

# THE RELIABILITY OF THE STAR EXCURSION BALANCE TEST AND LOWER QUARTER Y-BALANCE TEST IN HEALTHY ADULTS: A SYSTEMATIC REVIEW

Cameron J. Powden, PhD, LAT, ATC<sup>1</sup>

Teralyn K. Dodds, DAT, LAT, ATC<sup>2</sup>

Emily H. Gabriel, PhD, LAT, ATC<sup>3</sup>

## ABSTRACT

**Background:** Dynamic balance is often an important criterion used during lower extremity musculoskeletal injury prediction, prevention, and rehabilitation processes. Methods to assess lower extremity dynamic balance include the Star Excursion Balance Test (SEBT) and Lower Quarter Y-Balance Test (YBT). Due to the importance of dynamic balance it is imperative to establish reliable quantification techniques.

**Purpose:** To conduct a systematic review to assess the reliability and responsiveness of the SEBT/YBT.

**Study Design:** Systematic Review.

**Methods:** Electronic databases (PubMed, MEDLINE, CINAHL, and SPORTDiscus) were searched from inception to August 2018. Included studies examined the intra- and inter-rater reliability of the SEBT/YBT in healthy adults. Two investigators independently assessed methodological quality, level of evidence and strength of recommendation with the Quality Appraisal of Reliability Studies (QAREL) scale. Relative intra and inter-rater reliability was examined through intraclass correlation coefficients (ICC) and responsiveness was evaluated through minimal detectable change (MDC). Data was analyzed based on reach direction (Anterior, Posteromedial, and Posterolateral) and normalization (normalized and non-normalized). Additionally, data were then synthesized using the strength of recommendation taxonomy to provide a grade of recommendation.

**Results:** A total of nine studies were included in this review. Six studies examined the inter-rater reliability and seven assessed intra-rater reliability. The included studies had a median QAREL score of 66.89% (range = 55.56% to 75.00%) and 59.03% (range = 33.33 to 66.67%) for inter and intra-rater reliability respectively. Median ICC values for inter-rater reliability were 0.88 (Range = 0.83 – 0.96), 0.87 (range = 0.80 – 1.00), and 0.88 (range = 0.73 – 1.00) for the anterior, posteromedial, and posterolateral directions respectively. Median ICC values for intra-rater reliability were 0.88 (Range = 0.84 – 0.93), 0.88 (Range = 0.85 – 0.94), and 0.90 (Range = 0.68 – 0.94) for the anterior, posteromedial, and posterolateral directions, respectively.

**Conclusions:** There is grade A evidence to support that the SEBT/YBT have excellent inter and intra-rater reliability when used in healthy adults. Furthermore, minimal detectable change values have been provided that can be used in practice to aid clinical decision making. Future research is needed to assess the reliability, responsiveness, and validity of the SEBT/YBT in pathologic populations.

**Level of Evidence:** 1a

**Keywords:** Dynamic balance, intra-rater reliability, inter-rater reliability, test-retest reliability, movement system

## CORRESPONDING AUTHOR

Cameron J. Powden, PhD, LAT, ATC

Department of Athletic Training

University of Indianapolis

1400 Hanna Ave

Indianapolis, IN, 46227

Phone: (317) 788-2516

E-mail: powdenc@uindy.edu

<sup>1</sup> Department of Athletic Training, University of Indianapolis, Indianapolis, IN, USA

<sup>2</sup> Department of Applied Medicine and Rehabilitation, Indiana State University, Terra Haute, IN, USA

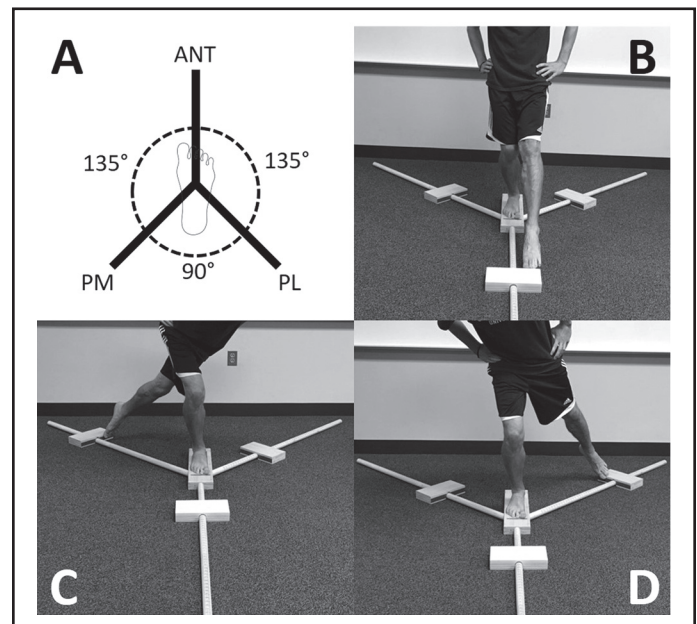
<sup>3</sup> Department of Athletic Training, Mercer University, Macon, GA, USA

**Conflict of Interest Statement:** The authors have no conflicts of interest to disclose.

## INTRODUCTION

It is estimated that approximately three to five million sport related injuries occur each year and primarily occur within the lower extremity.<sup>1,2</sup> These injuries result in significant time loss, medical costs, and often long term consequences such as an increased risk of osteoarthritis when joint trauma has occurred.<sup>3</sup> Due to the prevalence and burden of lower extremity injuries, it is imperative to develop screening tools to identify those at risk of injury and implement proper preventative interventions. Effective injury screening tools and subsequent preventative strategies can ultimately reduce the incidence of injuries, time loss from participation, and healthcare costs associated with the short and long-term treatment of these injuries.<sup>4,5</sup> Dynamic balance is thought to be essential for those participating in physical activity.<sup>6,7</sup> Therefore, deficits in balance have been widely investigated as a predictor of lower extremity injury.<sup>6-8</sup> Furthermore, dynamic balance is regularly used during the rehabilitation process to track progress and make return to play decisions.<sup>8,9</sup> The established clinical importance of dynamic balance for injury prediction, prevention, and decision making necessitates the establishment of reliable measurement tools.

The Star Excursion Balance Test (SEBT) and lower quarter Y-Balance Test (YBT) are two of the most prominent tools in the literature to measure dynamic balance of the lower extremity.<sup>10</sup> The SEBT began as a star comprised of four lines, all crossing at the same center point.<sup>11</sup> To complete the test, an individual stands at the center of the star then reaches with the contralateral leg as far as possible along one of the eight reach directions, while maintaining single leg squat stance.<sup>11</sup> The distance reached is measured in centimeters and typically normalized to the participant's height or leg length to quantify dynamic balance; however it can also be completed without leg length normalization.<sup>11</sup> In its current form, the SEBT has been reduced to three directions due to redundancy (Figure 1).<sup>12</sup> Additionally, an instrumented version, the YBT, was created with the intention to improve the reliability and uniformity of administration of the test.<sup>13</sup> Similar to the SEBT, the YBT consists of three reach directions (anterior, posteromedial, posterolateral) which require participants to



**Figure 1.** Star-Excursion Balance Test and Y-Balance Test Examples, A) Setup, B) Anterior Reach (ANT), C) Posterolateral Reach (PL), D) Posteromedial Reach (PM).

move in similar patterns (Figure 1). Although the movements necessary for both tests are similar, research has indicated the anterior reach distances have been different when comparing the two tests.<sup>13</sup> Therefore, the two instruments may not be directly comparable. Clinicians and researchers commonly use the SEBT and YBT to assess dynamic balance, track changes in performance after the introduction of an injury prevention or rehabilitation program, and to identify those that may be at a heightened risk of injury.<sup>8,14,15</sup>

With the SEBT and YBT commonly being used to assess dynamic balance, ensuring consistency is essential for clinical decision-making. Currently, there have been a variety of SEBT and YBT methodologies evaluated in the literature to assess the reliability of these tests.<sup>16-22</sup> Although there are several reliability studies, it is challenging to draw conclusions from the literature because of the varied assessment techniques and lack of evidence consolidation. Therefore, the purpose of this systematic review was to collect, critically appraise, and synthesize the published evidence describing the inter-rater and intra-rater reliability of the SEBT and YBT to measure dynamic balance in healthy adults.

## METHODS

### Search Strategy

The PRISMA guidelines were followed to conduct a systematic search of the literature to identify studies assessing intra-rater and/or inter-rater reliability of the SEBT and YBT as well as report those findings.<sup>23</sup> Electronic databases were searched using combinations of key words related to the SEBT and/or YBT and reliability (Table 1). Boolean operators “OR” and “AND” were employed to combine search terms.

The Boolean phrase and systematic search was derived and completed by investigators (CJP, TKD). The databases PubMed and EBSCO Host (CINHAL, MEDLINE, SportDiscus) were searched from inception through August 2018. Furthermore, the search was limited to full-text manuscripts written in English, which used human participants.

### Eligibility Criteria

Investigators (CJP, TKD) reviewed identified studies against eligibility criteria. Studies were screened for eligibility based on the criteria below. Initially, potential eligibility was determined by titles and abstracts. In cases in which eligibility was uncertain, the full text of the manuscript was reviewed for inclusion.

### Inclusion Criteria

The following inclusion criteria were used to select and screen studies for inclusion:

- Study purpose: Studies were included if the primary aim was to evaluate the intra-rater and or inter-rater reliability of the SEBT and or YBT.
- Type of participants: Studies on adult ( $\geq 18$  years of age) human participants were included. No restrictions were made in regards to health status of the participants.
- Type of outcome measures: SEBT and or YBT. Composite scores in the directions of Anterior (Ant), Posteromedial (PM), and Posterolateral (PL) were included for the review.
- Only peer reviewed, full text studies were included for the review.

### Exclusion Criteria

The following exclusion criteria were used to screen studies for inclusion:

- Studies that did not evaluate reliability using intra-class correlation coefficients (ICCs) or provide the data needed to calculate this statistic.
- Studies which included participants that were under the age of eighteen.
- Studies not published in English.
- Studies that did not use the SEBT or YBT to assess dynamic balance in the lower extremity

Step	Search Terms	Boolean Operator	EBSCO Host	PubMed
1	Y-Balance SEBT	OR	563	453
2	Star -Excursion Balance Test Reliability Consistency Agreement Inter-rater Intra-rater Accuracy Reproducibility Repeatability	OR	1,472,020	999,230
3	1,2	AND	86	35
Duplicates				30*
Total Identified				91
* Total number of duplicates between EBSCO and PubMed.				

---

## Data Extraction

Two reviewers (CJP, TKD) extracted data during the primary review. The extracted data included: study design, aims, population demographics, clinician demographics, methodology of the SEBT and YBT, reliability outcomes, statistical evaluations, and limitations. Discussion and the use of another independent reviewer (EHG) was used to resolve any discrepancies in interpretation if needed.

## Assessing Quality of Studies

The methodological quality of the included studies was assessed using the Quality Appraisal of Reliability Studies (QAREL) scale. The QAREL is specifically designed for reliability studies and evaluates statistical methods as well as internal and external validity.<sup>24,25</sup> The QAREL consists of an 11-item checklist, all 11 items are weighted equally and scored using *Yes*, *No*, *Unclear*, or *N/A* in accordance with scoring guidelines.<sup>24,25</sup> Included studies were considered to be of high quality if  $\geq 60\%$  of the checklist items were assigned as *Yes*.<sup>24,25</sup> Initially, two reviewers (CJP, TKD) scored the selected studies independently. Reviewers then met to develop a consensus for each study. Any disagreements that could not be brought to consensus through discussions were resolved using a third reviewer (EHG). Percent agreement was calculated to determine the initial agreement between the reviewers for each QAREL item.

## Data Analysis and Synthesis

Inter and intra-rater reliability of the SEBT and YBT were assessed using separate analyses of the anterior, posteromedial, and posterolateral directions. Studies could be included in both or one analysis based on the data presented. Relative reliability was evaluated through calculated or reported Interclass Correlation Coefficients (ICC) for both the inter-rater and intra-rater analyses. The confidence interval at 95% around the ICC was included when reported. Interclass Correlation Coefficients were interpreted in this manner: poor = 0.00-0.25, fair = 0.26-0.50, moderate = 0.51-0.75, and good = 0.76-1.00 reliability.<sup>26</sup> Standard error of measurement (SEM), a measure of measurement dispersion around a "true" score,<sup>26</sup> was used to examine absolute reliability. In cases where the SEM was not reported, it was calculated using the standard deviation and the square root of one minus

the ICC if the required data was reported ( $SEM = SD \cdot \sqrt{1-ICC}$ ).<sup>26</sup> Furthermore, minimal detectable change (MDC) was used to determine the amount of change needed to exceed measurement error at the 95% confidence level. In instances where the MDC was not reported or it was reported at a level of confidence other than 95%, the investigators calculated it if possible ( $MDC = SEM \cdot \sqrt{2} \cdot 1.96$ ). Descriptive analysis, through the use of mean, median, standard error, minimum, maximum, and z-skewness, were used to synthesize the ICC and MDC values from included studies (SPSS software, version 32.0; IBM Corporation, Armonk, NY).

## Level of Evidence and Grade of Recommendation

Data were then synthesized using the strength of recommendation taxonomy (SORT). The SORT method allows for the assessment of individual study level of evidence and a grade of recommendation based on quality, quantity, and consistency of the body of literature.<sup>27</sup> Individual studies were categorized into Level 1, 2, and 3 evidence based on the quality of the study.<sup>27</sup> An A-level recommendation is determined based on good quality patient-oriented evidence.<sup>27</sup> A B-level recommendation is based on limited-quality patient-oriented evidence.<sup>27</sup> A C-level recommendation is determined based on consensus, usual practice, opinion, and disease-oriented evidence.<sup>27</sup>

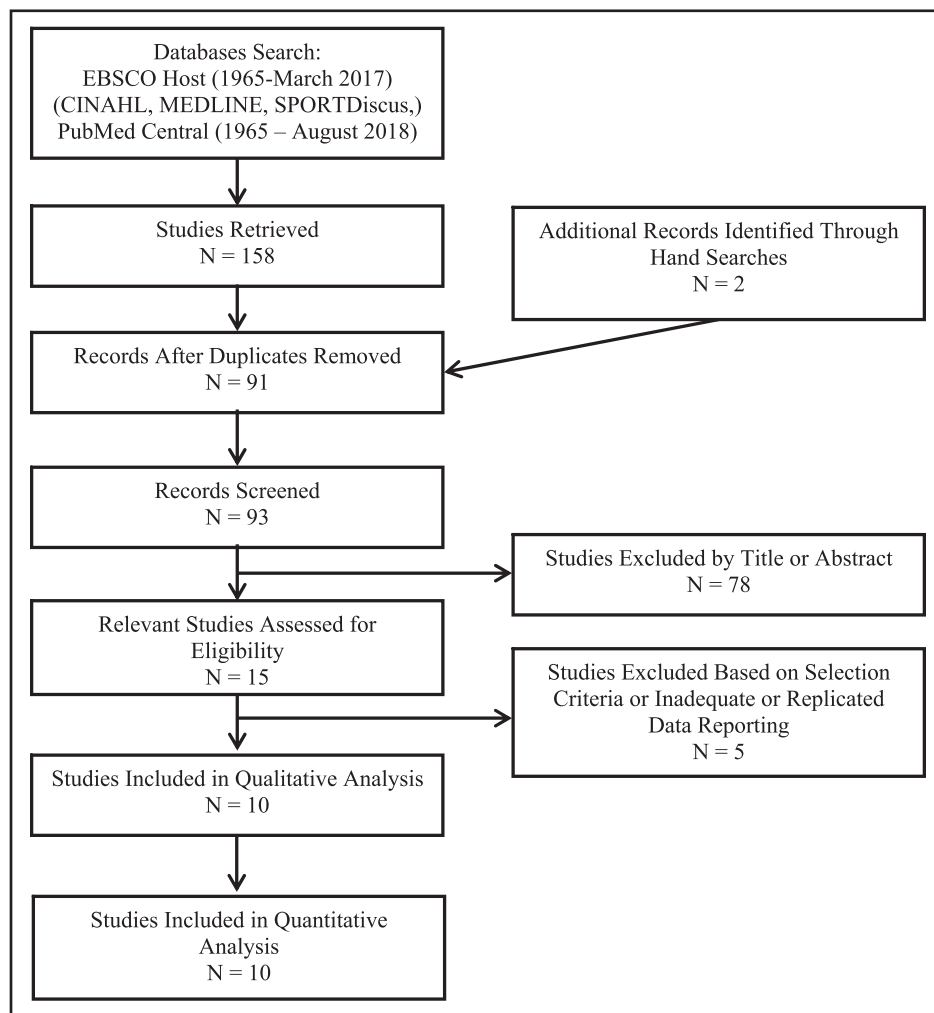
## Sensitivity Analysis

The effect of quality criteria on the assumptions of level of evidence for high quality studies ( $\geq 60\%$ ) was tested by subjecting the criteria to changes of  $\pm 10\%$  and determining the subsequent level of evidence change. Separate sensitivity analyses were conducted for intra- and inter-rater reliability for each direction (anterior, posteromedial, and posterolateral).

## RESULTS

### Literature Search

Figure 2 displays a diagram outlining the results of the search and study review process. A total of 93 studies were retrieved from electronic and hand searches. Of those, nine studies were identified as meeting the selection criteria. Eight studies were identified through electronic search.<sup>16-20,22,28,29</sup> One



**Figure 2.** Flow Chart of Literature Review.

study was identified through hand search.<sup>2</sup> Six studies examined inter-rater reliability.<sup>16,19-22,28</sup> Seven studies examined intra-rater reliability.<sup>19-22,24,30,31</sup>

### General Characteristics

The study characteristics of included studies are displayed in Table 2. All studies included participants free of lower extremity injury. Seven studies<sup>16-18,21,22,28,29</sup> quantified the SEBT using a tape measure and measurements in centimeters. Two studies<sup>19,20</sup> quantified the SEBT using the YBT instrumented kit and measured in centimeters.

*Inter-rater Studies.* Excellent inter-rater reliability was demonstrated within each investigation regardless of use of SEBT or YBT or quantification technique (normalized or non-normalized). Table 5 and 6 illustrate the mean, median, minimum, maximum,

standard error, and z-skewness of ICC and MDC values for both quantification techniques and each reach direction. Anterior reach inter-rater reliability had a median ICC of 0.88 (Range = 0.83 – 0.96). Posteromedial reach inter-rater reliability had a median ICC of 0.87 (range = 0.80 – 1.00). Posterolateral reach inter-rater reliability had a median ICC of 0.88 (range = 0.73 – 1.00).

*Intra-rater Studies.* Excellent intra-rater reliability was demonstrated within each investigation regardless of use of SEBT or YBT or quantification technique (normalized or non-normalized). Table 4 illustrates the mean, median, minimum, maximum standard error, and z-skewness of ICC and MDC values for both quantification techniques and each reach direction. Anterior reach intra-rater reliability had an overall median ICC of 0.88 (Range = 0.84



**Table 2. Characteristics of Included Studies.**

Author	Participants	Measurement Technique	Key Test Parameters	Directions	Number of Trials	Normalization	Reliability Design	Sample Size
Kinzey et al. 1998	9M, 11F 18-35y	SEBT (cm)	Reaching Leg: No Touch Down Stance Leg: Not Mentioned Hands: On hips	A, PM, PL	5 C	Non- Norm	Intra-rater (2 sessions, 1 week apart)	20 (healthy)
Hertel et al. 2000	8M, 8F 21.3±1.3y 171.2±6.7cm 70.3±10.0kg	SEBT (cm)	Reaching Leg: Light touch Stance Leg: Not Mentioned Hands: Not Mentioned	A, PM, PL	1 P/ 3 C	Non-Norm	Intra-rater (2 sessions, 1 week apart)	16 (healthy)
Pilsky et al. 2009	15M 19.7±0.81y	YBT (cm)	Reaching Leg: Not Mentioned Stance Leg: Not Mentioned Hands: Not Mentioned	A, PM, PL	6 P/3 C	Non-Norm	Inter-rater Intra-rater	15 (healthy)
Munro et al. 2010	11M, 11F 22.3±3.7y 167.7±6.2cm 79.6±10.0kg	SEBT (cm)	Reaching Leg: Light Touch Stance Leg: No Heel Lift Hands: On Hips	A, PM, PL	7 C	Non-Norm Norm leg length	Inter-rater Intra-rater (3 sessions, 1 week apart)	22 (healthy)
Gribble et al. 2013	10M, 19W 31.72±10.8y 169.52±8.8cm 65.58±12.3kg	SEBT (cm)	Reaching Leg: Not Mentioned Stance Leg: Not Mentioned Hands: On Hips	A, PM, PL	4 P/ 3 C	Non-Norm Norm leg length	Inter-rater Intra-rater (1 session)	29 (healthy)
Shaffer et al. 2013	53M, 11F 25.2±3.8y 175.5±9.6cm 77.5±12.5kg	YBT (cm)	Reaching Leg: Not Mentioned Stance Leg: Not Mentioned Hands: Not Mentioned	A, PM, PL	6 P/ 3 C	Norm leg length	Inter-rater (1 session)	64 (healthy)
van Lieshout et al. 2016	21M, 34F 24.0±2.9y	SEBT (cm)	Reaching Leg: Not Mentioned Stance Leg: Not Mentioned Hands: On Hips	A, PM, PL	3 C	Norm leg length	Inter-rater (1 session)	50 (healthy)
Lopez-Plaza et al. 2018	27M, 24.5±3.1y 176.8±6.0cm 75.3±10.0kg	SEBT (cm)	Reaching Leg: Light Touch Stance Leg: Not Mentioned Hands: On Hips	PM, PL	?P 3 C	Norm leg length	Intra-rater (2 sessions, 1 month apart)	27 (healthy)
Hyong et al. 2014	18M, 67F 20.7±0.9y 165.6±5.5cm 58.3±8.3kg	SEBT (cm)	Reaching Leg: Not Mentioned Stance Leg: Not Mentioned Hands: On Hips	A, PM, PL	6 P/3 C	Norm leg length	Inter-rater Intra-rater (1 session)	67 (healthy)

M=Male, F=Female, Y=Years of age, SEBT=Star Excursion Balance Test, YBT=Y- balance Test, A=Anterior, PM=Posteromedial, PL=Posterolateral, C=Collection Trials, P=Practice Trials, Non-norm=Not Normalized to Leg Length, Norm Leg=Normalized to Leg Length, ?=Unspecified practice trials

- 0.93). Posteromedial reach intra-rater reliability had an overall median ICC of 0.88 (Range = 0.85 - 0.94). Posterolateral reach intra-rater reliability had an overall median ICC of 0.90 (Range = 0.68 - 0.94).

### Methodological Quality

The two reviewers (CJP, TKD) agreed on 109/110 (99%) items on the QAREL checklist. The one difference in QAREL score was resolved by discussion between the reviewers.

*Inter-rater Studies.* There were a total of four high quality studies<sup>19-22</sup> and two low quality studies.<sup>16,28</sup> Quality appraisal tool for studies of diagnostic reliability scores for the inter-rater reliability studies ranged from 55.56% to 75.00% with a median of 66.69%. The internal validity portion of the scale ranged from 0% to 50% with a median of 40%.

Primarily, the included studies suffered from a lack of blinding of the raters to their own findings, others, and additional cues as well as a lack of testing order variation. The external validity portion of the scale ranged from 66.67% to 100.00% with a median of 100.00%.

*Intra-rater Studies.* There were a total of three high quality studies<sup>18,19,21</sup> and four low quality studies.<sup>16,17,28,29</sup> Quality appraisal tool for studies of diagnostic reliability scores for the intra-rater reliability studies ranged from 33.33% to 66.67% with a median of 59.03%. Primarily, the included studies suffered from a lack of blinding of the raters to their own findings, others, and additional cues as well as a lack of testing order variation. The internal validity portion of the scale ranged from 0% to 50% with a median of 40%. The external validity portion of the scale ranged from 66.67% to 100% with a median of 66.67%.

**Table 3. Quality Analysis of Reliability Studies Using the Quality Appraisal of Reliability Studies Tool.**

Questions	Gribble et al. (2013)	Hertel et al. (2000)	Hyong et al. (2014)	Kinzey et al. (1998)	van Lieshout et al. (2016)	Lopez-Plaza et al. (2018)	Munro et al. (2010)	Plisky et al. (2009)	Shaffer et al. (2013)
1. Representative sample	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Representative raters	Yes	Unclear	Unclear	Unclear	Yes	Unclear	Unclear	Yes	Yes
3. Blinding (other raters)	N/A	Unclear	Unclear	N/A	Yes	N/A	N/A	Yes	Yes
4. Blinding (own findings)	Unclear	Unclear	Unclear	Unclear	No	Unclear	Unclear	No	N/A
5. Blinding (reference/disease)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6. Blinding (clinical information)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7. Blinding (additional cues)	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
8. Examination order varied	Yes	Yes	No	Unclear	No	No	Yes	No	No
9. Appropriate time interval	Yes	Yes	Unclear	Yes	Yes	Unknown	Yes	Yes	Yes
10. Test appropriate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
11. Appropriate statistics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Internal validity (%) (Q: 3-9)	2/4 = 50.00	2/5 = 40.00	0/5 = 00.00	1/4 = 25.00	2/5 = 40.00	0/4 = 0.00	2/4 = 50.00	2/5 = 40.00	2/4 = 50.00
External validity (%) (Q: 1-2, 10)	3/3 = 100	2/3 = 66.67	2/3 = 66.67	2/3 = 66.67	3/3 = 100	2/3 = 66.67	2/3 = 66.67	3/3 = 100	3/3 = 100
Percentage of Yes (%)	6/8 = 75.00	5/9 = 55.56	3/9 = 33.33	4/8 = 50.00	6/9 = 66.67	3/8 = 37.50	5/8 = 62.50	6/9 = 66.67	6/8 = 75.00
Level of Evidence	2	2	2	2	2	2	2	2	2

**Table 4. Pooled Intraclass Correlation Coefficients (ICC) and Minimal Detectable Change (MDC).**

			Mean	Median	Standard Error	Minimum	Maximum	Z-Skewness
<i>Intra-rater</i>								
Anterior	ICC	Overall	0.89	0.88	0.01	0.84	0.93	0.04
	MDC	Normalized (%)	5.87	5.66	0.26	5.57	6.38	1.34
		Non Normalized (cm)	6.37	6.89	1.68	3.25	8.99	-0.63
Posteromedial	ICC	Overall	0.89	0.88	0.01	0.85	0.94	0.79
	MDC	Normalized (%)	7.84	7.44	0.91	6.10	10.36	1.12
		Non Normalized (cm)	7.12	7.42	1.16	4.11	9.55	0.63
Posterolateral	ICC	Overall	0.87	0.90	0.03	0.68	0.94	-2.75
	MDC	Normalized (%)	7.55	7.63	0.64	6.40	8.62	-0.26
		Non Normalized (cm)	8.76	8.13	1.11	7.02	11.75	1.02
<i>Inter-rater</i>								
Anterior	ICC	Overall	0.89	0.89	0.02	0.83	0.96	0.06
	MDC	Normalized (%)	6.70	5.50	1.78	4.40	10.20	1.22
		Non Normalized (cm)	5.68	5.68	3.74	1.94	9.42	-
Posteromedial	ICC	Overall	0.88	0.88	0.02	0.80	1.00	1.10
	MDC	Normalized (%)	8.84	8.60	0.85	7.50	10.41	0.58
		Non Normalized (cm)	5.10	5.10	3.08	2.02	8.18	-
Posterolateral	ICC	Overall	0.88	0.88	0.03	0.73	1.00	-0.61
	MDC	Normalized (%)	10.72	10.65	0.61	9.70	11.82	0.25
		Non Normalized (cm)	6.80	6.80	4.16	2.64	10.95	-

### Level of Evidence

*Inter-rater reliability.* The results of this review indicate that there is Grade A evidence to support excellent inter-rater reliability of the SEBT/YBT. This recommendation is based on consistent findings from four high quality studies<sup>19-22</sup> and two low quality studies<sup>16,28</sup> that are all level 2 investigations.

*Intra-rater reliability.* The results of this review indicate that there is Grade A evidence to support excellent intra-rater reliability of the SEBT/YBT. This recommendation is based on consistent findings from three high quality studies<sup>18,19,21</sup> and four low quality studies<sup>16,17,28,29</sup> that are all level 2 investigations.

**Table 5. Intra-Rater Reliability Statistics and Minimal Detectable Change.**

Author	Measurement Technique	Normalization	Test Information	Relative Reliability (ICC)	Absolute Reliability (SEM)	MDC	Classification of ICC
Hertel et al. (2000)	Tape Measure (cm)	Non-Norm	Anterior	0.93	0.23	6.38cm	Good
Hyong et al. (2014)	Tape Measure (cm)	Nom Leg Length	Anterior	0.88	3.24	8.99%	Good
van Lieshout et al. (2016)	Tape Measure (cm)	Norm Leg Length	Anterior	0.90	1.20	3.25%	Good
Munro (2010)	Tape Measure (cm)	Non-Norm	Anterior	0.88	2.04	5.66 m	Good
Munro (2010)	Tape Measure (cm)	Nom Leg Length	Anterior	0.84	2.48	6.89%	Good
Plisky et al. (2009)	Y-balance Kit (cm)	Non-Norm	Anterior	0.88	2.01	5.57cm	Good
Hertel et al. (2000)	Tape Measure (cm)	Non-Norm	Posteromedial	0.93	2.20	6.10cm	Good
Hyong et al. (2014)	Tape Measure (cm)	Nom Leg Length	Posteromedial	0.94	2.41	6.68%	Good
Kinzey et al. (1998)	Tape Measure (cm)	Non-Norm	Posteromedial	0.85	3.74	10.36cm	Good
van Lieshout et al. (2016)	Tape Measure (cm)	Nom Leg Length	Posteromedial	0.89	3.45	9.55%	Good
Lopez-Plaza et al. (2018)	Tape Measure (cm)	Norm Leg Length	Posteromedial	0.86	2.74	4.11%	Good
Munro (2010)	Tape Measure (cm)	Non-Norm	Posteromedial	0.90	2.54	7.04cm	Good
Munro (2010)	Tape Measure (cm)	Nom Leg Length	Posteromedial	0.86	2.94	8.15%	Good
Plisky et al. (2009)	Y-balance Kit (cm)	Non-Norm	Posteromedial	0.85	2.83	7.84cm	Good
Hertel et al. (2000)	Tape Measure (cm)	Non-Norm	Posterolateral	0.90	2.75	7.63cm	Good
Hyong et al. (2014)	Tape Measure (cm)	Nom Leg Length	Posterolateral	0.93	3.30	9.15%	Good
van Lieshout et al. (2016)	Tape Measure (cm)	Nom Leg Length	Posterolateral	0.89	4.25	11.75%	Good
Lopez-Plaza et al. (2018)	Tape Measure (cm)	Norm Leg Length	Posterolateral	0.68	4.68	7.02%	Moderate
Munro (2010)	Tape Measure (cm)	Non-Norm	Posterolateral	0.94	2.31	6.40cm	Good
Munro (2010)	Tape Measure (cm)	Nom Leg Length	Posterolateral	0.92	2.62	7.11%	Good
Plisky et al. (2009)	Y-balance Kit (cm)	Non-Norm	Posterolateral	0.86	3.11	8.62cm	Good

ICC = Intraclass Correlation Coefficient, SEM = Standard Error of Measurement, MDC= Minimum Detectable Change, Non- Norm= Non- Normalized Leg Length, Nom Leg Length= Normalized Leg Length

### Sensitivity Analysis

Changing the quality criterion for determining high or low quality studies by  $\pm 10\%$  did not affect the recommendation for inter-rater reliability. The recommendation for intra-rater reliability would change to a B if the criterion were increased by 10% because the authors' recommendation would be based upon the findings from one high quality study<sup>21</sup> and six low quality studies.<sup>8,16-18,28,29</sup> This indicates that the current available evidence is generally high quality and that the findings of this review are not likely biased by lower quality evidence.

## DISCUSSION

### Summary of Results

The purpose of this systematic review was to determine the inter- and intra-rater reliability of the SEBT/YBT. The results demonstrate that there is Grade A evidence indicating excellent inter- and

intra-rater reliability of the SEBT/YBT. This recommendation does not change when using normalized or non-normalized quantification techniques as well as when evaluating each reach direction. The findings demonstrate that the SEBT/YBT can be used consistently between one or more clinicians as well as over time. Additionally, summated MDC scores are provided that can be used to help guide clinical decisions by enhancing the determination of when patient change has occurred that exceed the error associated with the test. Furthermore, the results of the sensitivity analysis demonstrate that primarily high level of evidence supports the reliability and usefulness of the SEBT/YBT.

### Methodological Considerations

Included studies assessed the reliability of the SEBT/YBT in healthy populations. Primarily, healthy adults with a mean age range from 19 to 31 years old were



**Table 6. Inter-Rater Reliability Statistics and Minimal Detectable Change.**

Author	Measurement Technique	Normalization	Test Information	Relative Reliability (ICC)	Absolute Reliability (SEM)	MDC	Classification of ICC
Gribble et al. (2013)	Tape Measure (cm)	Non-Norm	Anterior	0.92	NR	NR	Good
Gribble et al. (2013)	Tape Measure (cm)	Norm Leg Length	Anterior	0.88	NR	NR	Good
Hertel et al. (2000)	Tape Measure (cm)	Non- Norm	Anterior	0.83	3.40	9.42cm	Good
Hyong et al. (2014)	Tape Measure (cm)	Norm Leg Length	Anterior	0.83	3.68	10.20%	Good
van Lieshout et al. (2016)	Tape Measure (cm)	Norm Leg Length	Anterior	0.92	1.60	4.40%	Good
Plisky et al. (2009)	Y-balance Kit (cm)	Non- Norm	Anterior	0.96	0.70	1.94cm	Good
Shaffer et al. (2013)	Y-balance Kit (cm)	Norm Leg Length	Anterior	0.88	2.00	5.50%	Good
Gribble et al. (2013)	Tape Measure (cm)	Non-Norm	Posteromedial	0.80	NR	NR	Good
Gribble et al. (2013)	Tape Measure (cm)	Norm Leg Length	Posteromedial	0.91	NR	NR	Good
Hertel et al. (2000)	Tape Measure (cm)	Non-Norm	Posteromedial	0.85	2.95	8.18cm	Good
Hyong et al. (2014)	Tape Measure (cm)	Norm Leg Length	Posteromedial	0.90	3.76	10.41%	Good
van Lieshout et al. (2016)	Tape Measure (cm)	Norm Leg Length	Posteromedial	0.87	3.10	8.60%	Good
Plisky et al. (2009)	Y-balance Kit (cm)	Non-Norm	Posteromedial	1.00	0.73	2.02cm	Good
Shaffer et al. (2013)	Y-balance Kit (cm)	Norm Leg Length	Posteromedial	0.86	2.70	7.50%	Good
Gribble et al. (2013)	Tape Measure (cm)	Non-Norm	Posterolateral	0.88	NR	NR	Good
Gribble et al. (2013)	Tape Measure (cm)	Norm Leg Length	Posterolateral	0.92	NR	NR	Good
Hertel et al. (2000)	Tape Measure (cm)	Non-Norm	Posterolateral	0.73	3.95	10.95cm	Moderate
Hyong et al. (2014)	Tape Measure (cm)	Norm Leg Length	Posterolateral	0.88	4.26	11.82%	Good
van Lieshout et al. (2016)	Tape Measure (cm)	Norm Leg Length	Posterolateral	0.87	3.80	10.65%	Good
Plisky et al. (2009)	Y-balance Kit (cm)	Non-Norm	Posterolateral	1.00	0.79	2.64cm	Good
Shaffer et al. (2013)	Y-balance Kit (cm)	Norm Leg Length	Posterolateral	0.85	3.50	9.70%	Good

ICC = Intraclass Correlation Coefficient, SEM = Standard Error of Measurement, MDC= Minimum Detectable Change, Non- Norm= Non Normalized Leg Length, Norm Leg Length= Normalized Leg, NR=Not Reported

included. The activity level of the participants included in this review varied. The majority of the participants in this review were general population or recreationally active. One study's participants<sup>19</sup> consisted of recreational collegiate soccer players while another study's participants<sup>20</sup> were individuals actively participating in military training. However, regardless of these variations there were consistent reliability measures demonstrated by the included studies. Based on the characteristic that all participants were healthy, it is unclear how lower extremity pathology may affect SEBT/YBT reliability. Additionally, there is limited evaluation of the SEBT/YBT outside of a physically active collegiate population. Therefore, further evaluations are needed to determine the reliability and utility of the SEBT/YBT in a wide range of populations.

Instrumentation techniques used to conduct the SEBT/YBT varied slightly between studies. The

most common quantification technique used a tape measure attached to the floor.<sup>16-18,21,22,28-30</sup> The other method used quantified the SEBT/YBT using the Y-Balance instrumented kit.<sup>19,20</sup> Other methodological variations that were noted between the studies included normalization of reaching limb, number of practice and test trials, and body positioning during the SEBT/YBT. The reliability of the SEBT/YBT was found to be excellent regardless of the quantification technique that was used. Five studies<sup>16,19,20,22,28,29</sup> allowed the participants to complete between one and six practice trials before completing the trials for collection. Body positioning of the participants while completing the SEBT/YBT varied between the studies. Six studies<sup>16-18,21,22,30</sup> required that the participants' hands must remain on hips while reaches were completed. Two studies<sup>18,30</sup> required the heel of the stance leg to remain flat on the ground while the reaches were completed. Additionally, the

---

positioning of the foot on the tape measure/block during testing varied between studies. The foot was most commonly placed behind or in front of the intersection of the reach directions<sup>16,18-20,28</sup> or so the foot was bisected by the reach directions.<sup>17,21,22,30</sup> When conducting the SEBT/YBT clinicians should allow for four practice trials. Studies in which at least four practice trials were permitted, saw more consistent results in fewer collected trials.<sup>16,19,20,22</sup> Although there were several methodological differences between the studies in terms of quantification technique, normalization, practice trials, and body position, these differences did not appear to affect the reliability of the SEBT/YBT. It is important to note though that if a clinician is using the non-normalized method, results can only be compared within the same patient and not across patients. Thus, it is important for a clinician to be consistent in the methodology used in their practice.

### **Practical Implications**

The results of this review indicate that quantification technique, normalization, practice trials, and body positioning do not appear to affect intra or inter-rater reliability. The results indicate that clinicians can perform the SEBT/YBT using their preferred technique with a high degree of consistency between clinicians and over time. However, clinicians should use the same methodology when attempting to compare scores as the use of different measurement techniques may produce different raw values. Furthermore, normalization to leg length should occur to allow for comparison across patients. In summary, the results indicate that the SEBT/YBT is a reliable tool that can provide comparable results between multiple raters during pre-participation injury screening as well as throughout the rehabilitation process.

The following summary MDCs should be used in clinical practice to determine patient change that exceeds the error associated with the SEBT/YBT. When evaluating changes in normalized reach distances over time MDCs of 5.87%, 7.84%, and 7.55% should be used for anterior, posteromedial, and posterolateral reach directions, respectively. When evaluating changes in non-normalized reach distances MDCs of 6.37cm, 7.12cm, and 8.76cm should

be used for anterior, posteromedial, and posterolateral reach directions, respectively. For example, if a patient's anterior reach increases or decreases 5.87% or more for a normalized reach, the change can be considered true change and potentially clinically meaningful change. The same is true for an increase or decrease in anterior reach of greater than 6.37cm for a non-normalized reach, the change can be considered true change and potentially clinically meaningful change. The ability to determine change that exceeds dynamic balance measurement error can assist a clinician in making prevention, rehabilitation, and return to play decisions. However, it is important to note that the included MDCs are based on healthy participants and that these values may not translate to pathologic populations.

### **Limitations of Review**

This systematic review is not without limitations. Following the inclusion criteria, only healthy adult participants were included in the review. Due to this, five studies<sup>30-34</sup> were excluded due to the participant group including individuals under the age of 18 years. This review did not include any studies in which the participant group was pathologic or injured due to limitations in the literature. By excluding these studies the authors may have unintentionally limited the scope of the reliability of the SEBT/YBT. Future research should investigate the reliability of the SEBT/YBT within participants with pathologic conditions or injuries. Lastly, this investigation was only able to assess ICCs and MDCs of the SEBT/YBT. Important clinical statistics such as minimally clinically important difference should be investigated in future studies.

### **CONCLUSION**

The results of this systematic review demonstrate that there is Grade A evidence to support excellent inter- and intra-rater reliability of the SEBT/YBT. These results infer that the SEBT/YBT should be used clinically to assess dynamic balance and provide consistent and repeatable results between one or more clinicians. Due to all of the included studies assessing dynamic balance in healthy populations, future research should determine the reliability of the SEBT/YBT in a pathologic population.

---

## REFERENCES

1. Dick R, Agel J, Marshall SW. National collegiate athletic association injury surveillance system commentaries: Introduction and methods. *J Athl Train*. 2007;42(2):173-182.
2. Kraus JF, Conroy C. Mortality and morbidity from injuries in sports and recreation. *Annu Rev Public Heal*. 1984;5(1):163-192.
3. Thomas AC, Hubbard-Turner T, Wikstrom EA, Palmieri-Smith RM. Epidemiology of posttraumatic osteoarthritis. *J Athl Train*. 2017;52(6):491-496.
4. Bird SP, Markwick WJ. Musculoskeletal screening and functional testing: considerations for basketball athletes. *Int J Sports Phys Ther*. 2016;11(5):784-802.
5. Marcoux V, Chouinard M-C, Diadiou F, Dufour I, Hudon C. Screening tools to identify patients with complex health needs at risk of high use of health care services: A scoping review. *PloS one*. 2017;12(11):e0188663.
6. Alentorn-Geli E, Myer GD, Silvers HJ, et al. Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 1: Mechanisms of injury and underlying risk factors. *Knee Surg Sport Traum Arthrosc*. 2009;17(7):705-729.
7. Granata KP, Lockhart TE. Dynamic stability differences in fall-prone and healthy adults. *J Electromyogr Kinesiol*. 2008;18(2):172-178.
8. Plisky PJ, Rauh MJ, Kaminski TW, Underwood FB. Star Excursion Balance Test as a predictor of lower extremity injury in high school basketball players. *J Orthop Sports Phys Ther*. 2006;36(12):911-919.
9. Gribble PA, Hertel J, Plisky P. Using the Star Excursion Balance Test to assess dynamic postural-control deficits and outcomes in lower extremity injury: a literature and systematic review. *J Athl Train*. 2012;47(3):339-357.
10. Gribble PA HJ, Plisky P. . Using the Star Excursion Balance Test to asses dynamic postural control defecits and outcomes in lower extremity injury: A literatue review and systematic review. *J Athl Train*. 2012;47(3):339-357.
11. GW. G. Lower Extremity Functional Profile. *Adrian, MI: Wynn Marketing, Inc*. 1995.
12. Hertel J, Braham RA, Hale SA, Olmsted-Kramer LC. Simplifying the star excursion balance test: analyses of subjects with and without chronic ankle instability. *J Orthop Sports Phys Ther*. 2006;36(3):131-137.
13. Coughlan GF, Fullam K, Delahunt E, Gissane C, Caulfield BM. A comparison between performance on selected directions of the star excursion balance test and the Y balance test. *J Athl Train*. 2012;47(4):366-371.
14. Smith CA, Chimera NJ, Warren M. Association of y balance test reach asymmetry and injury in division I athletes. *Med Sci Sport Exer*. 2015;47(1):136-141.
15. Hartley EM, Hoch MC, Boling MC. Y-balance test performance and BMI are associated with ankle sprain injury in collegiate male athletes. *J Sci Med Sport*. 2018;21(7):676-680.
16. Hyong IH, Kim JH. Test of intrarater and interrater reliability for the star excursion balance test. *J Phys Ther Sci*. 2014;26(8):1139-1141.
17. Kinzey SJ, Armstrong CW. The reliability of the star-excursion test in assessing dynamic balance. *J Orthop Sports Phys Ther*. 1998;27(5):356-360.
18. Munro AG, Herrington LC. Between-session reliability of the star excursion balance test. *Phys Ther Sport*. 2010;11(4):128-132.
19. 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*. 2009;4(2):92-99.
20. Shaffer SW, Teyhen DS, Lorenson CL, et al. Y-balance test: a reliability study involving multiple raters. *Mil Med*. 2013;178(11):1264-1270.
21. van Lieshout R, Reijnenveld EA, van den Berg SM, et al. Reproducibility of the modified star excursion balance test composite and specific reach direction scores. *Int J Sports Phys Ther*. 2016;11(3):356-365.
22. Gribble PA, Kelly SE, Refshauge KM, Hiller CE. Interrater reliability of the star excursion balance test. *J Athl Train*. 2013;48(5):621-626.
23. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*. 2009;6(7):e1000097.
24. Lucas N, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol*. 2013;13(1):111.
25. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol*. 2010;63(8):854-861.
26. Portney LG WM. *Foundations of Clinical Research: Applications to Practice (3rd ed.)*. 3rd ed. Upper Saddle River, New Jersey Pearson/Prentice Hall 2009.
27. Ebell MH, Siwek J, Weiss BD, et al. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *J Am Board Fam Pract*. 2004;17(1):59-67.

- 
28. Hertel J, Miller SJ, Denegar CR. Intratester and intertester reliability during the Star Excursion Balance Tests. *J Sport Rehabil.* 2000;9(2):104-116.
  29. López-Plaza D, Juan-Recio C, Barbado D, Ruiz-Pérez I, Vera-García FJ. Reliability of the Star Excursion Balance Test and two new similar protocols to measure trunk postural control. *Phys Med Rehabil.* 2018;10(12):1344-1352
  30. Demura S, Yamada T. Proposal for a practical star excursion balance test using three trials with four directions. *Sport Sci Health.* 2010;6(1):1-8.
  31. Shaikh AA, Walunjkar RN. Reliability of the Star Excursion Balance test (SEBT) in healthy children of 12-16 Years. *Indian J Physiother Occup Ther.* 2014;8(2):29.
  32. Faigenbaum AD, Myer GD, Fernandez IP, et al. Feasibility and reliability of dynamic postural control measures in children in first through fifth grades. *Int J Sports Phys Ther.* 2014;9(2):140-148.
  33. Linek P, Sikora D, Wolny T, Saulicz E. Reliability and number of trials of Y Balance Test in adolescent athletes. *Musculoskelet Sci Pract.* 2017;31:72-75.
  34. Kenny SJ, Palacios-Derflinger L, Owoeye O, Whittaker JL, Emery CA. Between-day reliability of pre-participation screening components in pre-professional ballet and contemporary dancers. *J Dance Med Sci.* 2018;22(1):54-62.