



Interrater, Test-retest Reliability of the Y Balance Test: A Reliability Study Including 51 Healthy Participants

FREDERIK N. FOLDAGER^{†1}, SOFIE ASLERIN^{†1}, SARAH BÆKDAHL^{†1}, LISA U. TØNNING^{‡1,2}, and INGER MECHLENBURG^{‡1,2,3}

¹Department of Orthopaedic Surgery, Aarhus University Hospital, Aarhus N, DENMARK;

²Department of Clinical Medicine, Aarhus University, Aarhus N, DENMARK, ³Department of Public Health, Aarhus University, Aarhus N, DENMARK

[†]Denotes graduate student author, [‡]Denotes professional author

ABSTRACT

International Journal of Exercise Science 16(4): 182-192, 2023. The aim was to determine the relative and absolute interrater, test-retest reliability of the Y-Balance Test (YBT) in a sample of healthy and active adults aged 18 to 50 years. The sample consisted of 51 healthy and active participants, 30 men and 21 women with a mean age of 28 ± 7 years. The YBT was performed on the right leg in the three test directions. Test and retest of the YBT were performed with a median interval of 15 days. The method for data collection was in line with the Y Balance Test Lower Quarter Protocol (YBT-LQ). The test was conducted by raters previously inexperienced in the use of the YBT. The relative reliability was reported as Intraclass Correlation Coefficient ($ICC_{(2,1)}$). The absolute reliability was reported as Standard Error of Measurement (SEM) and Minimal Detectable Change (MDC). The ICC ranged from 0.79-0.86. SEM ranged from 2%-4%, indicating the measurement error at group level and MDC ranged from 5%-11%, indicating the measurement error at individual level. The YBT showed good relative and absolute reliability. The YBT is therefore considered suitable at both group and individual level in physically active populations.

KEY WORDS: Postural balance, Star Excursion Balance Test (SEBT), Y Balance Test (YBT), intraclass correlation coefficient (ICC), standard error of measurement (SEM), minimal detectable change (MDC)

INTRODUCTION

The Y Balance Test (YBT) is a version of the Star Excursion Balance Test (SEBT) and was developed as a low-cost, time-efficient and reliable test to address some of the common errors in the SEBT (14). The YBT measures dynamic balance in only three directions: anterior,

posteromedial and posterolateral. A composite score is also calculated, which is the sum of the maximal reach in the three test directions (14).

The YBT is a performance-based tests requiring strength, range of movement, neuromuscular control, unilateral balance, proprioception and stability in the lower extremity joints (8, 14). The hip joint, which is a ball-and-socket joint allowing triplanar movement of the femur relative to the pelvis, serves as a central pivot point for the entire body (13). During the YBT unilateral stance is kept while supporting the body over one limb, which demands strong and specific activation of the hip muscles (13). Thus the YBT also measure hip stability and strength (21). Fullam et al. (6) found that the kinematics and muscular demands differed between the SEBT and the YBT and the tests should be considered as independent assessments. There has been limited research on the YBT measurement properties, including interrater, test-retest reliability, which mimics testing in clinical practice.

Previous studies on the interrater, test-retest reliability of the YBT indicate moderate to excellent relative reliability (ICC 0.68-0.97) for samples of less than 50 participants (5, 9, 11, 18). Previous studies have primarily been conducted in school-aged children and older adults (5, 9, 18). Shaffer et al. (17) examined the interrater, test-retest reliability of the YBT in a population of 64 healthy and active adult males aged 25 ± 4 years and found good relative reliability (ICC 0.80-0.85) and absolute reliability, where the Standard Error of Measurement (SEM) was 4-5% of the measurement and Minimal Detectable Change (MDC) was 11-15% of the measurement in the three test directions as well as the composite score.

Few studies have investigated the interrater, test-retest reliability of the YBT in a sample size above 50 participants as recommended in the COSMINS checklist for reliability and validity studies in young, healthy and active adults of both sexes (19). The only previous study on healthy adults with more than 50 participants by Shaffer et al. (17) studied a quite narrow age range, while this study targeted a much broader range of individuals based on age.

The aim of this study was to examine the relative and absolute interrater, test-retest reliability of the YBT in a sample of healthy and active adults of both sexes between 18 and 50 years of age. It was hypothesized that the relative reliability reported as ICC would be good, corresponding to an ICC > 0.75 indexed by Koo et al. (11) and that the absolute reliability reported as SEM and MDC would be on the same level as found by Shaffer et al. (17).

METHODS

Participants

A sample size of at least 50 participants was chosen a priori for the interrater, test-retest reliability study to achieve a minimum ICC of 0.8 with an expected 95% confidence interval width of 0.1 (3, 7, 10, 19).

Participants were recruited among students at Aarhus University, employees at Aarhus University Hospital, and through social media between November 2020 and April 2021. Participants were eligible for inclusion if they were 18-50 years old and fluent in reading and understanding Danish. Participants were excluded if they had 1) previous hip surgery, 2) lower extremity joint replacement, 3) rheumatoid, neurological or diseases of other etiology affecting the function of the hip muscles or joint or 4) vestibular dysfunction or other conditions affecting balance. Eligible participants were thoroughly informed about the aim of the study before consenting to participate. Eligibility criteria for inclusion and exclusion were investigated prior to testing.

Protocol

This interrater, test-retest reliability study was conducted and reported using the Guidelines for Reporting Reliability and Agreement Studies (12). To examine the interrater, test-retest reliability of the YBT the extent of agreement and reproducibility was calculated between two repeated measurements by two different raters. Test and retest were a priori scheduled with a 14-day interval to avoid genuine changes in dynamic balance and to avoid memory effect (15). Due to COVID-19 restrictions in Denmark, not all participants complied with the predetermined time interval between test and retest. Fifteen of the 51 included participants had follow-up times exceeding 30 days. These participants were not exposed to COVID-19, but retest wasn't possible due to extensive lock downs and the fact that participants wanted to exercise caution. Overall, this resulted in a median of 15 days between test and retest, ranging from 4 to 134 days. The tests were conducted by raters previously inexperienced in the use of the YBT.

Methods and procedures were in accordance with applicable guidelines, laws and the declaration of Helsinki II (1). All participants gave informed written consent to participate. The study was registered at the Danish Data Protection Agency (1-16-02-632-20).

Prior to testing, all participants completed questions of demographics and the University of California Los Angeles (UCLA) questionnaire which is a questionnaire where the patient chooses 1 out of 10 options that best describes their current physical activity level (22). Leg length measurement was performed and participants did a short warm-up consisting of two sets of ten repetitions each of alternating sets of calf-raises and chair stands.

Leg length measurement was inspired by Plisky et al. (14) Participants were instructed to lie in a supine starting position. The participants then lifted their hips and returned to the starting position. The rater passively straightened the legs to equalize the pelvis. The participants' right leg was measured in centimetres from the anterior superior iliac spine (ASIS) to the most distal part of the medial malleolus with a cloth tape. Leg length of the right leg was used to normalize data. Leg length measurement was conducted at the first trial.

The YBT was conducted and instructed according to the Y Balance Test Lower Quarter (YBT-LQ) Protocol (14). The only deviation from the YBT-LQ was that the test was performed on the right leg only. The YBT was completed using the Y Balance Test Kit™.

Prior to testing, the rater provided a verbal and visual demonstration of the test to each participant. Subsequently, each participant performed six practice trials in each direction to minimize learning effect and stabilize reliability. After six practice trials, participants performed three test trials in each direction. All trials were performed barefooted, standing on the right leg. Each participant was instructed to stand on the centre footplate, with the distal aspect of the right foot behind the red line marking the starting position. The predetermined test sequence was anterior, posteromedial, and posterolateral (Figure 1). While maintaining single leg stance, participants were instructed to push the reach-indicator block as far as possible with the contralateral leg and return to the starting position while maintaining postural control. Reach distance was recorded to the nearest 0.5 cm and was observed on the proximal side of the reach-indicator block. Attempts were discarded if the participant failed to return to the starting position, failed to maintain unilateral stance, used the reach-indicator block to maintain unilateral stance or accelerated the reach-indicator block to gain more distance. If a trial was discarded, participants repeated the trial. Each participant had six attempts to complete three valid trials. The maximal reach distance between the three valid trials was recorded in each direction and used for data analysis.

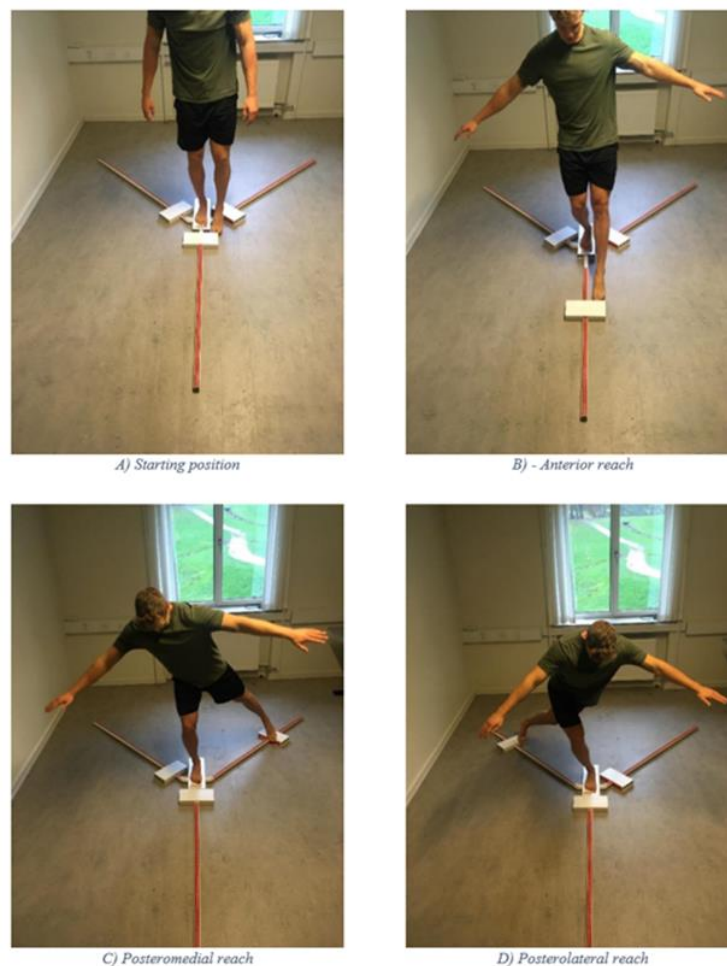


Figure 1. The Y Balance Test: a) starting position b) anterior reach c) posteromedial reach and d) posterolateral reach.

Raters: All participants were tested twice (test and retest) in the YBT on their right leg only by two different raters blinded to the previous test result. The time between testing sessions was planned to be 14 days apart.

The raters consisted of three Master of Health Science students. Before testing, all three raters conducted pilot tests, including use of the equipment and the YBT-LQ protocol (14). Raters were assigned to tests using a convenience sample method. The three raters conducted 39, 33 and 30 YBT tests or retests, respectively. Data were aggregated for analysis.

Statistical Analysis

Maximal reach in each direction and composite score were calculated as absolute values and values normalized to leg length. Normalization was calculated as maximal reach divided by leg length and multiplied by 100. Normalization allowed comparison between participants within the study and with previous studies. Maximal reach at the first test was compared to the maximal reach at retest using a paired t-test for each of the three test directions. A significant difference between test and retest was defined as a p-value < 0.05.

Relative reliability of the maximal reach in each direction and of the composite score was analyzed calculating ICC. According to Shrout and Fleiss (20) nomenclature, a two-way-analysis of variance, random effect model (ICC_(2,1)) was conducted. Absolute reliability of the maximal reach in each direction and of the composite score was analyzed calculating SEM and MDC (20). Calculations of SEM and MDC are presented in equations 1 and 2, wherein SD was the pooled SD in which the ICC was determined (20).

$$SEM = SD * \sqrt{1 - ICC} \quad (1)$$

$$MDC = 1.96 * \sqrt{2} * SEM \quad (2)$$

The SEM quantifies the precision of individual test scores and can be used to calculate the MDC, which can be used to define the amount of change in dynamic balance performance between test and retest to reflect a true difference (20). The SEM and MDC are reported as absolute values in centimeters and as a percentage of the grand mean to allow comparison with other studies. The grand mean is the average of several subsamples (test and retest).

A graphical illustration of repeatability was calculated plotting the mean difference, 95% CI to the mean difference and 95% Limits of Agreement (95% LoA) in Bland-Altman Plots (2).

Statistical Considerations: Continuous data were assessed for normality using histograms and probability plots. Normally distributed data were presented as means with standard deviations (SD). Non-normally distributed data were presented as medians with Inter-Quartile Range (25th-75th percentile). Categorical data were presented as number of events with percentages of total events. Statistical software, Stata 16.1, ©Statacorp LLC, College Station, Texas, was used to analyse data.

RESULTS

Descriptive statistics for participants are displayed in Table 1.

Table 1. Characteristics for participants performing the Y Balance Test.

	Minimum; Maximum	
Sex, n (%)	Male 30 (58.2) Female 21 (41.8)	
Age (years), Mean ± SD [95% CI]*	28 ± 7.2 [26.0 ; 30.0]	19 ; 50
Leg length (cm), Mean ± SD [95% CI]*	95.2 ± 5.5 [93.6 ; 96.7]	79 ; 105
UCLA Activity Score, median [IQR]†	7 [5 ; 9]	2 ; 10
Time between test-retest (days), median [IQR]†	15 [13 ; 69]	4 ; 134

*Age and leg length are presented with mean, standard deviation (SD) and 95% confidence intervals [95% CI].

†UCLA Activity Score and time between test and retest are presented with median and interquartile range [IQR].

Table 2. Y Balance Test: The absolute and normalized maximal reach at test and retest in the three test directions (anterior, posteromedial and posterolateral) and the composite score.

Test direction	Absolute Reach (cm)	Normalized Reach (%)	P-value‡
	Mean ± SD [95% CI]*	Mean ± SD [95% CI]*,†	
Anterior, Test	62.4 ± 7.8 [60.2 ; 64.6]	66 ± 6.9 [63.6 ; 67.4]	0.15
Anterior, Retest	61.4 ± 8.6 [58.9 ; 63.8]	64 ± 8.0 [62.2 ; 66.7]	
Difference, Anterior	1.0 ± 5.1 [-0.4 ; 2.5]	1 ± 5.3 [-0.4 ; 2.6]	
Posteromedial, Test	104.7 ± 9.6 [102.0 ; 107.5]	110 ± 8.2 [107.8 ; 112.4]	0.45
Posteromedial, Retest	105.4 ± 9.8 [102.7 ; 108.2]	111 ± 8.2 [108.3 ; 113.3]	
Difference, Posteromedial	-0.7 ± 6.1 [-2.4 ; 1.1]	-1 ± 6.4 [-2.5 ; 1.1]	
Posterolateral, Test	103.1 ± 9.4 [100.5 ; 105.7]	108 ± 9.0 [105.9 ; 110.9]	0.56
Posterolateral, Retest	102.8 ± 8.9 [100.1 ; 105.1]	108 ± 9.2 [105.4 ; 110.6]	
Difference, Posterolateral	0.4 ± 5.2 [-1.0 ; 1.9]	0.4 ± 5.4 [-1.1 ; 1.9]	
Composite Score, Test	270.3 ± 23.6	284 ± 20.2	
Composite Score, Retest	269.5 ± 24.6	283.3 ± 22.8	
Difference, Composite Score	0.8 ± 12.6	0.7 ± 13.2	

*Values are presented with mean ± standard deviation (SD) and 95% confidence intervals [95% CI].

†Normalized reach was calculated as (absolute reach/leg length)*100.

‡The paired t-test tests the null hypothesis that the difference for the absolute reach (cm) between test and retest was 0. A significant difference between test and retest was defined as a p-value < 0.05.

The maximal reach during YBT is reported in Table 2. There was no significant difference between test and retest for the absolute maximal reach, anterior reach ($p = 0.15$), posteromedial reach ($p = 0.45$), posterolateral reach ($p = 0.56$).

The relative reliability, ICC, and absolute reliability, SEM and MDC, for interrater, test-retest reliability for absolute maximal reach values are presented in Table 3.

Table 3. Relative (ICC) and absolute (SEM and MDC) interrater, test-retest reliability of the Y Balance Test.

Test direction	ICC [95% CI]	SEM stated in cm (% of the grand mean)	MDC stated in cm (% of the grand mean)
Anterior	0.79 [0.69 ; 0.90]	2.4 (4%)	6.5 (11%)
Posteromedial	0.79 [0.69 ; 0.90]	2.8 (3%)	7.8 (7%)
Posterolateral	0.83 [0.75 ; 0.92]	2.4 (2%)	6.6 (6%)
Composite Score	0.86 [0.78 ; 0.93]	4.7 (2%)	13.1 (5%)

95% CI: 95% confidence interval, ICC: Intra Class Correlation, SEM: Standard Error of Measurement, SEM (%): SEM in percent of the grand mean (i.e., the mean of test and retest), MDC: Minimal Detectable Change and MDC (%): MDC in percent of the grand mean (i.e., the mean of test and retest).

ICC varied from 0.79 to 0.86 in the three test directions and the composite score, which is considered a good relative reliability (11). The SEM and MDC showed the measurement error in absolute values, using the same units (centimeters) as the test and as a percentage of the grand mean in the three test directions and the composite score. SEM was 2-5 centimeters, corresponding to 2-4% of the grand mean. MDC was 7-13 centimeters, corresponding to 5-11% of the grand mean. The Bland Altman Plots showed that the difference between test-retest plotted against the mean difference was close to 0 in each of the three test directions and the composite score. The Bland Altman Plots did not show signs of heteroscedasticity in the data (Figure 2).

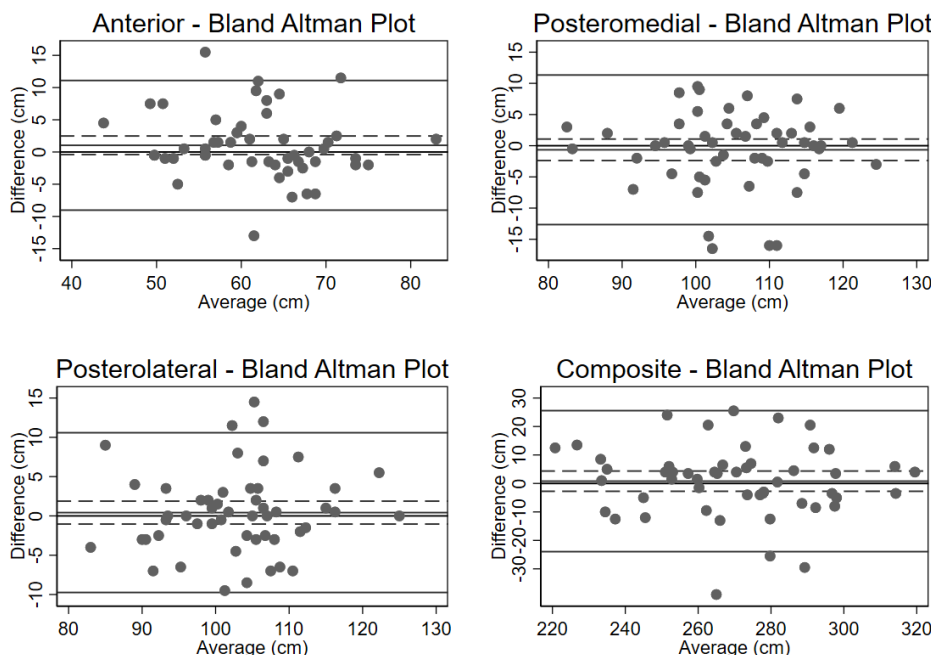


Figure 2. Bland Altman Plots for each of the three test directions (anterior, posteromedial and posterolateral) and the composite score. The mean of test and retest (x-axis) plotted against the difference between test and retest (y-axis). The two dashed lines around the mean difference is a 95% CI for the mean difference. The two thick lines above and below the mean difference represent the 95% limits of agreement, indicating the size of measurement errors.

DISCUSSION

The YBT demonstrated a good relative interrater, test-retest reliability with an ICC > 0.75 in the three test directions (anterior, posteromedial and posterolateral) and for the composite score, as the ICC varied from 0.79-0.86. The YBT also demonstrated a good absolute interrater, test-retest reliability as the SEM corresponded to 2-4% of the measurement and the MDC corresponded to 5-11% of the measurement. Thus, the YBT is a reliable test suitable to use at both group and individual level even among raters with no experience in the YBT. The clinical applicability of these results goes into how day-to-day and rater-to-rater variation in YBT measurements should be interpreted. Changes in measurements equivalent or greater than the SEM (2-4%) should be interpreted as changes larger than what is caused by statistical variation (3). Changes in measurements equivalent or greater than the MDC (5-11%) should be interpreting as changes that fall outside the measurement error and thus interpreted as a genuine and meaningful change, relevant for the participants (3).

Although direct comparison of the results of this study with other studies is limited to the specific timespan between test and retest, sample and raters, our results strengthen previous YBT reliability research, indicating that the YBT has a good relative and absolute reliability. Test-retest reliability of the YBT has been reported in several studies using different study designs (intrarater and interrater, test-retest reliability). Previous investigations have indicated moderate to excellent intrarater, test-retest reliability among 188 children aged 7-12 years old (ICC 0.75-0.90; (4), 21 children aged 12-13 years old (ICC 0.68-0.90; (9), and 178 adolescents aged 11-19 years old (ICC 0.40-0.97; (16). These results indicate that the YBT is a reliable tool to identify changes of dynamic balance performance over time at group level, especially in younger populations using the same rater.

Plisky et al. (14) reported excellent interrater reliability (ICC 0.97-1.00) in a population of 15 younger adults aged 19-20 years with raters scoring the same trials within the same test session. The excellent reliability, indicating almost perfect reliability, should be interpreted in that context. The ICC presented by Plisky et al. (14) is thus not directly comparable to the ICC (0.79-0.86) in this study. Freund et al. (5) reported a good to excellent interrater, test-retest reliability (ICC 0.82-0.97) in a subset population of 8 females aged 50-79 years. In that study there was only five minutes between test and retest and the sample size of 8 was the basis for calculating the interrater, test-retest reliability. Sipe et al. (18) reported excellent interrater, test-retest reliability (ICC 0.93-0.95) in a population of 30 older adults aged 67 ± 5 years. The presented ICCs by Plisky et al., Freund et al. and Sipe et al. should be interpreted with caution, as there is evidence that the strength of evidence of reliability studies is much stronger when the sample includes at least 50 participants (19).

Shaffer et al. (17) examined the interrater, test-retest reliability of the YBT during a 48-hour period in a population of 64 healthy and active adult males aged 25 ± 4 years and found a good relative reliability (ICC 0.80-0.85) and absolute reliability, where SEM was 4-5% of the measurement and MDC was 11-15% of the measurement in the three test directions and the

composite score. The ICC reported by Shaffer et al. was almost identical to the ICC (0.79-0.86) presented in this study, indicating that the YBT produces consistent results at group level for healthy and active adults among raters with no experience in use of the YBT. The absolute reliability reported as SEM and MDC by Shaffer et al. were slightly higher than those presented in this study. This discrepancy between the two studies may be due to the use of another formula to calculate SEM. When using the same equation as Shaffer et al., the absolute interrater, test-retest reliability in the present study was at the same level as reported by Shaffer et al. Thus similar result was obtained despite the greater variation in time between test and retest in the present study and despite the difference in study sample.

This study was conducted using a standardized testing protocol by Plisky et al. (14) which minimizes systematic errors between participants and raters in the YBT and therefore increases the internal validity of the study.

However, this study has some limitations. The timespan between test and retest was a median of 15 days and varied from a minimum of 4 days to a maximum of 134 days. This means that the dynamic balance may have changed between test and retest for some of the participants. The interrater, test-retest reliability would possibly have been better if investigated within a shorter and more uniform timespan as in other studies (5, 17, 18). We conducted a sub-analysis of the participants data due to the wide range of days between test-retest and the possibility of skewed results. We conducted a sub-analysis with the 36 participants who had completed test-retest within 30 days, which was assessed to reflect a time period where no genuine change in dynamic balance would have occurred. The results for the 36 participants weren't skewed as the ICC was 0.76-0.90 compared to 0.79-0.86 presented in this study for the sample of 51 participants. Thus, the ICC in both cases would reflect moderate to good reliability according to the classification by Koo et al. (11). Based on the sample of 36 participants, SEM (4-6%) and MDC (11-16%) was determined with higher uncertainty as opposed to the presented SEM (2-4%) and MDC (5-11%) for the sample of 51 participants due to the smaller sample size when excluding the participants with longer than 30 days between test and retest. Thus, it was assessed reasonable to present the results based on the $n = 51$ sample, due to having sufficient statistical power in the analyses.

Another limitation in this study was the appropriateness of the statistical procedure as the t-tests do not account for variability between raters, as well as the fact that repeated direction measurements are observed on each subject. However, the use of the t-test increases this study's results comparability to other studies as it is common in the literature.

In this study the YBT was only performed on the right leg, which was not in compliance with the YBT-LQ Protocol where the test is performed on the right and the left leg alternately in each test direction to reduce muscular fatigue. The authors of this study hypothesized that muscular fatigue was accumulating and therefore that the results in the last test direction posterolateral would be affected the most. The SEM indicating the measurement error was, however, lower in the posterolateral direction (2%) compared with the other test directions (3-4%). Muscular fatigue did thus not seem to affect the results.

The YBT showed a good relative and absolute interrater, test-retest reliability among raters with no experience in use of the YBT. The YBT is thus considered suitable at both group and individual level in young, healthy, and physically active populations. Future research should determine the reliability of the YBT in clinical or pathologic settings. Furthermore, future research should focus on examining the relationship between the SEBT and YBT techniques.

ACKNOWLEDGMENTS

The Association of Danish Physiotherapists supported the study by funding the two Y-Balance Boards used. The authors report there are no competing interests to declare.

REFERENCES

1. World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *JAMA* 310(20):2191-2194, 2013.
2. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Int J Nurs Stud* 47(8):931-936, 2010.
3. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge University Press, 2011.
4. Faigenbaum AD, Myer GD, Fernandez IP, Carrasco EG, Bates N, Farrell A, Ratamess NA, Kang J. Feasibility and reliability of dynamic postural control measures in children in first through fifth grades. *Int J Sports Phys Ther* 9(2):140-148, 2014.
5. Freund JE, Stetts DM, Oostindie A, Shepherd J, Vallabhajosula S. Lower quarter Y-Balance Test in healthy women 50-79 years old. *J Women Aging* 31(6):475-491, 2019.
6. Fullam K, Caulfield B, Coughlan GF, Delahun E. Kinematic analysis of selected reach directions of the Star Excursion Balance Test compared with the Y-Balance Test. *J Sport Rehabil* 23(1):27-35, 2014.
7. Giraudeau B, Mary JY. Planning a reproducibility study: How many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med* 20(21):3205-3214, 2001.
8. Gonell AC, Romero JAP, Soler LM. Relationship between the Y Balance Test scores and soft tissue injury incidence in a soccer team. *Int J Sports Phys Ther* 10(7):955-966, 2015
9. Greenberg ET, Barle M, Glassmann E, Jung M-K. Interrater and test-retest reliability of the Y Balance Test in healthy, early adolescents female athletes. *Int J Sports Phys Ther* 14(2):204-213, 2019.
10. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 30(1):1-15, 2000.
11. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15(2):155-163, 2016.
12. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 64(1):96-106, 2011.
13. Neumann DA. Kinesiology of the hip: A focus on muscular actions. *J Orthop Sports Phys Ther* 40(2):82-94, 2010.

14. Plisky PJ, Gorman PP, Butler RJ, Kiesel KB, Underwood FB, Elkins B. The reliability of an instrumented device for measuring components of the Star Excursion Balance Test. *N Am J Sports Phys Ther* 4(2):92-99, 2009.
15. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. F.A. Davis Company, 2015
16. Schwartz G, Brueckner D, Schedler S, Kiss R, Muehlbauer T. Performance and reliability of the lower quarter Y Balance Test in healthy adolescents from grade 6 to 11. *Gait Posture* 67:142-146, 2019.
17. Shaffer SW, Teyhen DS, Lorensen CL, Warren RL, Koreerat CM, Straseske CA, Childs JD. Y-Balance Test: A reliability study involving multiple raters. *Mil Med* 178(11):1264-1270, 2013.
18. Sipe CL, Ramey KD, Plisky PP, Taylor JD. Y-Balance Test: A valid and reliable assessment in older adults. *J Aging Phys Act* 27(5):663–669, 2019.
19. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Qual Life Res* 21(4):651-657, 2012.
20. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 19(1):231, 2005.
21. Wilson BR, Robertson KE, Burnham JM, Yonz MC, Ireland ML, Noehren B. The relationship between hip strength and the Y Balance Test. *J Sport Rehabil* 27(5):1-450, 2018.
22. Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. *J Arthroplasty* 13(8):890-895, 1998.

