ORIGINAL RESEARCH

REPRODUCIBILITY OF THE MODIFIED STAR EXCURSION BALANCE TEST COMPOSITE AND SPECIFIC REACH DIRECTION SCORES

IJSPT

Remko van Lieshout, PT, MSc[1]
Elja A.E. Reijneveld, PT, MSc[1]
Sandra M. van den Berg, PT, MSc[1]
Gijs M. Haerkens, PT, MSc[1]
Niek H. Koenders, PT, MSc[1]
Arina J. de Leeuw [1]
Roel G. van Oorsouw, PT, MSc[1]
Davy Paap, PT, MSc[1]
Else Scheffer, MSc[1]
Stijn Weterings, PT, MSc[1]
Mirelle J. Stukstette, PT, PhD[1]

ABSTRACT

Background: The mSEBT is a screening tool used to evaluate dynamic balance. Most research investigating measurement properties focused on intrarater reliability and was done in small samples. To know whether the mSEBT is useful to discriminate dynamic balance between persons and to evaluate changes in dynamic balance, more research into intra- and interrater reliability and smallest detectable change (synonymous with minimal detectable change) is needed.

Purpose: To estimate intra- and interrater reliability and smallest detectable change of the mSEBT in adults at risk for ankle sprain.

Study Design: Cross-sectional, test-retest design

Methods: Fifty-five healthy young adults participating in sports at risk for ankle sprain participated (mean ± SD age, 24.0 ± 2.9 years). Each participant performed three test sessions within one hour and was rated by two physical therapists (session 1, rater 1; session 2, rater 2; session 3, rater 1). Participants and raters were blinded for previous measurements. Normalized composite and reach direction scores for the right and left leg were collected. Analysis of variance was used to calculate intraclass correlation coefficient values for intra- and interrater reliability. Smallest detectable change values were calculated based on the standard error of measurement.

Results: Intra- and interrater reliability for both legs was good to excellent (intraclass correlation coefficient ranging from 0.87 to 0.94). The intrarater smallest detectable change for the composite score of the right leg was 7.2% and for the left 6.2%. The interrater smallest detectable change for the composite score of the right leg was 6.9% and for the left 5.0%.

Conclusion: The mSEBT is a reliable measurement instrument to discriminate dynamic balance between persons. Most smallest detectable change values of the mSEBT appear to be large. More research is needed to investigate if the mSEBT is usable for evaluative purposes.

Level of Evidence: Level 2

Keywords: Ankle, Dynamic Balance; mSEBT; Reliability; smallest detectable change

CORRESPONDING AUTHOR

Mirelle J. Stukstette,
Clinical Health Sciences
University Medical Center Utrecht, Utrecht University
P.O. Box 85500, room HB4.05, 3508 GA Utrecht, The Netherlands
Tel: +31887556764, Fax: +31 887 55 34 09
E-mail: m.j.p.m.stukstette@umcutrecht.nl

[1] Physical Therapy Science, Program in Clinical Health Sciences, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

BACKGROUND

Ankle sprains are common in young adults, participating in competitive, court and team sports such as soccer, volleyball, tennis, hockey and running.[1,2] Patients with ankle sprains suffer from pain, limitations in activities and participation restrictions, resulting in high costs for individual sport participants and society.[1,3-5] Considering the impact of ankle sprains, injury prevention is important in sports. Dynamic balance deficits, defined as limitations in the ability to hold a stable base of support while making purposeful movements like reaching,[6] are related to an increased risk for ankle sprain.[7-11] For the development of effective prevention programs, a screening tool to detect and evaluate persons with dynamic balance deficits is needed.

The modified Star Excursion Balance Test (mSEBT) is a screening tool, widely used by physical therapists to detect dynamic balance deficits, and to evaluate dynamic balance improvement in individuals after following a preventive training program.[7,12-15] The mSEBT measures dynamic balance, while a person maintains balance on a single leg and simultaneously reaches as far as possible with the other leg along the reach line of three directions (anterior, posteromedial and posterolateral). Originally composite scores (sum scores of the three directions) of the mSEBT were used as an index for dynamic balance. Recently, there is emerging evidence that specific reach directions of the mSEBT correlate with some specific lower extremity impairments.[7,16-18] Therefore, scores of separate reach directions as well as composite score are used as indices of dynamic balance.

To determine whether the mSEBT is a reliable measurement instrument to evaluate individual dynamic balance performance, insight into measurement properties of the test is necessary. For diagnostic purposes reliability (the degree until which the mSEBT can differentiate balance performances between persons) is an important parameter. For evaluative purposes, information about measurement error and the minimal amount of change above measurement error (smallest detectable change synonymous with minimal detectable change) is needed to know whether changes in performance should be attributed to measurement error or to true changes in performance.[19] In the current study reliability, measurement error

and smallest detectable change (SDC) are referred to under the umbrella term reproducibility. For practitioners and researchers it is important to gain insight in reproducibility of the mSEBT under several conditions. Namely, when the mSEBT is used by different raters (interrater), and by the same raters (intrarater) on different occasions. Intra- and interrater reliability of the mSEBT were investigated in different studies and seem to be good to excellent.[6,20-24] However, commonly identified flaws in these studies were small sample sizes, absence of a clear measurement protocol or missing information regarding measurement error and SDC values. Information about intra- and interrater reproducibility, based on high quality studies is needed to confirm intra- and interrater reliability of the mSEBT and to investigate whether the mSEBT is useful to evaluate changes in performance in persons at risk for ankle sprain over time. Therefore the aim of the current study is to estimate the intra- and interrater reproducibility in healthy young adults participating in sports at risk for ankle sprain.

METHODS

Participant characteristics

A convenience sample of volunteers was recruited in the Netherlands at Utrecht School of Medicine, University Medical Center Utrecht, Utrecht University; and at University of Applied Sciences Utrecht; and from researchers' personal networks. Students and workers from Utrecht School of Medicine and University of Applied Sciences Utrecht were informed about this research and asked to participate via e-mail, and their digital learning environment. Furthermore, friends and colleagues of the researchers were informed about the research and asked to participate. Included were healthy persons between 18 and 30 years old who participated at least once a week in a selection of sports with increased risk for ankle sprain. Based on the number of ankle injuries treated in emergency departments, the following sports were defined as sports with increased risk for ankle sprain: running, soccer, volleyball, futsal, hockey, badminton, tennis, martial arts, handball, squash, gymnastics, korfball, basketball.[25] Excluded were persons who reported lower extremity injury within the past twelve months, cerebral concussion within the previous three months, history of prior ankle surgery, and history of neurological disorders affecting balance.

Participant characteristics such as age, gender, dominant leg (preferred kicking leg), history of ankle sprain, participation in sports with increased risk for ankle sprain and sport frequency were collected. The study was approved by the University Medical Center Utrecht Medical Ethics Committee, METC-protocol number 13-156/C. Informed written consent was obtained from all subjects prior to participation in the study and the rights of the subjects were protected throughout the study.

Rater characteristics

Four physical therapists were randomly chosen from a group of 10 physical therapists that participated in the master program Physical Therapy Science at University Medical Centre Utrecht and were selected by convenience sampling. The chosen physical therapists were trained during two standardization sessions, each session took two hours. During the standardization sessions a specifically developed measurement protocol was discussed and practiced. This protocol also included verbal instructions (test instructions and encouragements) given to the participant.

Measurement procedure and protocol mSEBT

A cross-sectional, test-retest design (rater 1, rater 2, rater 1) was used to examine the intra- and inter-rater reproducibility of the mSEBT (Figure 1). The four raters were rotated in two couples of two raters with one test couple simultaneously measuring two participants (Figure 1). Each participant performed three test sessions by two raters (session 1, rater 1; session 2, rater 2; session 3, rater 1) within one hour. To ensure blinding of raters for previous measurement, forms with the recorded data were collected after each test session (by a research assistant). In order to prevent missing items all data were checked by a research assistant before participants left the test location.
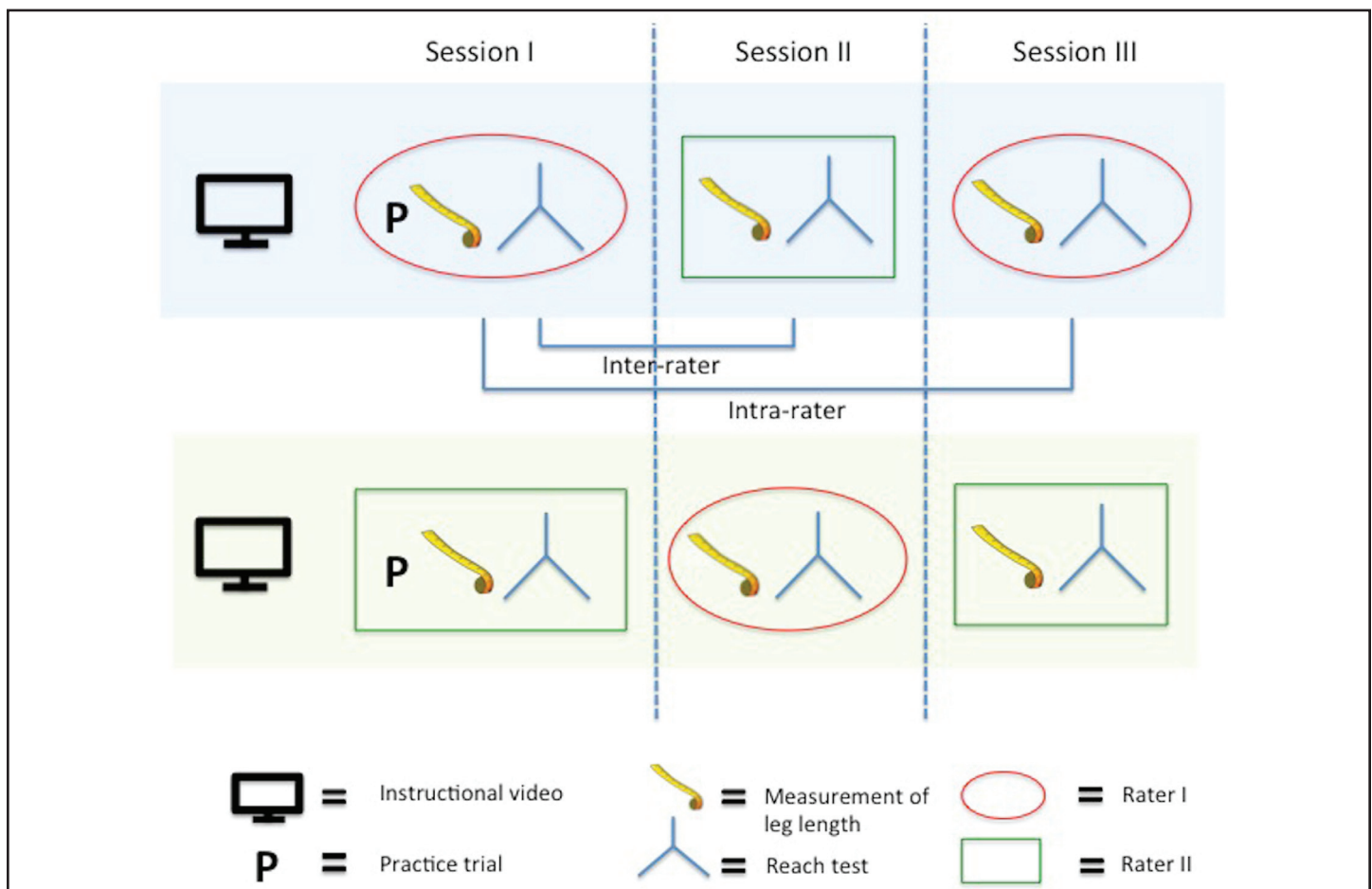
Figure 1. *Study design used to investigate reproducibility of the mSEBT.*
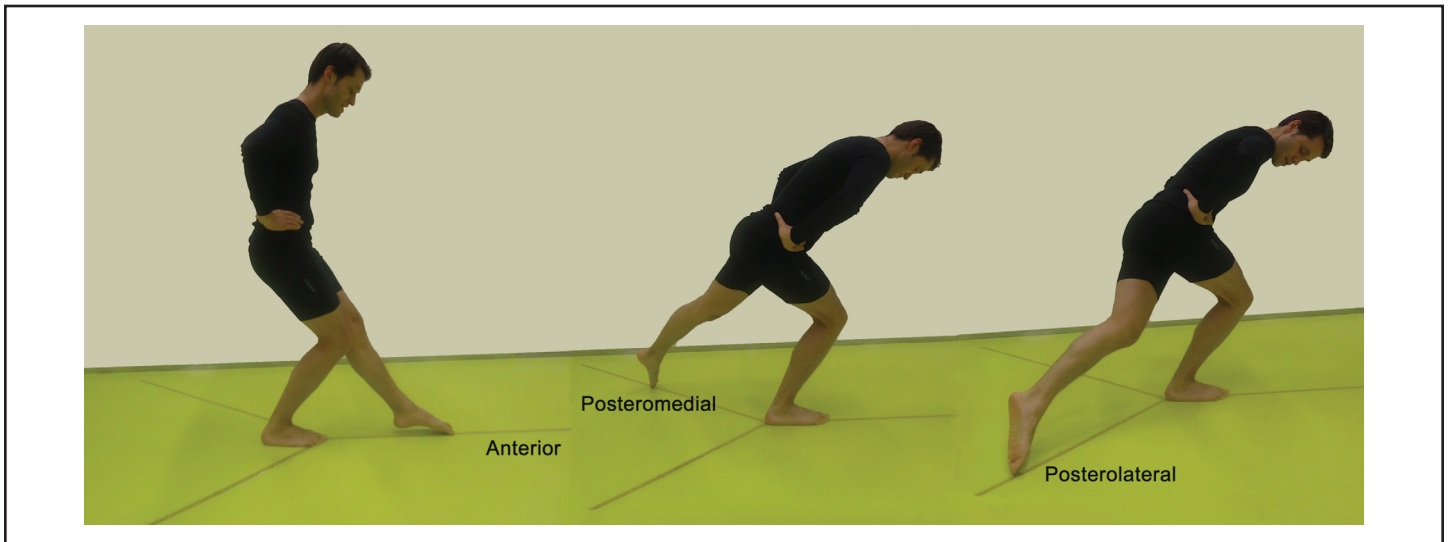
Figure 2. *Participant performing the mSEBT in the three reach directions that are labeled in reference to right stance leg.*

Before testing was started, participants viewed an instructional video, which demonstrated the test and the testing procedure, followed by four practice trials in each of the three reach directions on each leg.[23] Participants performed the mSEBT standing at the center of a grid laid on the floor with three reach lines in the form of an Y. The three reach lines were labeled in relation to the stance leg as anterior (A), posteromedial (PM) and posterolateral (PL) directions, with two angles of 135° (between the A and PM line and between the A and PL line) and one angle of 90° (between the PM and PL line) (Figure 2). The lines were constructed with standard tape measures and transparent tape on the floor. For the anterior reach direction participants were standing with the most distal part of the big toe at the cross of the Y at the beginning of the anterior tape measure. For the posterior reach directions participants were standing with the most posterior part of the heel at the cross of the Y at the beginning of the posterior tape measures,[6] were barefooted and asked to place both hands on the hips. Participants were instructed to reach as far as possible along each of the three reach lines, make a light touch on the line with the most distal part of the big toe and return the reaching leg back to the center while maintaining a single-leg stance with the other leg. Participants performed three trials in each direction on each leg. The test started with the right leg as stance leg followed by the left leg in successively the A, PM

and PL reach directions. Ten seconds of rest were provided between the different trials of one reach direction. A trial was discarded and repeated if participants (1) took weight on the reaching foot; (2) failed to bring back the reaching foot to the starting position without losing control; (3) failed to keep both hands on hips; (4) failed to keep the stance foot at the same place; or (5) failed to keep the forefoot or heel of the stance foot on the floor.[6,23,26] A maximum of six attempts per reach direction was allowed to obtain three valid scores. The three valid scores were averaged and used to calculate normalized mSEBT scores. The rater recorded scores of each trial in each reach direction in centimeters.

Calculating normalized mSEBT scores

To normalize reach distances the participant's leg length was measured from the anterior superior iliac spine to the distal tip of the medial malleolus with the participant lying supine.[26,27] Right and left leg length were measured two times and averaged. A standard tape measure was used to quantify the distance in centimeters.

For each reach direction, the mean out of the scores of three trials was calculated and normalized. Normalization was accomplished by dividing the mean reach distance by the participants (stance) leg length and then multiplying by 100%.[27] In order to calculate a composite score, the mean of the three normalized reach direction scores was calculated.

Data analysis

According to the recommendations of the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group and a sample size calculation as described by Giraudeau and Mary, at least 50 participants should be included in the study.[28,29] The COSMIN initiative aims to improve the selection of health measurement instruments. As part of this initiative, the COSMIN group developed a critical appraisal tool (a checklist) containing standards for evaluating the methodological quality of studies on the measurement properties of health measurement instruments (http://www.cosmin.nl).

Descriptive statistics including frequencies, percentages, means, and standard deviations were used to describe the research population. Measurement properties (intra- and interrater reliability, measurement error and SDC values) were estimated according to the recommended methods of the COSMIN group.[30]

Reliability

Reliability was assessed with the intraclass correlation coefficient (ICC). Analysis of variance was used to calculate $ICC_{consistency}$ (model 3.1) for intrarater reliability, and $ICC_{agreement}$ (model 2.1) for interrater reliability.[19,31,32] Interpretation of ICC scores was as following: Poor reliability ICC < 0.40, Fair reliability ICC 0.40 - 0.70, Good reliability ICC 0.70 – 0.90 and Excellent reliability ICC > 0.90.[33]

Measurement error

Measurement error (intra- and interrater agreement) was investigated calculating Bland and Altman 95% limits of agreement (LoA). LoA illustrates the range and magnitude of the differences in measurements between or within raters.[34]

Smallest detectable change

SDC refers to the smallest amount of change, which falls outside the measurement error of the measurement instrument.[19] To calculate SDC values, first intra- and interrater measurement error were expressed in the units of measurement error: standard error of measurement agreement ($SEM_{agreement}$). Analysis of variance was used to calculate $SEM_{agreement}$.[19,31] SDC was based on SEM and calculated using the following formula: SDC = 1.96*2*SEM.[19]

All statistical analyses were performed using SPSS version 20 (SPSS Inc., Chicago, IL). All data entry was double checked by another researcher. In case of missing data, no further actions were taken and available data were analyzed when missing data remains below 5%.

RESULTS

Participant characteristics

Sixty-six adults between the ages of 18 to 30 years signed up to participate. Nine did not met the inclusion criteria and two did not show up. Therefore, 55 participants (62% female) were included in this study, with a mean age of 24.0 (SD 2.9) years; a mean leg length of 90.9 (SD 6.4) cm, and quarter of participants had a history of ankle sprain. Participants performed sports with an increased risk for ankle sprain on average 2.3 (SD 1.2) times a week. See Table 1 for participant characteristics.

Intrarater and interrater reliability and agreement parameters were assessed on 55 pairs of observations by four physical therapists (3 males, working experience ranged from six months to 11 years). Missing data was below 2%. Therefore, no further actions were taken and available data were analyzed.

Reliability

Intrarater reliability for both the right and the left leg was good to excellent (ICC's for the right leg ranged

Table 1. *Participant characteristics (N = 55)*	
Characteristics	
Men / Female, n/n (%/%)	21/34 (38/62)
Age (y), mean (SD)	24.0 (2.9)
Right/left leg length (cm), mean (SD)	90.9 (6.3)/ 90.9 (6.4)
Dominant leg right / dominant leg left, n/n (%/%)	51/4 (92.7/7.3)
History of ankle sprain, n (%)	14 (25.5)
Weekly frequency participation in risk sports, mean (SD)	2.3 (1.2)
Participation in risk sports, n (%)	
Running	36 (65)
Soccer	10 (18)
Volleyball	6 (11)
Futsal	5 (9)
Hockey	3 (5)
Badminton	3 (5)
Tennis	3 (5)
Martial arts	2 (4)
Handball	2 (4)
Squash	2 (4)
Gymnastics	1 (2)
Korfball	1 (2)

from 0.87 (PM and PL) to 0.91 (Composite and A) and for the left leg from 0.89 (A) to 0.93 (Composite)).

Interrater reliability for the right leg was good to excellent (ICC values ranged from 0.87 (PM and PL) to 0.92 (A)). For the left leg excellent interrater reliability was found for all separate direction scores as well as the composite score (ICC values ranged from 0.92 (PL) to 0.94 (Composite)) (See Table 2).

Measurement error
Intrarater LoA for the composite score of the right leg demonstrated a mean difference in normalized scores within raters of -1.4%. For the left leg the mean difference in composite score within raters was -1.6% (See Table 3).

Interrater LoA for the right leg demonstrated a mean difference between raters for the composite scores of -0.3%. For the left leg the mean difference in composite scores between raters was -2.8% (See Table 3).

Smallest detectable change
Intrarater SDC values for the right leg ranged from 4.7% (A) to 12.7% (PL) and the value for the composite score of the right leg was 7.2%. Intrarater SDC values for the left leg ranged from 1.8% (A) to 10.8% (PL) and the SDC value for the composite score of the left leg was 6.2%.

Interrater SDC values for the right leg ranged from 4.4% (A) to 12.3% (PL) and the SDC value for the composite score of the right leg was 6.9%. Interrater

Table 2. *Intra- and interrater reliabilty and agreement for composite and separate reach direction scores of the mSEBT (N = 55)*

Direction	Mean (SD)	Intrarater			Interrater		
		ICC[*] Consistency	SEM Agreement	SDC Agreement	ICC[*] Agreement	SEM Agreement	SDC Agreement
Right leg							
Composite score	72.3 (2.9)	0.91	2.6	7.2	0.91	2.5	6.9
Anterior	65.3 (5.4)	0.91	1.7	4.7	0.92	1.6	4.4
Posteromedial	78.0 (9.6)	0.87	3.9	10.7	0.87	3.7	10.3
Posterolateral	73.6 (11.7)	0.87	4.6	12.7	0.87	4.4	12.3
Left leg							
Composite score	73.0 (7.3)	0.93	2.2	6.2	0.94	1.8	5.0
Anterior	66.0 (5.4)	0.89	0.7	1.8	0.93	1.6	4.4
Posteromedial	80.0 (8.8)	0.90	3.0	8.4	0.93	2.5	6.9
Posterolateral	73.1 (10.8)	0.91	3.9	10.8	0.92	3.2	9.0

[*]All values except ICC are normalized excursion reach distance (reach distance / leg length x 100%).

Table 3. *Mean differences in normalized reach direction scores and 95% limits of agreement within and between raters*

Direction	Mean (SD)	Intrarater			Interrater		
		Mean difference within raters	Upper bound of 95% LoA	Lower bound of 95% LoA	Mean difference between raters	Upper bound of 95% LoA	Lower bound of 95% LoA
Right leg							
Composite score	72.3 (2.9)	-1.4	5.4	-8.1	-0.3	6.6	-7.2
Anterior	65.3 (5.4)	-0.2	4.5	-5.0	0.3	4.7	-4.2
Posteromedial	78.0 (9.6)	-2.0	8.0	-12.0	-0.4	9.9	-10.8
Posterolateral	73.6 (11.7)	-2.1	10.0	-14.2	-5.0	8.7	-18.7
Left leg							
Composite score	73.0 (7.3)	-1.6	3.8	-7.0	-2.8	4.0	-5.6
Anterior	66.0 (5.4)	-0.2	5.0	-5.4	-0.2	4.4	-4.5
Posteromedial	80.0 (8.8)	-1.6	6.3	-9.5	-1.6	6.4	-7.4
Posterolateral	73.1 (10.8)	-2.8	6.5	-12.2	-2.8	6.4	-10.1

LoA = limits of agreement

SDC values for the left leg ranged from 4.4% (A) to 9.0% (PL) and the interrater SDC value for the composite score of the left leg was 5.0%. See Table 2 for all SDC values for the mSEBT.

DISCUSSION

In the current study, intra- and interrater reproducibility of the mSEBT were investigated in a sufficiently large sample of 55 healthy young adults participating in sports at risk for ankle sprain. The results show that the mSEBT has good to excellent intra- and interrater reliability despite the large variation in work experience (from 6 months to 11 years) between the four trained physical therapists who administered the test. Therefore, the current study results suggest that anyone can administer the test if trained properly. SDC values for the composite scores appear to be quite large. Changes in normalized composite scores of at least 6.9% for the right leg and 5.0% for the left leg are needed to observe a true change in performance of dynamic balance.

Earlier studies investigating reliability of the mSEBT gave indications that the mSEBT is a reliable measurement instrument.[6,20-24] However, these studies did not fully accomplish current methodological recommendations for clinimetric research as described in the COSMIN standard.[29] Other authors used small sample sizes or a poor description of the used measurement procedures. The current study performed according to the COSMIN standard, sufficiently powered, and used a transparent measurement protocol according to the recommendations of Gribble and Plisky,[6] confirmed that the mSEBT is a reliable measurement instrument. Recapitulating, the results of the current study support previous findings, that the mSEBT is a reliable measurement instrument, which can be used in practice by one or more raters.

Questions arise whether the found SDC values are acceptable and the mSEBT is an appropriate tool to evaluate dynamic balance performance and risk for ankle sprain in individuals. SDC values for composite reach directions seem to be large and therefore small changes on the mSEBT might occur due to measurement error instead of actual improvement in dynamic balance. Currently, there is no research studying the required minimal improvement in dynamic balance to decrease risk for ankle sprain. To

draw conclusions whether the mSEBT is usable tool for evaluative purposes, further research is needed to investigate the required minimal clinically important difference.

It should be noted that the reach distances obtained in the current study are relatively low compared to normative data from healthy subjects in earlier research. In other studies the PM reach direction scores are close to 90% or 100% from leg length,[16,22,35] where the scores in the current study and the study of McCann[36] are around 80% of leg length. The differences in reach distance scores compared with earlier results could be explained by several reasons. First, the population in our study is not completely comparable to the populations in the previous studies. In the current study, a large proportion of female participants, of recreational activity level, and mostly active in running participated, which may have resulted in relatively low reach distance scores. Second, differences in calculations and protocols used in the studies may have resulted in lower reach distance scores compared to previous studies. In the current study the average score was used for calculation, while in some other studies the highest score out of three trials was used.[16] The protocol used in the current study was a very strict protocol (e.g. both hands had to stay on the hips, forefoot and heel had to keep contact with the floor and start position of the great toe and heel at the beginning of the tape) where other studies handled a less stringent protocol.[16,22,35] Moreover, the current authors also looked into whether fatigue could have affected the observed reach distances. None of the participants reported fatigue symptoms during the test sessions and post hoc analyses on the composite scores of the left and right leg showed no significant differences in composite scores between the different sessions. This indicates that the mSEBT performance was stable across the test sessions and there were no signs of a deterioration of performance due to fatigue within the hour testing.

In the current study, injury risk stratification based on the cut-off points found by Plisky et al. and Butler et al. does not seem useful.[16,17] Plisky found that in high school female basketball players athletes with a composite reach distance less than 94.0% leg length were 6.5 times more likely to have a lower extrem-

ity injury including ankle sprains.[16] Butler reported that male college football players who scored below 89.6% leg length were 3.5 times more likely to get injured.[17] However, risk stratification based on these cut-off points seems not plausible because almost all participants in the current study performed below the cut-off points and it is not likely that all of them had an elevated risk for lower extremity injuries. As earlier mentioned an important explanation for the lower test scores in the current study are the differences between the study populations. As suggested by Butler et al. there is need for developing population-specific cut-off points to screen athletes for injury risk.[17]

A remarkable finding of the current study is that SDC values for the left leg are systematically lower than SDC values for the right leg. A possible explanation for the difference in SDC values could be the effect of cross-education. Earlier research has shown that unilateral balance training was effective in improving neuromuscular reactions to perturbations during single-leg stance for the trained but also for the untrained leg. [37] In the current study each test was started with the right leg as the stance leg and therefore a cross-education effect due to training of the right stance leg may have caused a more stable performance of the leg. Furthermore, it could be hypothesized that because of right feet dominance in the study population (right leg dominance in 51 out of 55 participants) the left leg is mostly the stance leg during balance tasks and therefore is more trained in dynamic balance. This could have led to a more stable performance in maintaining balance for the left leg, resulting in lower SDC values. Although, previous studies do not show differences in mean scores for dynamic balance between the dominant and non-dominant leg,[38-40] no research has investigated differences in measurement error between the dominant and non-dominant leg. In summary, it is unclear whether the differences in SDC values between the left and right leg found in the current study should be attributed to the test sequence used in the measurement protocol or to differences between the dominant and non-dominant leg.

It should be noted that in the current measurement protocol the leg length test was part of the evaluation of the mSEBT. The leg length test is a potential contributor to measurement error. In our study protocol the raters measured the participant's leg length in each of the three test sessions. Since the mSEBT is used in practice to evaluate change in dynamic balance an investigator is not interested in change in leg length. Therefore, measuring leg length at baseline and using these values across the successive tests would be more efficient and decreases the measurement error between mSEBT measurements. Nevertheless, post hoc analyses on measurement error of the leg length measurements showed only a minor contribution to measurement errors. In the current sample inter-rater (average) leg length SEM agreement values were respectively 0.90cm and 0.93cm for the right and the left leg and for both legs ICC values for inter-rater reliability were higher than 0.99. Looking at intra-rater (average) leg length, SEM consistency values were respectively 0.46cm and 0.43cm for the right and the left leg and for both legs inter-rater reliability ICC values were 0.98.

When interpreting the data it has to be taken into account that the SDC values investigated in the current study are based on a sample of healthy young adults participating in sports with increased risk for ankle sprain. SDC values are population specific and may be different in a population suffering from ankle or other lower extremity injuries.[41,42] Therefore, the authors recommend that future research should study the measurement properties of the mSEBT using the same measurement protocol but in different populations at risk for lower extremity injuries and in populations suffering from different lower extremity injuries. Another consideration is that since each test was started with the right leg, the possible influence of a cross education effect could not have been ruled out. The authors therefore recommend that future research should focus on the influence of possible cross education effects, and on side to side differences in measurement error and SDC values, when using a controlled test order.

CONCLUSION
The mSEBT is a reliable measurement instrument used to discriminate dynamic balance between healthy young adults, participating in sports with increased risk for ankle sprain. For evaluative purposes changes of the composite score of at least 6.9% and 5.0% for the right stance leg and the left

stance leg respectively are needed to feel confident that real improvement has occurred in an individual's dynamic balance. More research is needed to investigate if the mSEBT is usable for evaluative purposes of individual dynamic balance, and whether differences in SDC values are due to leg dominance or cross education effect.

REFERENCES

1. Fong DT, Hong Y, Chan LK, Yung PS, Chan KM. A systematic review on ankle injury and ankle sprain in sports. *Sports Med*. 2007;37(1):73-94.

2. Waterman BR, Owens BD, Davey S, Zacchilli MA, Belmont PJ,Jr. The epidemiology of ankle sprains in the united states. *J Bone Joint Surg Am*. 2010;92(13):2279-2284.

3. Gerber JP, Williams GN, Scoville CR, Arciero RA, Taylor DC. Persistent disability associated with ankle sprains: A prospective examination of an athletic population. *Foot Ankle Int*. 1998;19(10):653-660.

4. Kerkhoffs GM, van den Bekerom M, Elders LA, et al. Diagnosis, treatment and prevention of ankle sprains: An evidence-based clinical guideline. *Br J Sports Med*. 2012;46(12):854-860.

5. Yeung MS, Chan KM, So CH, Yuan WY. An epidemiological survey on ankle sprain. *Br J Sports Med*. 1994;28(2):112-116.

6. 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.

7. de Noronha M, Franca LC, Haupenthal A, Nunes GS. Intrinsic predictive factors for ankle sprain in active university students: A prospective study. *Scand J Med Sci Sports*. 2013;23(5):541-7.

8. Fong DT, Chan YY, Mok KM, Yung PS, Chan KM. Understanding acute ankle ligamentous sprain injury in sports. *Sports Med Arthrosc Rehabil Ther Technol*. 2009;1:14.

9. Trojian TH, McKeag DB. Single leg balance test to identify risk of ankle sprains. *Br J Sports Med*. 2006;40(7):610-3.

10. Willems TM, Witvrouw E, Delbaere K, Mahieu N, De Bourdeaudhuij I, De Clercq D. Intrinsic risk factors for inversion ankle sprains in male subjects: A prospective study. *Am J Sports Med*. 2005;33(3):415-423.

11. Willems TM, Witvrouw E, Delbaere K, Philippaerts R, De Bourdeaudhuij I, De Clercq D. Intrinsic risk factors for inversion ankle sprains in females--a prospective study. *Scand J Med Sci Sports*. 2005;15(5):336-345.

12. Steffen K, Emery CA, Romiti M, et al. High adherence to a neuromuscular injury prevention programme (FIFA 11 +) improves functional balance and reduces injury risk in canadian youth female football players: A cluster randomised trial. *Br J Sports Med*. 2013;47(12):794-802.

13. Zech A, Klahn P, Hoeft J, zu Eulenburg C, Steib S. Time course and dimensions of postural control changes following neuromuscular training in youth field hockey athletes. *Eur J Appl Physiol*. 2014;114(2):395-403.

14. McLeod TC, Armstrong T, Miller M, Sauers JL. Balance improvements in female high school basketball players after a 6-week neuromuscular-training program. *J Sport Rehabil*. 2009;18(4):465-481.

15. Filipa A, Byrnes R, Paterno MV, Myer GD, Hewett TE. Neuromuscular training improves performance on the star excursion balance test in young female athletes. *J Orthop Sports Phys Ther*. 2010;40(9):551-558.

16. 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.

17. Butler RJ, Lehr ME, Fink ML, Kiesel KB, Plisky PJ. Dynamic balance performance and noncontact lower extremity injury in college football players: An initial study. *Sports Health*. 2013;5(5):417-422.

18. Dallinga JM, Benjaminse A, Lemmink KA. Which screening tools can predict injury to the lower extremities in team sports?: A systematic review. *Sports Med*. 2012;42(9):791-815.

19. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10):1033-1039.

20. Clark RC, Saxion CE, Cameron KL, Gerber JP. Associations between three clinical assessment tools for postural stability. *N Am J Sports Phys Ther*. 2010;5(3):122-130.

21. Gribble, Phillip Kelly, Sarah Refshauge,Kathryn Hiller, Claire. Interrater reliability of the star excursion balance test. *Journal of athletic training*. 2013;48(5):621-626.

22. Munro AG, Herrington LC. Between-session reliability of the star excursion balance test. *Phys Ther Sport*. 2010;11(4):128-132.

23. Robinson RH, Gribble PA. Support for a reduction in the number of trials needed for the star excursion balance test. *Arch Phys Med Rehabil*. 2008;89(2):364-370.

24. 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.

25. Factsheet enkelblessures consument en veiligheid. http://www.veiligheid.nl/csi/veiligheidnl.nsf/0/360 6F8C12637A14DC1257B4100271197/$file/ Factsheet%20enkelblessures.pdf. Updated 2012. Accessed March 7, 2013.

26. 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.

27. Gribble PA, Hertel J. Considerations for normalizing measures of the star excursion balance test. *Measurement in Physical Education & Exercise Science*. 2003;7(2):89-100.

28. 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*. 2001;20(21):3205-3214.

29. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651-657.

30. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol*. 2010;10:22.

31. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40(2):171-178.

32. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.

33. Coppieters M, Stappaerts K, Janssens K, Jull G. Reliability of detecting 'onset of pain' and 'submaximal pain' during neural provocation testing of the upper quadrant. *Physiother Res Int*. 2002;7(3):146-156.

34. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310.

35. 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.

36. McCann RS, Kosik KB, Beard MQ, Terada M, Pietrosimone BG, Gribble PA. Variations in star excursion balance test performance between high school and collegiate football players. *J Strength Cond Res*. 2015;29(10): 2765-70.

37. Oliveira AS, Brito Silva P, Farina D, Kersting UG. Unilateral balance training enhances neuromuscular reactions to perturbations in the trained and contralateral limb. *Gait Posture*. 2013; 38(4):894-9.

38. Alonso AC, Brech GC, Bourquin AM, Greve JM. The influence of lower-limb dominance on postural balance. *Sao Paulo Med J*. 2011;129(6):410-413.

39. Hoffman M, Schrader J, Applegate T, Koceja D. Unilateral postural control of the functionally dominant and nondominant extremities of healthy subjects. *J Athl Train*. 1998;33(4):319-322.

40. Matsuda S, Demura S, Nagasawa Y. Static one-legged balance in soccer players during use of a lifted leg. *Percept Mot Skills*. 2010;111(1):167-177.

41. Gjorup T. The kappa coefficient and the prevalence of a diagnosis. *Methods Inf Med*. 1988;27(4):184-186.

42. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Stat Med*. 1994;13(23-24):2465-2476.