Reliability Testing of the Balance Error Scoring System in Children Between the Ages of 5 and 14

Colby Hansen, MD,
colby.hansen@hsc.utah.edu Division of Physical Medicine and Rehabilitation University of Utah Salt Lake City, UT U.S.A.

Dan Cushman, MD,
Division of Physical Medicine and Rehabilitation University of Utah Salt Lake City, UT U.S.A.

Wei Chen, PhD,
Wei.Chen@utah.edu Division of Epidemiology University of Utah Salt Lake City, UT U.S.A.

Jerry Bounsanga, BS, and
jerry.bounsanga@utah.edu Department of Orthopaedics University of Utah Salt Lake City, UT U.S.A.

Man Hung, PhD
man.hung@hsc.utah.edu Department of Orthopaedics University of Utah Salt Lake City, UT U.S.A.

Abstract

Objective—Pediatric concussion is an extensive public health concern with a complex clinical presentation. Balance assessment has been well-studied in the adult population, but has been limited in children. We aimed to assess the reliability and minimum detectable change (MDC) of the Balance Error Scoring System (BESS) in healthy children.

Design—This prospective observational study included 373 healthy children aged 5 to 14. Inter-rater reliability was assessed by having 4 assessors review videos of 50 random subjects distributed evenly by age and gender across the entire cohort. Intra-rater reliability was done by having assessors review videos of other assessors' live evaluations. Test-retest reliability was obtained by comparing BESS scores as recorded live at the two separate time points by the same rater.

Setting—Local elementary and junior high schools

Participants—373 healthy children between the ages of 5 and 14.

Interventions—The BESS was performed on all children.

CORRESPONDING AUTHOR: Dan Cushman, MD, 590 Wakara Way, Salt Lake City, UT 84108, dan.cushman.work@gmail.com, 801-213-3476 (phone), 801-587-5482 (fax).

Main Outcome Measures—Intraclass Correlation Coefficients and Minimum Detectable Change Scores

Results—The overall inter-rater intraclass correlation coefficient (ICC) was determined to be 0.93 (95% confidence interval [CI] 0.79 – 0.97) and intra-rater ICC was 0.96 (95% CI 0.95 – 0.97) with individual intra-rater ICCs ranging between 0.69 and 0.99. The test-retest reliability was 0.90 (95% CI 0.88 – 0.92). The MDCs were 9.6, 4.6, and 7.3 points at the 95% confidence levels for inter-rater, intra-rater, and test-retest comparisons, respectively. No learning effect was seen.

Conclusions—The BESS demonstrates excellent reliability in the pediatric population without evidence of a learning effect.

Keywords

Brain Concussion; Sports; Pediatrics; Child; Reproducibility of Results

INTRODUCTION

Pediatric concussion is a widespread public health concern that has received significant media attention in recent years. Between 1997 and 2007, emergency department visits for concussion have increased by 100% in children age 8 to 13 and by 200% in those age 14 to 19.[1] Clinical research supports the idea that children and adolescents take longer to recover from a concussion.[2] A recent Institute of Medicine committee statement suggested that more research should be performed to "establish objective, sensitive, and specific metrics and markers of concussion diagnosis, prognosis, and recovery in youth".[3] Evaluation of concussion includes balance assessment,[4-6] often involving the Balance Error Scoring System (BESS).[4,7-9]

Administering the BESS involves testing postural control in three different positions (double leg stance, tandem stance, and single leg stance) on both a firm and foam surface. The reliability of the BESS has been well-studied in the adult population. The intra-rater reliability in adult and high school populations has been demonstrated to have intraclass correlation coefficients (ICC) that range between 0.63 and 0.92.[10-14] Inter-rater reliability has been shown to be between 0.44 and 0.96 for each portion of the test,[11,14,15] with an overall test ICC of 0.57 in an adult population.[11] The minimum detectible change (MDC) has also been studied in an adult population with the mean intrarater and interrater MDCs across 3 testers were 7.3 and 9.4 points, respectively.[11]

Data for BESS testing in the pediatric population is much more scarce, however.[16] Two small studies performed by Valovich-McLeod et al. have looked at the BESS in youth aged 9 to 14. These studies used a modest sample size of 50 children and displayed high intrarater ICCs ranging from 0.87 to 0.95 for the various subcomponents of the test, with an ICC of 0.98 for the total score,[17] and a test-retest reliability of 0.70.[18]

To our knowledge, the reliability of the BESS has not been studied in a large pediatric population. Our aim was to examine the inter- and intra-rater reliability, the test-retest reliability, the Minimum Detectable Change (MDC), and evaluate for the presence of a

learning effect for the BESS in a large pediatric population between the ages of 5 and 14 years.

ETHICAL CONSIDERATIONS

This prospective observational study was approved by the University of Utah Institutional Review Board. Written parental permission and verbal child permission were required from all subjects.

METHODS

Children aged 5 to 14 were recruited from local school districts (elementary and junior high schools), after approval from their school board, by approaching principals and physical education teachers. Children without a history of any known visual, vestibular, or balance disorder were included in the study. Potential subjects with a lower extremity injury or pain within the last six months were excluded, as were children with active concussion symptoms. Parents of subjects were asked about prior head injuries and further information about the number, perceived recovery, and timing of the concussions were recorded. Parental opinion of the child's balance ability was also recorded.

After obtaining parental and subject consent, children were screened for any active sources of lower extremity pain, and height and weight were measured. The BESS was performed, as below, at the child's school as part of their Physical Education class. Testing took place in a quiet location separate from the gymnasium, free of distractions. Up to four testing stations were used at any one time, with portable dividers used to minimize distractions from nearby subjects. The testing surface was generally equivalent to a concrete or linoleum floor, and a non-slip grip mat was used under the foam pad for those respective conditions to prevent the foam pad from slipping under the chidren's feet. Children, who were dressed in regular school attire, were asked to remove their shoes and identify which foot they preferentially used to kick a ball. This was used as the dominant foot in the BESS testing protocol. If they could not reliably identify which foot they preferred to kick a ball with, they were asked which hand they write with; in these instances the ipsilateral foot was used as the dominant foot. Children were videotaped and re-tested 2-3 weeks later in a similar fashion for the purposes of reliability determinations. This timeframe was chosen for logistic and feasibility purposes, but also because we feel it mirrors a reasonable timeframe within which a health care provider may reassess a young concussed athlete. Videotaping was done in order to facilitate the completion of inter-rater assessments.

The BESS testing, as described by Riemann et al.,[15] was performed – a summary is provided in table 1. Briefly, subjects performed three 20-second standing balance conditions with hands on the iliac crests and eyes closed: one with the feet together, one with the dominant leg lifted, and one in a tandem position with a leading dominant leg. These conditions were repeated with the participants standing on a piece of medium-density foam (Airex Balance Pad, Alcan, Inc., Montreal, Canada). Single points were given for errors in balance – lifting the hands off the iliac crests, opening the eyes, stepping, stumbling, falling, remaining out of position for more than 5 seconds, moving the hip into more than 30 degrees

of flexion or abduction, or lifting the forefoot or heel. A maximum of 10 points per condition is allowed[15], for a total score ranging from 0 to 60. If the subject was unable to maintain stability for 5 seconds at any point during the trial, a maximum score of 10 was given for that condition. A total score was determined from the sum of the error points given for each of the six conditions.

Test-retest reliability was obtained by comparing BESS scores as recorded live at the two separate time points by the same rater. Intra-rater reliability was done by having assessors review videos of their live assessments. Only assessors who tested a minimum of 10 subjects were asked to review their own videos. Inter-rater reliability was assessed by having 4 assessors review videos of 50 random subjects distributed evenly by age and gender across the entire cohort. All assessors, for both intra-rater and inter-rater reliability determinations had experience administering the BESS in clinical encounters and received additional training in performing the BESS. This training took place in the form of live clinical assessments with the lead author who routinely evaluates patients in the age demographic of this study, as well as by reviewing an online course produced by the lead author in which BESS video demonstrations of actual patients in this age demographic were viewed, scored, and discussed. All subjects were part of a project aimed at deriving normative data for the BESS and modified BESS in children; results on the normative data are discussed elsewhere[19].

Statistical analysis

We described participant demographic characteristics using mean, standard deviation, median and proportion as appropriate. We also generated matrix scatter plots displaying the BESS scores recorded by the four reviewers.

In order to examine intra-rater, inter-rater, and test-retest reliabilities of the BESS, Intraclass Correlation Coefficients (ICCs) and their associated 95% Confidence Intervals were estimated. The inter-rater observations were based on the BESS total scores across the four reviewers overall; the test-retest observations were based on the first live and the second live tests of all the reviewers; the intra-rater observations were based on the first live test and first video within individual reviewer. Furthermore, the MDC for the intra-rater, inter-rater and test-retest reliabilities were computed for the total BESS scores; individual stance data was not analyzed.. For reliability testing, a power analysis determined that a sample size of 46 kids with 4 raters rating each subject on video review achieves 99% power to detect an ICC of 0.3 using an F-test with an alpha of 0.05.

A p-value of less than 0.05 was considered to be statistically significant. All analyses were performed using SPSS 22 (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.) and SAS 9.2 (SAS Institute Inc., Cary, NC, USA).

Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Utah.[20] REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and

export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources.

RESULTS

A total of 388 potential subjects were screened for the study, of which 373 participated. Six potential subjects were excluded for previous diagnosis of a visual, vestibular, or balance disorder; eight were excluded for having a lower extremity injury within the last six months; and one child was excluded due to a recent, symptomatic concussion. The median age of participants was 9 years of age (range 5-14), with 58.7% male gender, and 83.7% participating in team or individual sports. A total of 10.5% of the participants had sustained a concussion in the past, of which 84.6% of those participants had sustained only one concussion, 66.7% had occurred greater than 1 year of ago, 94.9% reported complete recovery, and 5.13% reported that they were mostly recovered. These subjects were included in the study. 16 total reviewers participated in test administration. A thorough description of subject demographics and BESS test results is available separately[19].

Of 373 subjects, 331 (89%) completed both the first and second assessments. The test-retest reliability was determined to be 0.90 (95% confidence interval [CI] 0.88 – 0.92). Regarding the possibility of a learning effect, the mean difference in BESS score between the two live test administrations was 0.036 (95% CI −0.53 – 0.60), and the paired t-test result (p=0.900) revealed no significant learning effect as a group. However, in some individuals there were rather large differences between the first and second test. Specifically, in 9% of subjects the BESS score did not change, in 44% of subjects the BESS score worsened (mean change 4.43, st dev 3.70, range 1-27), while in 47% of subjects the BESS score improved (mean change 4.22, st dev 3.00, range 1-15).

Overall intra-rater ICC was 0.96 (95% CI 0.95 – 0.97) with individual intra-rater ICCs ranging between 0.69 and 0.99. Eight of 16 reviewers had completed enough live assessments and subsequent video reviews to be eligible for analysis of intra-rater reliability. See table 2 for a summary of intra-rater ICCs.

Overall inter-rater ICC was 0.93 (95% CI 0.79 – 0.97). Figure 1 graphically demonstrates the BESS correlations between raters, for which correlation coefficients ranged from 0.85 to 0.95.

MDC was calculated according to various scoring comparisons. MDC based on the first and second live test administration was 6.2 at the 90% CI, 7.3 at the 95% CI, and 9.6 at the 99% CI. MDC based on the intra-rater scoring variation was 3.9 at the 90% CI, 4.6 at the 95% CI, and 6.0 at the 99% CI. Lastly, MDC based on the inter-rater variation was 8.1 at the 90% CI, 9.6 at the 95% CI, and 12.6 at the 99% CI. These are illustrated in Table 3 for reference.

DISCUSSION

A good test has two important features: validity and reliability. To achieve validity, a good test needs to measure what it claims to measure. There are several types of validities – criterion validity, content validity and construct validity. In a pediatric concussion

population, the criterion validity has ranged from poor to adequate in comparisons to the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) and Post-Concussion Symptom Scale (PCSS).[21] Excellent content validity has been demonstrated in adults.[22] And the BESS has shown good construct validity in identifying balance deficits in different conditions.[22] A good test also needs to be reliable; that is, how consistently a test produces information about the subjects. Test scores should not be affected by measurement errors. Validity does not necessarily imply reliability and vice versa. Yet, for a test to be considered as good, both validity and reliability evidences are required.

These results give valuable information on the reliability of the BESS in a pediatric population. Though well-studied in an adult population,[10-15] literature regarding the reliability of this test in the pediatric population is scarce.[16] This data expands upon previous work by Valovich-McLeod et al.,[17,18] using a larger population of healthy children between the ages of 5 and 14.

The test-retest reliability of the BESS has previously been demonstrated in a population of 50 pediatric athletes over 60 days to have an intraclass correlation coefficient of 0.70 ,[18] which is much lower than our value of 0.90. The difference may arise from the fact that our children were retested much sooner (2-3 weeks after initial testing). Valovich et al.[17] suggested that there may be a learning effect for the BESS in a pediatric population when retested within days; this has also been demonstrated in young adult populations.[23-25] We found no evidence for this over this longer time period, as scores for initial and repeat testing, performed over 2-3 weeks, showed no significant difference. We used a similar population, but used a much longer break between test administrations, which could explain the loss of a potential learning effect that may have been present at an earlier time point. Interestingly, Valovich et al.[18] tested 50 youth sport participants over 60 days and demonstrated a significant difference in scores between tests, which does not appear in our similar study. One difference may be that the prior study specifically recruited youth athletes who may see the testing as a challenge and something to improve upon, or may have a better ability to train their bodies based off a single test.

Intra-rater reliability of the BESS has been commonly studied in the young adult and adult populations, and ranges from 0.63 to 0.92.[10-14] Our reviewers had similar reliability values, ranging from 0.69 to 0.99. Most ICC values were above 0.9 and the overall ICC was found to be 0.96, which is similar to prior work in the pediatric population, which demonstrated an ICC of 0.98.[17] Valovich et al. also studied a relatively small group of 50 athletes between the ages of 9 and 14. In our study, 250 subjects of our total sample between the ages of 5 and 14 were used to derive this intra-rater data.

Inter-rater reliability was high in our population of 50 healthy children, moreso than those recorded in the young adult populations.[11,14] This is an interesting phenomenon, as young adult populations (Iverson 2008) tend to have lower amounts of errors in the BESS compared to pediatric populations.[17-19] It is possible that the types of errors seen in pediatric populations are more exaggerated in appearance and easier to identify than those in adults. We also used different populations of raters (physicians, physical therapists, and certified

athletic trainers), which suggests that different types of clinicians can administer the test reliably.

We looked at MDC scores according to repeated assessments over time to identify what constitutes a significant change in BESS score, as well as the MDC if different assessors are conducting the examination. For serial BESS test administrations done by the same reviewer, our data (95% CI for MDC of 7.3) suggest that a change of 8 points in the BESS for the same reviewer has a greater than 95% probability that it is due to a change in postural stability rather than intra-rater variability. Valovich et al.[18] used a reliable change index (RCI) method rather than MDC in their study of 50 student athletes between the ages of 9 and 14 that were administered the BESS approximately 60 days apart; their data was reported in 70%, 80%, and 90% confidence intervals. They reported a significant worsening (or increase) in BESS score at a 90% confidence interval was 5.3 units, or 6 points on the BESS. They reported a significant improvement (or decrease) in BESS score at a 90% CI was 9.4 units, or 10 points.

For intra-rater variability scoring the same test administration, our MDC of 4.58 (95% CI) is slightly better than that reported by Finnoff et al.[11] who reported an intra-rater MDC of 7.3 points in an adult population. Their study did not report, however, the confidence interval at which they established their intra-rater MDC.

Regarding MDC for assessments done by different raters, our MDC of 9.6 points at a 95% CI suggests that a change of 10 points in the BESS when viewed by different reviewers is due to a change in postural stability with a >95% certainty. This is almost identical to the findings of Finnoff et al.[11] who demonstrated an inter-rater MDC of 9.4 points when measured in an adult population. Both of our findings suggest that the majority of variability in testing comes from differences between raters.

Clearly, future studies need to investigate the performance of children with concussion on BESS testing to establish discriminant validity, (the ability of the BESS to accurately distinguish concussed from non-concussed children at the time of initial evaluation) and convergent validity (whether the BESS is related to similar instruments). It will also be important to determine the timing of normalization of BESS performance in children with concussion, and how these measures compare to other measures, such as neuropsychological tests or symptom scores. Our study was limited by the fact that our data were limited to a sample from local schools. Hence, the results may not be generalizable to other populations. As testing was performed on a group of healthy children, for whom one could expect relatively little variability, this reliability may be lower in an affected population. Finally, it should be noted that the comparatively broad range of ages in our subjects could contribute to a relatively lower reported reliability, particularly as the younger children (ages 5-10), which were included in our cohort, tend to have greater variability of their BESS score and make more errors[19].

In conclusion, this is the first large study examining the reliability of the BESS in a pediatric population. The BESS was demonstrated to have a test-retest reliability of 0.90 (MDC 7.3 points), an intra-rater ICC of 0.96 (MDC 4.6 points), and an inter-rater ICC of 0.93 (MDC

9.6 points). This suggests that the BESS is a reliable measure for balance assessment in the pediatric population, but may lack responsiveness as relatively large MDCs are required to assess change. Our study should increase the confidence for healthcare providers to utilize the BESS as a standard part of the physical exam in children with suspected concussion, even as young as age 5.

ACKNOWLEDGEMENTS

REFERENCES

1. Bakhos LL, Lockhart GR, Myers R, Linakis JG. Emergency department visits for concussion in young child athletes. Pediatrics. 2010; 126(3):e550–556. [PubMed: 20805145]

2. Field M, Collins MW, Lovell MR, Maroon J. Does age play a role in recovery from sports-related concussion? A comparison of high school and collegiate athletes. The Journal of pediatrics. 2003; 142(5):546–553. [PubMed: 12756388]

3. Youth CoS-RCi. Sports-Related Concussions in Youth: Improving the Science, Changing the Culture. Graham, R.; Rivara, FP.; Ford, MA.; Spicer, CM., editors. National Academies Press (US); Washington (DC): 2014.

4. Scat3. Br J Sports Med. 2013; 47(5):259. [PubMed: 23479480]

5. Giza CC, Kutcher JS, Ashwal S, et al. Summary of evidence-based guideline update: evaluation and management of concussion in sports: report of the Guideline Development Subcommittee of the American Academy of Neurology. Neurology. 2013; 80(24):2250–2257. [PubMed: 23508730]

6. McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on concussion in sport - The 3rd international conference on concussion in sport held in Zurich, November 2008. Pm r. 2009; 1(5):406–420. [PubMed: 19627927]

7. McCrea M, Guskiewicz KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study. JAMA : the journal of the American Medical Association. 2003; 290(19):2556–2563. [PubMed: 14625332]

8. Murray N, Salvatore A, Powell D, Reed-Jones R. Reliability and validity evidence of multiple balance assessments in athletes with a concussion. J Athl Train. 2014; 49(4):540–549. [PubMed: 24933431]

9. Riemann BL, Guskiewicz KM. Effects of mild head injury on postural stability as measured through clinical balance testing. J Athl Train. 2000; 35(1):19–25. [PubMed: 16558603]

10. Erkmen N, Ta kın H, Kaplan T, Sanio lu A. The effect of fatiguing exercise on balance performance as measured by the balance error scoring system. Isokinetics & Exercise Science. 2009; 17(2):121–127.

11. Finnoff JT, Peterson VJ, Hollman JH, Smith J. Intrarater and interrater reliability of the Balance Error Scoring System (BESS). Pm r. 2009; 1(1):50–54. [PubMed: 19627872]

12. Hunt TN, Ferrara MS, Bornstein RA, Baumgartner TA. The reliability of the modified Balance Error Scoring System. Clin J Sport Med. 2009; 19(6):471–475. [PubMed: 19898074]

13. Susco TM, Valovich McLeod TC, Gansneder BM, Shultz SJ. Balance Recovers Within 20 Minutes After Exertion as Measured by the Balance Error Scoring System. J Athl Train. 2004; 39(3):241–246. [PubMed: 15496993]

14. Valovich 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. [PubMed: 20108849]

15. Riemann BL, Guskiewicz KM, Shields EW. Relationship Between Clinical and Forceplate Measures of Postural Stability. Journal of Sport Rehabilitation. 1999; 8(2):71.

16. Davis GA, Purcell LK. The evaluation and management of acute concussion differs in young children. Br J Sports Med. 2014; 48(2):98–101. [PubMed: 23613516]

17. Valovich McLeod TC, Perrin DH, Guskiewicz KM, Shultz SJ, Diamond R, Gansneder BM. Serial administration of clinical concussion assessments and learning effects in healthy young athletes. Clin J Sport Med. 2004; 14(5):287–295. [PubMed: 15377968]

18. Valovich McLeod TC, Barr WB, McCrea M, Guskiewicz KM. Psychometric and measurement properties of concussion assessment tools in youth sports. J Athl Train. 2006; 41(4):399–408. [PubMed: 17273465]

19. Hansen, C.; Cushman, D.; Anderson, N., et al. (Unpublished) A Normative Data Set of the Balance Error Scoring System in Children Between the Ages of 5 and 14. 2015.

20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009; 42(2):377–381. [PubMed: 18929686]

21. Barlow M, Schlabach D, Peiffer J, Cook C. Differences in change scores and the predictive validity of three commonly used measures following concussion in the middle school and high school aged population. International journal of sports physical therapy. 2011; 6(3):150–157. [PubMed: 21904694]

22. Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error scoring system. Sports health. 2011; 3(3):287–295. [PubMed: 23016020]

23. Mancuso J. An investigation of the learning effect for the Balance Error Scoring System and its clinical implications [abstract]. J Athl Train. 2002; 37(S-10)

24. Mulligan IJ, Boland MA, McIlhenny CV. The balance error scoring system learned response among young adults. Sports health. 2013; 5(1):22–26. [PubMed: 24381697]

25. Valovich TC, Perrin DH, Gansneder BM. Repeat Administration Elicits a Practice Effect With the Balance Error Scoring System but Not With the Standardized Assessment of Concussion in High School Athletes. J Athl Train. 2003; 38(1):51–56. [PubMed: 12937472]
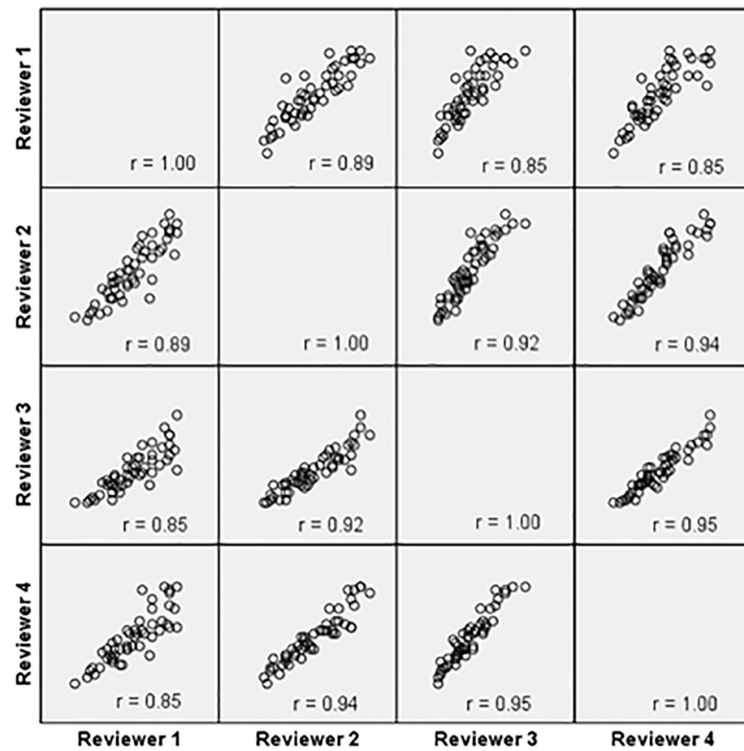
Figure 1.
Matrix scatter plot comparing four reviewers. Each of the 4 × 4 squares represents an individual graph comparing Balance Error Scoring System (BESS) total scores for 50 subjects. Within each square, each circle represents one reviewer's total BESS score on the y-axis with the other reviewer's total BESS score on the x-axis, for the same subject. Perfect correlation would demonstrate circles that create a straight line with a slope of 1.

Table 1

Balance Error Scoring System outline, adapted from Reimann et al.[15]

Errors	
Lifting hands off iliac crests	Remaining out of the test position for more than 5 seconds
Opening eyes	Moving hip into more than 30° of flexion or abduction
Stepping, stumbling, or falling	Lifting forefoot or heel
Scoring system – one point given for each error above, with a maximum of 10 points per test. Total scores range from 0 to 60.	
1. Double-firm – Narrow bilateral stance with medial aspects of feet together	
2. Single-firm – Standing on non-dominant leg	
3. Tandem-firm – Standing feet in line, with the heel of dominant foot touching the toes of non-dominant foot	
4. Double-foam – Narrow bilateral stance with medial aspects of feet on midline of foam	
5. Single-foam – Standing on non-dominant leg in the middle of the foam	
6. Tandem-foam – Standing diagonally across the middle of the foam, feet in line, with the heel of dominant foot touching the toes of non-dominant foot	

Table 2

Intra-rater intraclass correlation coefficients (ICCs) for all Balance Error Scoring System reviewers viewing a pediatric population.

	First Live Vs First Video		Second Live Vs Second Video	
	Intraclass Correlation Coefficient	95% Confidence Interval (Lower & Upper Bound)	Intraclass Correlation Coefficient	95% Confidence Interval (Lower & Upper Bound)
Reviewer 1	0.988	0.962-0.992	0.991	0.971-0.997
Reviewer 2	0.971	0.941-0.986	0.962	0.915-0.983
Reviewer 3	0.981	0.970-0.988	0.969	0.933-0.984
Reviewer 4	0.919	0.748-0.969	0.953	0.813-0.984
Reviewer 5	0.728	-0.222-0.940	0.920	0.699-0.980
Reviewer 6	0.964	0.914-0.984	0.919	0.786-0.970
Reviewer 7	0.973	0.950-0.986	0.989	0.979-0.995
Reviewer 8			0.690	-0.257-0.928

Table 3

Summary table of minimum detectable change (MDC) values for different reliability values at various confidence intervals (CIs) for the Balance Error Scoring System in a pediatric population.

	MDC at 90% CI	MDC at 95% CI	MDC at 99% CI
Test-retest	6.17	7.33	9.64
Intra-rater	3.85	4.58	6.02
Inter-rater	8.06	9.57	12.60