other than Association Football as well as in non-athletic populations. Future studies should explicitly state the type of ICC used for the reliability data analysis to allow for better between-study comparisons.

Notes

Compliance with Ethical Standards

Funding

No sources of funding were used to assist in the preparation of this article.

Conflict of interest

Jozo Grgic, Luca Oppici, Pavle Mikulic, Jens Bangsbo, Peter Krustrup and Zeljko Pedisic declare that they have no conflicts of interest relevant to the content of this review.

Author Contributions

JG and ZP conceived the idea for the review. JG and LO conducted the study selection, data extraction, and quality assessment. JG drafted the initial manuscript. LO, PM, JB, PK, and ZP contributed to writing the manuscript. All authors approved the final version of the manuscript.

References

1. Bangsbo J, Iaia FM, Krustrup P. The Yo-Yo intermittent recovery test: a useful tool for evaluation of physical performance in intermittent sports. Sports Med. 2008;38(1):37–51.
2. Bradley PS, Mohr M, Bendiksen M, et al. Sub-maximal and maximal Yo-Yo intermittent endurance test level 2: heart rate response, reproducibility and application to elite soccer. Eur J Appl Physiol. 2011;111(6):969–78.
3. Krustrup P, Mohr M, Amstrup T, et sl. The yo-yo intermittent recovery test: physiological response, reliability, and validity. Med Sci Sports Exerc. 2003;35(4):697–705.
4. Krustrup P, Mohr M, Nybo L, et al. The Yo-Yo IR2 test: physiological response, reliability, and application to elite soccer. Med Sci Sports Exerc. 2006;38(9):1666–73.
5. Thomas A, Dawson B, Goodman C. The yo-yo test: reliability and association with a 20-m shuttle run and VO(2max). Int J Sports Physiol Perform. 2006;1(2):137–49.
6. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26(4):217–38.
7. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. Sports Med. 2008;38(4):297–316.
8. Ahler T, Bendiksen M, Krustrup P, et al. Aerobic fitness testing in 6- to 9-year-old children: reliability and validity of a modified Yo-Yo IR1 test and the Andersen test. Eur J Appl Physiol. 2012;112(3):871–6.
9. Bradley PS, Bendiksen M, Dellal A, et al. The application of the Yo-Yo intermittent endurance level 2 test to elite female soccer populations. Scand J Med Sci Sports. 2014;24(1):43–54.

10. Castagna C, Bendiksen M, Impellizzeri FM, et al. Reliability, sensitivity and validity of the assistant referee intermittent endurance test (ARIET) - a modified Yo-Yo IE2 test for elite soccer assistant referees. J Sports Sci. 2012;30(8):767–75.
11. Castagna C, Krustrup P, D'Ottavio S, et al. Ecological validity and reliability of an age-adapted endurance field test in young male soccer players. J Strength Cond Res. 2018. doi: 10.1519/JSC.0000000000002255.
12. Deprez D, Coutts AJ, Lenoir M, et al. Reliability and validity of the Yo-Yo intermittent recovery test level 1 in young soccer players. J Sports Sci. 2014;32(10):903–10.
13. Deprez D, Fransen J, Lenoir M, et al. The Yo-Yo intermittent recovery test level 1 is reliable in young high-level soccer players. Biol Sport. 2015;32(1):65–70.
14. Dobbin N, Hunwicks R, Highton J, et al. Reliable testing battery for assessing physical qualities of elite academy rugby league players. J Strength Cond Res. 2018;32(11):3232–38.
15. Ehlert AM, Cone JR, Wideman L, et al. Evaluation of a goalkeeper-specific adaptation to the Yo-Yo intermittent recovery test level 1: reliability and variability. J Strength Cond Res. 2019;33(3):819–24.
16. Enright K, Morton J, Iga J, et al. Reliability of "in-season" fitness assessments in youth elite soccer players: a working model for practitioners and coaches. Sci & Med in Football. 2018;2(3):177–83.
17. Fanchini M, Castagna C, Coutts AJ, et al. Are the Yo-Yo intermittent recovery test levels 1 and 2 both useful? Reliability, responsiveness and interchangeability in young soccer players. J Sports Sci. 2014;32(20):1950–7.
18. Kong Z, Qi F, Shi Q. The influence of basketball dribbling on repeated high-intensity intermittent runs. J Exerc Sci Fit. 2015;13(2):117–22.

19. Póvoas SC, Castagna C, da Costa Soares JM, et al. Reliability and construct validity of Yo-Yo tests in untrained and soccer-trained schoolgirls aged 9-16. Pediatr Exerc Sci. 2016;28(2):321–30.
20. Póvoas SC, Castagna C, Soares JM, et al. Reliability and validity of Yo-Yo tests in 9- to 16-year-old football players and matched non-sports active schoolboys. Eur J Sport Sci. 2016;16(7):755–63.
21. Silva CD, Natali AJ, Lima JRP, et al. Yo-Yo IR2 test and margaria test: validity, reliability and maximum heart rate in young soccer players. Rev Bras Med Esporte. 2011;17(5):344–9.
22. Yanci J, Granados C, Otero M, et al. Sprint, agility, strength and endurance capacity in wheelchair basketball players. Biol Sport. 2015;32(1):71–8.
23. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264–9.
24. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63.
25. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539–49.
26. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1–15.
27. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30–46.

28. Trevethan R. Intraclass correlation coefficients: clearing the air, extending some cautions, and making some requests. Health Serv Outcomes Res Methodol. 2017;17(2);127–43.

29. Grgic J, Grgic I, Pickering C, et al. Wake up and smell the coffee: Caffeine supplementation and exercise performance—an umbrella review of 21 published meta-analyses. Br J Sports Med. 2019. doi: 10.1136/bjsports-2018-100278

30. Mohr M, Nielsen JJ, Bangsbo J. Caffeine intake improves intense intermittent exercise performance and reduces muscle interstitial potassium accumulation. J Appl Physiol. 2011;111(5):1372–9.

31. Chtourou H, Hammouda O, Souissi H, et al. The effect of ramadan fasting on physical performances, mood state and perceived exertion in young footballers. Asian J Sports Med. 2011;2(3):177–85.

32. Grgic J, Lazinica B, Garofolini A, et al. The effects of time of day-specific resistance training on adaptations in muscular hypertrophy and strength: a systematic review and meta-analysis. Chronobiol Int. 2019;36(4):449–60.

33. Chtourou H, Hammouda O, Souissi H, et al. Diurnal variations in physical performances related to football in young soccer players. Asian J Sports Med. 2012;3:139–44.

34. Mejri MA, Yousfi N, Mhenni T, et al. Does one night of partial sleep deprivation affect the evening performance during intermittent exercise in Taekwondo players? J Exerc Rehabil. 2016;12(1):47–53.

Figure 1. Flow diagram of the search and study selection process

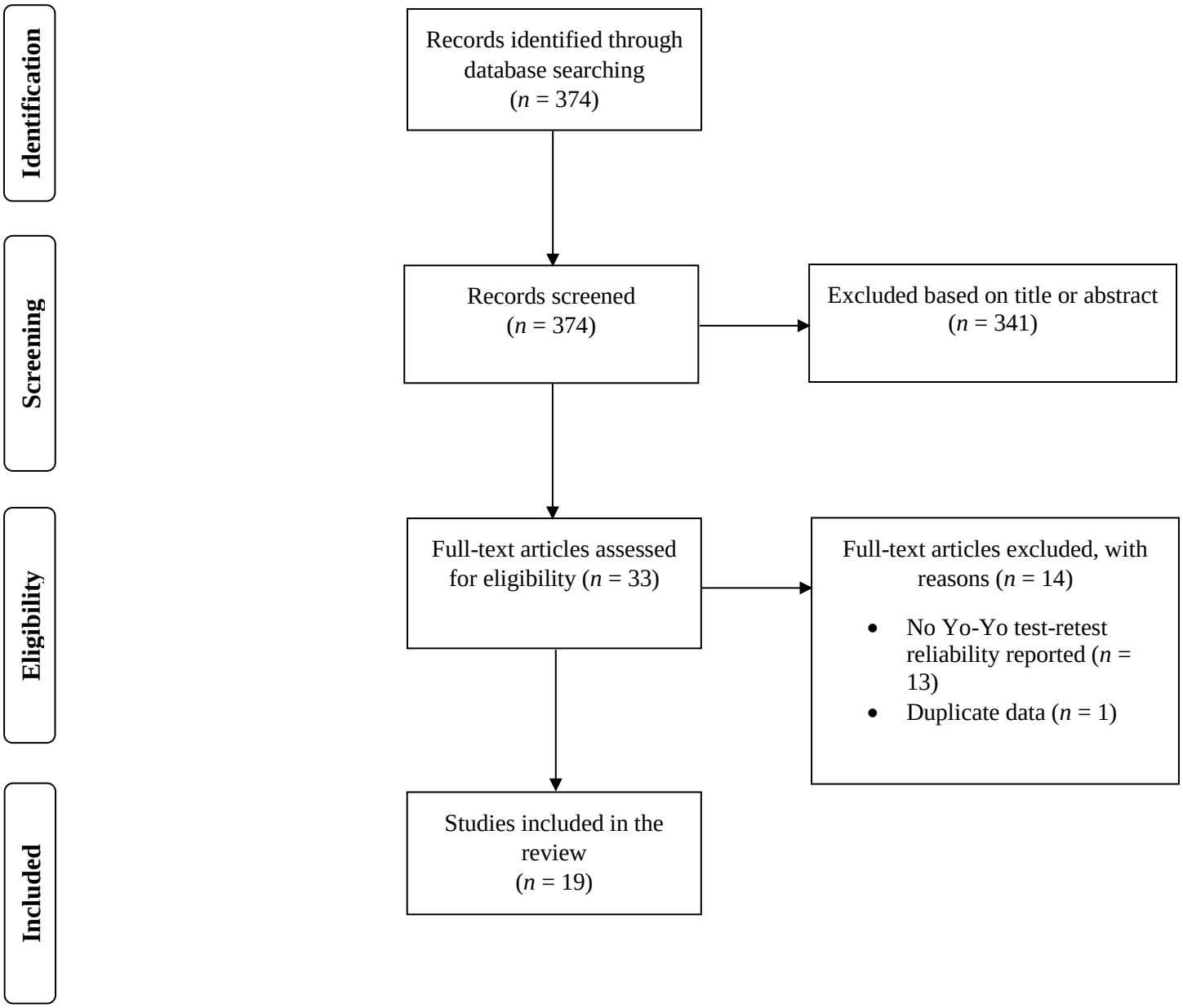

Table 1. Summary of the studies included in the review

Study	Sample	Sports experience	Yo-Yo test employed	Testing environment	Familiarization with the test	Days between tests	ICC	CV
Ahler et al. [8]	31 boys and girls	None	Intermittent recovery level 1, modified for children	Indoors, handball court	None[a]	2 to 3 days	Not provided	19%
Bradley et al. [9]	27 adult females	Association Football players	Intermittent endurance level 2	Indoors, artificial grass surface	Had previous experience	7 days	Not provided	4.5%
Bradley et al. [2]	37 males (classified as "youth")	Association Football players	Intermittent endurance level 2	Indoors, artificial grass surface	None	7 days	Not provided	3.9%
Castagna et al. [10]	41 adult males	Association Football assistant referees	Intermittent endurance level 2, modified for Association Football assistant referees	Artificial surface	None	7 days	0.98	4.3%
Castagna et al. [11]	28 boys	Association Football players	Intermittent recovery level 1, modified for children	Natural grass Association Football pitch	One familiarization session	7 days	0.94	5.1%
Deprez et al. [12]	35 U13 male Association Football players, 32 U15 male Association Football players and 11 U17 male Association Football players	Association Football players	Intermittent recovery level 1	Indoor venue	Had previous experience	7 days	U13: 0.82 U15: 0.85 U17: 0.94	U13: 17.3% U15: 16.7% U17: 7.9%
Deprez et al. [13]	22 U15 male Association Football players, 10 U17 male Association Football players and 4	Association Football players	Intermittent recovery level 1	Outdoors, artificial turf	Had previous experience	7 days	Not provided	U15: 6.8% U17: 4.3% U19: 4.1%

	U19 male Association Football players							
Dobbin et al. [14]	50 males (age mean ± SD: 17.1 ± 1.1 years)	Academy rugby league players	Intermittent recovery level 1, modified for rugby players	Outdoors, synthetic grass pitch	None[a]	8 days	0.97	9.7%
Ehlert et al. [15]	8 men and 8 adult women	Association Football goalkeepers	Intermittent recovery level 1, modified for goalkeepers	Not specified	None	5 to 7 days	Men: 0.98 Women: 0.96	Men: 5.8% Women: 9.6%
Enright et al. [16]	19 adult males	Association Football players	Intermittent recovery level 2	Outdoor, grass pitch	Had previous experience	7 days	0.96	4.2%
Fanchini et al. [17]	24 males (age mean ± SD: 17 ± 1 years)	Association Football players	Intermittent recovery level 1 and intermittent recovery level 2	Not specified	Had previous experience	7 days	Level 1: 0.78 Level 2: 0.93	Level 1: 7.3% Level 2: 7.1%
Kong et al. [18]	36 males (aged 13-18 years)	Basketball players	Intermittent endurance level 2, with and without dribbling	Outdoor, basketball court	Had previous experience	Up to 7 days	With dribbling: 0.92 Without dribbling: 0.83	With dribbling: 12.6% Without dribbling: 15%
Krustrup et al. [3]	13 adult males	None	Intermittent recovery level 1	Indoor, running lanes	One familiarization session	7 days	Not provided	4.9%
Krustrup et al. [4]	29 adult males	16 “normally” trained and 13 Association Football players	Intermittent recovery level 2	Indoor, artificial turf	None	7 days	Not provided	9.6%

Póvoas et al. [19]	132 girls (aged 9-16 years)	67 classified as "untrained" and 65 Association Football players	Intermittent recovery level 1, modified for children; intermittent endurance level 1; and intermittent endurance level 2	Indoor, futsal court for the untrained group; artificial turf Association Football pitch for the Association Football players	None	7 days	*Untrained groups* Intermittent recovery level 1: 0.89 Intermittent endurance level 1: 0.87 Intermittent endurance level 2: 0.80 *Trained groups* Intermittent recovery level 1: 0.97 Intermittent endurance level 1: 0.95 Intermittent endurance level 2: 0.97	*Untrained groups* Intermittent recovery level 1: 9% Intermittent endurance level 1: 11% Intermittent endurance level 2: 9% *Trained groups* Intermittent recovery level 1: 10% Intermittent endurance level 1: 9% Intermittent endurance level 2: 8%
Póvoas et al. [20]	142 boys (aged 9-16 years)	72 classified as "untrained" and 70 Association Football players	Intermittent recovery level 1, modified for children; intermittent endurance level 1; and intermittent endurance level 2	Indoor, futsal court for the untrained group; artificial turf Association Football pitch	The Association Football players had previous experience	7 days	*Untrained groups* Intermittent recovery level 1: 0.95 Intermittent endurance level 1: 0.85	*Untrained groups* Intermittent recovery level 1: 9.3%

				for the Association Football players			Intermittent endurance level 2: 0.84 *Trained groups* Intermittent recovery level 1: 0.98 Intermittent endurance level 1: 0.95 Intermittent endurance level 2: 0.93	Intermittent endurance level 1: 10.2% Intermittent endurance level 2: 8.5% *Trained groups* Intermittent recovery level 1: 11.1% Intermittent endurance level 1: 10.2% Intermittent endurance level 2: 8.5%
Silva et al. [21]	14 boys	Association Football players	Intermittent recovery level 2	Natural turf	Had previous experience	7 days	Not provided[b]	11%
Thomas et al. [5]	33 adult men	None	Intermittent recovery level 1 and intermittent recovery level 2	Indoors, wooden sprung floor	None	3 to 7 days	level 1: 0.95 level 2: 0.86	level 1: 8.7% level 2: 12.7%
Yanci et al. [22]	16 adult males	Wheelchair basketball players	Intermittent recovery level 1, modified for	Indoors, synthetic court	One familiarization session	7 days	0.94	Not presented

			wheelchair basketball players					

ICC intra-class correlation coefficient
CV coefficient of variation
SD standard deviation
U13 under 13 years old
U15 under 15 years old
U17 under 17 years old
U19 under 19 years old
[a] familiarization was incorporated as a part of the warm-up
[b] incorrect ICC values presented in the manuscript

Table 2. Results of the methodological quality assessment of the included studies using the COSMIN checklist

Study	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Total score
Ahler et al. [8]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	8
Bradley et al. [9]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	8
Bradley et al. [2]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	8
Castagna et al. [10]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9
Castagna et al. [11]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9
Deprez et al. [12]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9
Deprez et al. [13]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	8
Dobbin et al. [14]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9
Ehlert et al. [15]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9
Enright et al. [16]	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	10
Fanchini et al. [17]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9
Kong et al. [18]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9
Krustrup et al. [3]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	8
Krustrup et al. [4]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	8
Póvoas et al. [19]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	10
Póvoas et al. [20]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	10
Silva et al. [21]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	8
Thomas et al. [5]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9
Yanci et al. [22]	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	9