Journal of Athletic Training 2011:46(6):600–606

www.nata.org/jat

Validity and Reliability of Limits-of-Stability Testing: A Comparison of 2 Postural Stability Evaluation Devices

Marie L. Pickerill, PhD, ATC*; Rod A. Harter, PhD, ATC†

*Department of Kinesiology, DePauw University, Greencastle, IN; †Department of Health and Human Performance, Texas State University, San Marcos

Context: A lack of published comparisons between measures from commercially available computerized posturography devices and the outcome measures used to define the limits of stability (LOS) makes meaningful interpretation of dynamic postural stability measures difficult.

Objectives: To compare postural stability measures between and within devices to establish concurrent and construct validity and to determine test-retest reliability for LOS measures generated by the NeuroCom Smart Balance Master and the Biodex Balance System.

Design: Cross-sectional study.

Setting: Controlled research laboratory.

Patients or Other Participants: A total of 23 healthy participants with no vestibular or visual disabilities or lower limb impairments.

Intervention(s): The LOS were assessed during 2 laboratory test sessions 1 week apart.

Main Outcome Measure(s): Three NeuroCom LOS variables (directional control, endpoint excursion, and movement velocity) and 2 Biodex LOS variables (directional control, test duration).

Results: Test-retest reliability ranged from high to low across the 5 LOS measures (intraclass correlation coefficient [2,k] = 0.82 to 0.48). Pearson correlations revealed 4 significant relationships ($P < .05$) between and within the 2 computerized posturography devices ($r = 0.42$ to -0.65).

Conclusions: Based on the wide range of intraclass correlation values we observed for the NeuroCom measures, clinicians and researchers alike should establish the reliability of LOS testing for their own clinics and laboratories. The low to moderate reliability outcomes observed for the Biodex measures were not of sufficient magnitude for us to recommend using the LOS measures from this system as the gold standard. The moderate Pearson interclass correlations we observed suggest that the Biodex and NeuroCom postural stability systems provided unique information. In this study of healthy participants, the concurrent and construct validity of the Biodex and NeuroCom LOS tests were not definitively established. We recommend that this study be repeated with a clinical population to further explore the matter.

Key Words: postural control, balance, measurements

Key Points

- The NeuroCom Smart Balance Master provided high test-retest reliability, supporting its use in assessing dynamic postural stability in healthy participants.
- The NeuroCom Smart Balance Master offered more information on dynamic postural stability than did the Biodex Stability System, but each system offered unique information toward an overall postural stability assessment.
- Until a standardized definition of dynamic postural stability is agreed on by researchers and clinicians, neither the NeuroCom nor the Biodex can be considered the criterion standard for assessment.

Postural stability has been defined as the ability to control the body's center of gravity (COG) within a given base of support[1,2] and has been extensively researched.[3–12] To date, postural stability researchers[1,13] have defined the continuum of postural stability from static stability to functional stability. However, the understanding of postural stability control (ie, balance) that is essential for performing activities of daily living and achieving success in sports remains complicated by vague terminology and numerous outcome measures.

Postural stability measures used to evaluate postinjury and postsurgical musculoskeletal somatosensation have gained support from the sports medicine community,[4,5,7,8] and the effects of prophylactic ankle bracing, foot orthotics, balance training, and skill training on postural control and athletic performance have all been investigated.[6,12,14,15] Yet despite the recent advances in postural stability measurement and the increased applicability of research findings to clinical practice, 3 key problems remain: nomenclature, criterion standards, and technology.

The first challenge is the lack of a standardized postural stability nomenclature. The interchangeability of the terms *balance* and *postural stability* contributes to this concern; however, for the focus of this study, the construct of dynamic balance is our primary interest. Clinicians need to be familiar with the current nomenclature in order to properly assess patient outcomes and recognize differences in dynamic postural stability testing protocols to make appropriate testing and treatment decisions. We selected the operational definition supported by Nashner and McCollum[2] to review 2 criterion tests developed to evaluate

this construct. In limits-of-stability (LOS) testing, the person's foot position does not change relative to the platform; however, the platform may move relative to the horizontal surface in one testing design. Tests designed to measure the same construct should show convergent validity, given that the test designs were based on similar definitions.

A second concern is the lack of a criterion standard or gold standard for dynamic postural stability (ie, a single evaluative construct that defines good or normal dynamic balance). The Berg Balance Scale formalized the assessment of dynamic and functional postural stability[16] and established the criteria for dynamic balance. The Berg Scale is a subjective assessment validated only for evaluating older adults.[16] As more advanced computerized tests of dynamic balance are developed, it is important to evaluate both the construct and concurrent validity of these tests in order to improve our clinical capacity to accurately evaluate human movement.

The manner in which the numerous variables derived from advanced postural stability kinetic and kinematic technology relate to postural stability has given rise to questions about the validity, reliability, and objectivity of test measures.[4,14,17–20] Additionally, questions of clinical applicability and meaningfulness of test measures to aid in the evaluation and rehabilitation of clinic clientele must be addressed.[3] The foci here are the LOS tests and the quantification of dynamic stability, calculated using ground reaction force data to locate the center of pressure. Conversion of these data to COG sway angles suggests that regardless of height, the ultimate LOS for adults range from 6.25° to 8° forward, 4° to 4.45° backward, and 8° laterally.[2]

In order to accurately assess a criterion for dynamic postural stability, the outcome measures obtained with computerized posturography instrumentation must be both valid and reliable. To date, the reliability of many of the outcome measures used to assess postural stability has not been established.[8,17,19–24] Reports of postural stability studies[14,25] often include several significant outcome measures, with each variable analyzed individually and without mention of the multifaceted nature of dynamic balance. Published evidence of validity for the numerous manual and computerized assessment devices currently in clinical use for measuring postural stability is also lacking. Even when measures are reliable, no clear indication is provided as to which component of stability is affected (eg, visual, vestibular, somatosensory), nor is the location of the deficit or improvement in the somatosensory-neuromuscular system identified.[26] The lack of reliability and validity data is problematic for clinicians and researchers interested in postural stability assessment.

Reliability, the consistency of scores and the lack of measurement error, is a component of validity. Validity is a more complex concept involving multiple components that all provide evidence that the applied measures truly assess and offer information about the stated attribute.[27] Construct and concurrent validity are 2 components of dynamic postural stability that can be investigated through interclass correlation analysis. Construct validity evidence is provided when similar variables are correlated with and predictive of the given construct,[27] in this instance dynamic postural stability. We use concurrent validity, a submeasure of criterion-related evidence for validity, when trying to demonstrate that similar tests measure the same thing, or what researchers often call the gold standard, by demonstrating the highest predictive validity of the theoretical construct.[27] A gold standard is not required, but at minimum, a theoretical construct must be established.[27] Given that no current gold standard exists for dynamic postural stability, outcomes from these tests are used as criteria to compare one test against another. Using the operational definition of dynamic stability, researchers also compare these tests with the definition to assess their construct validities.

Dynamic postural stability outcome measures produced by various commercial testing devices should be compared to provide much-needed information about the quantification of LOS and the construct of dynamic postural stability. To date, we are aware of no authors who have reported comparisons of commercially available computerized posturography devices and the testing outcome measures used to determine dynamic LOS, leaving both the construct validity and concurrent validity of these dynamic postural stability tests in question.

Our intention was to evaluate whether 2 patented computerized posturography testing devices that quantify LOS assessed similar or different components of postural stability. Therefore, the purposes of this study were to determine the magnitude of the relationships between clinical measures from 2 commercially available postural stability testing devices used to assess dynamic LOS to establish concurrent and construct validity and to identify the test-retest reliability of outcomes from both devices.

METHODS

Participants

Twenty-three healthy university students (13 men, 10 women; age = 23.8 ± 5.7 years, height = 175.4 ± 9.0 cm, mass = 79.2 ± 15.1 kg) with no lower limb surgery or trauma during the previous 12 months were recruited via campus fliers for participation. The primary author (M.L.P.) obtained an oral medical history from each volunteer in order to eliminate those with visual conditions (eg, blindness, amblyopia, astigmatism) or vestibular conditions (eg, inner ear infection, Ménière disease, hearing loss). The Oregon State University Institutional Review Board approved the research study. Volunteers who met the inclusion criteria were briefed on the protocol and provided consent before participating.

Protocol

All participants reported to the Sports Medicine Laboratory for 2 testing sessions 1 week apart. Demographic data were collected from all volunteers before testing to determine the general characteristics of the sample population. A randomization table was used to assign 12 participants to test dynamic LOS (DLOS) on the Biodex Balance System (Biodex Medical Systems, Shirley, NY) at the first test session, whereas 11 participants began the study by testing LOS on the NeuroCom Smart Balance Master (NeuroCom International, Clackamas, OR). The reverse test order was used during the second testing session for all volunteers. At each testing session, participants were provided verbal instruction and 3 to 5 minutes of practice with each device before testing. All were barefoot and performed 2 trials of the LOS on each device separated by 5 minutes, rested for approximately 10 minutes, and then performed 2 trials on the other device, also separated by 5 minutes of rest.[28–30]

Biodex Balance System. The Biodex Balance System uses a microprocessor-based actuator to adjust the stability of a

suspended circular force plate. The force platform has a maximum of 20° tilt in any direction when completely unstabilized and determines a participant's stability based on the variance of the platform from center during a given task using a sampling rate of 100 Hz.[29]

The Biodex DLOS test prompts participants to move a cursor, viewed on a liquid crystal display, by leaning toward a target while standing on the fully unstable platform (level 1 of 12 levels, using current model software). Volunteers were instructed to "complete the test as quickly and accurately as possible, keeping your body in a straight line, using the ankles as the primary axis of rotation." The DLOS test measures the time and accuracy with which participants transfer their estimated COG (from ground reaction force and height data), moving the cursor to intercept each of 8 successive targets on the display screen. The targets are positioned at 45° intervals around a central target that represents the participant's center of pressure under static conditions. Each target is randomly highlighted, and the volunteer reaches the target by leaning and returning to the center position before the next target is selected and displayed on the screen. The test is complete when all 8 targets have been reached. Target placement was preset by the manufacturer at 50% of the LOS, based on the height of each volunteer.[29] This process takes into account the conversion of the angular motion of leaning to linear movement of the COG represented on the screen. The 2 dependent measures from the DLOS test were time (seconds) and directional control (based on 100% being a straight line from the center of pressure to the intended target).

NeuroCom Smart Balance Master. The NeuroCom Smart Balance Master assesses dynamic postural stability with the LOS test. The device sampled at a frequency of 100 Hz using a 2–force plate structure connected by a pin joint in the vertical center of the anterior-posterior center line of each plate, with 4 transducers oriented vertically and 1 transducer oriented horizontally.[30] During the LOS test protocol, the NeuroCom force plate remains fixed.

The NeuroCom LOS test required participants to transfer their COG, while standing on stable force plates, toward 8 targets spaced at 45° intervals around the body's COG, as represented on a computer monitor. Volunteers were instructed to "keep their body in a straight line, using their ankle joints at the primary axis of motion and to move toward each target as directly and quickly as possible." They were visually cued to each target separately, as independent subtests (8 seconds each), with targets preset by the manufacturer at 100% of a person's LOS based on height.[30] Target placement takes into account the conversion of the angular motion of leaning to linear movement of the COG represented on the screen. The 3 NeuroCom dependent measures were directional control (as defined previously), endpoint excursion (percentage of the distance achieved toward a target on the initial movement), and movement velocity (average speed of COG movement based on the middle 90% of the distance, measured in degrees per second).

Operational Definitions

For the purposes of this study, the following operational definitions, supported by Nashner and McCollum,[2] were used. *Static postural stability* referred to the ability to limit the movement of the COG when the base of support remained fixed. *Dynamic postural stability* was defined as the ability to shift and control the COG within a fixed base of support. *Functional postural stability* characterized the ability to move and control the COG within a changing base of support. The *base of support* in each case referred to both the foot position and surface condition remaining stationary.

Data Analysis

We assessed the association of outcome measures between devices and within each LOS test. An interclass correlation matrix was calculated using the Pearson *r* procedure to evaluate the correlation of outcome measures. The same scale used for the interclass correlation matrix was used to determine strong, moderate, or weak correlation given statistical significance. Statistical significance was set at $\alpha = .05$ for all comparisons, with data analyzed using the SPSS statistical software package (version 16; SPSS Inc, Chicago, IL).

Test-retest reliability of the data was determined using 5 repeated-measures analyses of variance. A 2-way random effects model was used to calculate intraclass correlation coefficients (ICC [2,k]) for each performance measure to assess reliability.[31] An ICC value ≥0.80 reflected high reliability; ≥0.60, moderate reliability; and <0.60, poor reliability.[32] The 95% confidence intervals (CIs) were established for ICC values at the $P < .05$ level of significance, and standard error of the measurement (SEM) values were calculated for use in interpreting the intraclass reliability of the outcome measures. We used the formula $SEM = s_x\sqrt{1 - r_{xx}}$ to calculate SEM, where s_x is the group's largest SD for the test scores and r_{xx} is the reliability of the test.

RESULTS

Interclass Correlations

All 23 participants completed the testing, and because the data were normally distributed, no data were excluded from the analysis. The Pearson correlation matrix of the combined sessions' outcome measures is presented in Table 1. Initial analyses by session revealed no differences ($P > .05$), so the data were collapsed across sessions and outcome variable means analyzed. A moderate inverse relationship was found between the Biodex outcome measures of directional control and test duration ($r = –0.65$, $P = .001$). Three other outcome measure relationships reached statistical significance, but they were only weakly correlated (0.42 to –0.46). An inverse relationship was observed between the measures of NeuroCom movement velocity and Biodex test duration ($r = –0.46$, $P = .028$), and positive correlations existed between the NeuroCom endpoint excursion and Biodex directional control measures ($r = 0.46$, $P = .029$), and NeuroCom measures of movement velocity and endpoint excursion ($r = 0.42$, $P = .047$). Direct comparison of Biodex and NeuroCom directional control measures revealed a correlation of $r = 0.33$, but this relationship was not significant ($P = .12$).

Intraclass Correlations (Test-Retest Reliability)

The data demonstrated moderate repeatability on the Biodex outcome measure of directional control (ICC [2,k] = 0.72, 95% CI = 0.45, 0.87) and poor reliability for test duration (ICC [2,k] = 0.54, 95% CI = 0.17, 0.78). Confidence intervals are not often reported with ICC values; the application here is that statistical significance is established such that the true population ICC value for the variable of interest will fall within the specified range of ICC values at the 95% CI.[33] The SEM values were

Table 1. Pearson Interclass Correlation Matrix (*P* Value) for Outcome Variables

		Biodex Medical System[a]		NeuroCom Smart Balance Master[b]		
	Variable	Directional Control	Test Duration	Directional Control	Endpoint Excursion	Movement Velocity
Biodex Medical System	Directional control	1.00	NA	NA	NA	NA
	Test duration	–0.65 (.001)[c]	1.00	NA	NA	NA
NeuroCom Smart Balance Master	Directional control	0.33 (.120)	–0.13 (.541)	1.00	NA	NA
	Endpoint excursion	0.46 (.029)[d]	–0.41 (.055)	0.314 (.145)	1.00	NA
	Movement velocity	0.17 (.435)	–0.46 (.028)[d]	–0.31 (.152)	0.42 (.047)[d]	1.00

Abbreviation: NA, not applicable.

[a] Biodex Medical Systems, Shirley, NY.

[b] NeuroCom International, Clackamas, OR.

[c] Correlation significant at the .01 level.

[d] Correlation significant at the .05 level.

±6.38% for Biodex directional control and ±12.03 seconds for test duration.

Similar results were calculated for outcome measures from the NeuroCom testing, with high test-retest reliability observed for endpoint excursion (ICC [2,k] = 0.88, 95% CI = 0.73, 0.94), high reliability observed for movement velocity (ICC [2,k] = 0.80, 95% CI = 0.59, 0.91), and moderate reliability observed for directional control (ICC [2,k] = 0.69, 95% CI = 0.40, 0.86). The SEM values were ±3.01% for endpoint excursion, 0.70°/s for movement velocity, and ±2.79% for directional control.

The data from all LOS testing are summarized in Table 2, with the reliability ICC values and probability values provided.

DISCUSSION

The comparison of the computerized posturography devices used in this study is quite complex. Theoretically, LOS tests that are based on the same dynamic postural stability definition should have outcome variables that correlate highly with one another to demonstrate construct and concurrent validity of postural stability assessment. The interclass correlation process used here helps us to assess the concurrent validity of these devices. This type of validity provides evidence in the area of criterion validity, often referenced as the criterion or gold standard of a test. The absence of a criterion standard against which to measure dynamic postural stability creates a unique structure for validating these tests. We acknowledge different points of view on this process but believe that this starting position in the investigation of concurrent and construct validity of dynamic postural stability is justified based on the available measurement[27,32,33] and postural stability[2,3,20,25] evidence.

Liston and Brouwer[24] used the Berg Balance Scale and gait velocity as criterion standards to establish validity for measures of the NeuroCom LOS test with hemiparetic patients, but the use of these criterion standards was not well justified, particularly because the Berg scale was developed for older adults. Our purpose was to assess the construct and concurrent validity of 2 computerized posturography LOS tests, so each was used as a criterion standard test and sheds light on the criterion standards for the measurement of dynamic postural stability.

The 2 devices provided only 1 similar measure of dynamic postural stability: directional control. We were surprised that the correlation between these variables ($r = 0.333$) was not significant with a greater than 2-fold difference between the Biodex (34%) and NeuroCom (80%) directional control means. Even with large mean differences, we expected these variables to correlate highly, with participants who scored well on 1 device also scoring well on the other, to establish concurrent and construct validity. This lack of correlation does not support construct validity, but we believe that differences in testing protocols were responsible for this discrepancy, and, in fact, the devices assess different components of postural stability: dynamic and functional. Both tests require that the COG is moved, meeting the operational definition of dynamic stability, although only the NeuroCom LOS maintains a fixed base of support. Although the Biodex DLOS maintains foot position, the unstable platform causes continual changes in the size of the base of support, given the perpendicular relationship of

Table 2. Outcome Measures

		Session 1 (Mean ± SD)		Session 2 (Mean ± SD)		Intraclass Correlation Coefficient
Device	Measure	Trial 1	Trial 2	Trial 1	Trial 2	(2,k) (*P* Value)
Biodex Balance System[a]	Directional control, %	27.09 ± 10.96	34.13 ± 11.17	34.61 ± 10.45	39.7 ± 12.16	0.72 (.001)
	Test duration, s	90.48 ± 17.77	75.52 ± 15.78	75.22 ± 14.00	63.78 ± 11.33	0.54 (<.001)
NeuroCom Smart Balance Master[b]	Directional control, %	80.26 ± 4.16	79.52 ± 5.09	79.39 ± 4.39	79.52 ± 4.60	0.69 (.514)
	Endpoint excursion, %	72.57 ± 8.35	75.13 ± 8.17	74.87 ± 8.69	78.00 ± 7.22	0.88 (.004)
	Movement velocity, °/s	4.67 ± 1.38	5.04 ± 1.51	5.62 ± 1.58	5.98 ± 1.43	0.80 (<.001)

[a] Biodex Medical Systems, Shirley, NY.

[b] NeuroCom International, Clackamas, OR.

the COG to the ground. Because the base of support is measured on a transverse plane relative to the COG, even though a participant's feet remained fixed, the base of support on the unstable Biodex platform changed, actually satisfying Nashner and McCollum's[2] definition of functional postural stability. Therefore, we should reconsider replacing the term *dynamic* with *functional* for the Biodex DLOS test. The answer to the fundamental question, "What exactly is dynamic postural stability?" remains elusive and depends on the field of study; for biomechanics, motor control, and biomedical engineering, the operational definition and answer vary.[3,16,17,20,25]

The profound variations in LOS test administration are sources of variability that contribute to the differences in test results. The Biodex system is a continuous test using randomized ordering of targets, whereas the NeuroCom system uses individual subtests within the LOS test administration for each target in succession, thereby allowing small breaks in testing between subtests. The ongoing nature of the Biodex DLOS test does not allow for recouping lost focus or concentration or a loss of balance; the NeuroCom LOS test guards against these problems. Although test fatigue and practice effects are probably much smaller concerns than is testing protocol variation, clinicians and researchers must consider these factors when interpreting test results.

Only 2 comparisons of outcome variables between the devices were significant. We observed weak to moderate correlations between Biodex directional control and NeuroCom endpoint excursion ($r=0.46$, $P=.029$) and between Biodex test duration and NeuroCom movement velocity ($r=-0.46$, $P=.028$). The relationship between test duration and movement velocity is readily explained, because both measures involve a time component: The more quickly a person moves, the less time it takes to complete the test.

Considering the time component relative to these LOS tests is intriguing. Time is not an evaluative component of knowing one's LOS, although given the element of fatigue, one's LOS plays a role in determining balance. The Biodex DLOS incorporates a fatigue component, which, although it does not fit the definition of dynamic balance, certainly provides information about the spatiotemporal nature of balance. Pai et al[14] suggested that the nervous system controls both COG and COG velocity for dynamic postural stability and that various strategies may be selected to achieve high positive outcomes. Hertel et al[25] also suggested that a spatiotemporal characteristic is a more relevant measure of postural control than a single-dimension characteristic. Their positions support our belief that dynamic postural stability outcome measures must be assessed cooperatively rather than comparatively. Our higher observed ICC value of the movement velocity measure (0.80) than that of the test duration measure (0.54) indicates that velocity was the more reliable temporal component of the dynamic postural stability construct. Additionally, given the present set of outcome measures, the NeuroCom LOS provides endpoint excursion, which evaluates the theorized LOS based on height, providing in relative terms the participant's comfortable LOS on initial movements. The Biodex DLOS offers a directional control measure on a changing base of support, giving information on functional balance capabilities but no actual values for determining LOS.

The relationship between the NeuroCom endpoint excursion and the Biodex directional control is more difficult to explain. One possibility for this result comes from the use of the 50% LOS target placement in the Biodex DLOS test. A person's COG is more stable and easier to control when it is central over the base of support because of the time and distance between the COG and the boundary of the base of support; as the COG moves toward the boundary of that base, control and stability are challenged.[25] The NeuroCom endpoint excursion variable quantifies the initial displacement of the participant's COG toward a target. Consistent with the suggestion by Hertel et al[25] about the spatiotemporal nature of postural stability, an uninjured person is likely to make an initial movement toward a target that is at least 50% of the hypothesized total distance, which, when transferred to the DLOS test, equals the target placement. The more distant target placement on the NeuroCom LOS leaves more chance for inaccurate predictions of movement and directional control that are less accurate as the COG is moved further to the edge of the base of support.

Owings et al[21] and Pai et al[14] supported this hypothesis of increasing error and movement predictions in their attempts to model stepping and perturbation behaviors. From a practical standpoint, this relationship may be an indication of effort, reflecting the challenge of the Biodex unstable platform and the initial movement distance used by the NeuroCom. These findings support theory related to concurrent validity. Outcome measures must predict dynamic balance and differentiate between the range of dynamic balance levels, such as the use of targets that challenge predictions of movement at all levels, thus supporting a criterion measure.

The lack of significant relationships between the NeuroCom outcome variables was unexpected. When evaluating dynamic postural stability, we anticipated that a specific pattern among the variables would emerge, particularly when similar testing instructions were provided to participants. Outcomes of dynamic postural stability modeling studies[7,8,21,23,25] indicating that multivariable assessments are needed to evaluate dynamic postural stability are consistent with our findings. We acknowledge that our sample size ($N=23$) and limited range of outcome measure values from healthy volunteers might have affected this outcome, although we are unaware of any published findings that assess the concurrent and construct validity of the computerized testing of dynamic postural stability.

Reliability of LOS tests should demonstrate a level of repeatability that instills confidence in the clinician or researcher using this information based on our current understanding of dynamic postural stability assessment. The level of reliability necessary to ensure confidence is subjective, but without any measure of reliability, the interpretation of data and of relationships between LOS outcome measures would be futile.

Several groups have conducted reliability studies using the NeuroCom LOS test. Brouwer et al[19] used 2 measures (movement time and path sway [directional control]) in the anterior, posterior, medial, and lateral directions and reported low ICC values, ranging from 0.18 to 0.48. Their study included 33 participants (age range, 20 to 32 years) and 3 test sessions at 1-week intervals, with set target distances at 75% of the predicted LOS distance based on height. Brouwer et al[19] assessed individual components of these directional measures, which would not reflect dynamic postural control as a whole. As noted previously, the poor performance or large variability on 1 sublet of the test may decrease ICC values and not represent the global dynamic balance. In contrast, we used composite scores of the outcome measures for reliability assessment. Participants' error variability is typically more sensitive in each component part than as an overall composite measure. Using more than 1 trial allows measurement error to balance out, so taking

the composite score (from more trials) results in less effect of error on the universal score. The SEM values from our study support this hypothesis.

Hageman et al[28] reported NeuroCom ICC values of 0.78 for path sway and 0.83 for movement time. Their test-retest study involved 12 healthy participants, 24 to 68 years of age, and 2 test sessions 1 week apart. Path sway and movement time were evaluated with target placement at 75% of the predicted LOS distance based on height. We used 4 trials across 2 test sessions, 1 week apart, and found ICC values of 0.55 and 0.69, respectively.

Both groups[19,28] set target placement at 75% of the predicted LOS distance based on height. In our study, the target was placed at 100% of the predicted LOS based on height. Use of the 75% target distance may not be a sufficient challenge for normal participants, which may result in skewed ICC values from the small ranges of ICC values for the given outcome measures. In contrast, the 100% target distance we used was the maximum LOS range possible with the NeuroCom device and provided a more difficult challenge to our volunteers. Results need to be transferable from one assessment session to the next, so that improvement can be monitored. Without sufficient challenge, continued improvement would not be measurable and would not provide the reliability evidence needed for criterion validity.

Our NeuroCom ICC values were comparable (0.55 to 0.82) with those observed in previous studies[19,24,28]; however, the NeuroCom LOS test can be conducted on at least 3 different systems sold by that manufacturer. The various systems use a freestanding long force plate, a short dual force plate design with open visual field, or a short dual force plate design with a visual surround. Because somatosensory, visual, and vestibular input can all affect postural stability, the reliability for each of these devices must be established independently. The data in our study were collected on the Smart Balance Master system with a gray-blue speckled pattern visual surround. The NeuroCom device used in the 3 previously referenced studies[19,24,28] was not specified.

To date and to our knowledge, no reliability measures for the Biodex DLOS test have been reported in the literature. Thus, our findings help establish the reliability evidence for the device and the test protocol, showing moderate test-retest reliability (ICC=0.70) for directional control and poor test-retest reliability (ICC=0.48) for test duration. These reliability data provide evidence that caution must be taken when the Biodex device is used to assess dynamic postural stability. Evaluation of the outcome measures' SEM values and confidence limit breadths indicates large amounts of unexplained error and a potential lack of sensitivity to detect change. Using various levels of platform stiffness could help reduce the variability of the data across participant pools and result in stronger ICC values, reducing extraneous error and improving test sensitivity.

Reference should also be made to the CIs and SEM values for test-retest reliability on the NeuroCom outcome measures. Although it is important to note that a healthy population displays small amounts of variability, which can be a rationale for smaller than expected ICC values, these values must still be considered in concert with the SEM and CI for each variable. Evaluation of these values indicates the presence or absence of precision for each outcome variable. We set CIs at 95%, meaning that if we conducted the same study repeatedly, the ICC value for a particular measure would fall within the calculated range 95% of the time. The breadth of the directional control CI (0.35, 0.74) points to a high potential for random error and a lack of test precision. For endpoint excursion (0.69, 0.91) and movement velocity (0.51, 0.95), the intervals remain in the moderate to high range and provide reassurance of test accuracy. Additionally, test sensitivity is supported by comparatively small SEM values. Individual scores that fall outside ±2 SEMs on repeated testing indicate true change in the abilities of the participant and demonstrate the sensitivity of the test to detect this change. Other authors have not reported these values, so direct comparisons cannot be made. The results of both statistical procedures are needed to assess the reliability and sensitivity of any outcome measure.

CONCLUSIONS

Based on the moderate to high test-retest reliability we observed, clinicians and researchers can be confident about the repeatability of LOS test results with the NeuroCom device in similar populations. Our results provide strong support for valid use of the NeuroCom device when assessing healthy people.

The interclass correlation assessment offered support for using particular measures that support a collaborative construct of dynamic postural stability assessment. Given the construct of dynamic postural stability, it appears that the NeuroCom LOS outcome measures provide more assessment information about dynamic postural stability than do the outcome measures of the Biodex DLOS. Because of the unexplained variance associated with significant interclass correlations between the NeuroCom and the Biodex outcome measures (ranging from 52.4% to 81.5%), we concluded that convergent validity was not supported, indicating that each system provided unique postural stability information. Indeed, the lack of strong relationships between variables on the same test suggest the need for more advanced dynamic postural stability models and conceptualizations for the way in which postural stability is defined and evaluated. We further suggest that given these findings, a more widespread agreement on postural stability terminology should be pursued by researchers and clinicians.

Of the 2 computerized posturography devices in this study, the NeuroCom demonstrated higher reliability as an assessment tool for LOS testing, but until researchers and clinicians can agree on a standardized definition of dynamic postural stability, neither the NeuroCom nor Biodex system can be considered the criterion standard for dynamic postural stability assessment. The validity and reliability of the postural stability component of human motor control warrant continued research.

ACKNOWLEDGMENTS

We used the following document in conducting the study and writing the manuscript: Moher D, Schulz KF, Altman D; CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285(15):1987–1991.

REFERENCES

1. Ag C. The influence of the lower extremities upon the structural integrity of the body. *J Am Osteopath Assoc*. 1950;49(9):443–446.
2. Nashner LM, McCollum G. The organization of human postural movements: a formal basis and experimental synthesis. *Behav Brain Sci*. 1985;8:135–150.
3. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test

(BESTest) to differentiate balance deficits. *Phys Ther.* 2009;89(5):484–498.
4. Ross SE, Guskiewicz KM, Gross MT, Yu B. Balance measures for discriminating between functionally unstable and stable ankles. *Med Sci Sports Exerc*. 2009;41(2):399–407.
5. Harringe ML, Halvorsen K, Renstrom P, Werner S. Postural control measured as the center of pressure excursion in young female gymnasts with low back pain or lower extremity injury. *Gait Posture*. 2008;28(1):38–45.
6. Nagy E, Toth K, Janositz G, et al. Postural control in athletes participating in an Ironman Triathlon. *Eur J Appl Physiol*. 2004;92(4–5):407–413.
7. Guskiewicz KM, Perrin DH. Effect of orthotics on postural sway following inversion ankle sprain. *J Orthop Sports Phys Ther.* 1996;23(5):326–331.
8. Hertel J, Buckley WE, Denegar CR. Serial testing of postural control after acute lateral ankle sprain. *J Athl Train*. 2001;36(4):363–368.
9. McKeon PO, Hertel J. Systematic review of postural control and lateral ankle instability, part 1: can deficits be detected with instrumented testing? *J Athl Train*. 2008;43(3):293–304.
10. Arnold BL, De La Motte S, Linens S, Ross SE. Ankle instability is associated with balance impairments: a meta-analysis. *Med Sci Sports Exerc*. 2009;41(5):1048–1062.
11. Eisenhauer MH, Ognibene J, Dutton S, McMahon K. The effects of proprioceptive perturbation training on postural sway and joint reaction times on healthy subjects [abstract]. *J Athl Train*. 2000;35(suppl 2):S52.
12. Ochsendorf DT, Mattacola CG, Arnold BL. Effect of orthotics on postural sway after fatigue of the plantar flexors and dorsiflexors. *J Athl Train*. 2000;35(1):26–30.
13. Saunders JB, Inman VT, Eberhart HD. The major determinants in normal and pathological gait. *J Bone Joint Surg Am.* 1953;35(3):543–558.
14. Pai YC, Maki BE, Iqbal K, McIlroy WE, Perry SD. Thresholds for step initiation induced by support-surface translation: a dynamic center-of-mass model provides much better prediction than a static model. *J Biomech*. 2000;33(3):387–392.
15. Stude DE, Brink DK. Effects of nine holes of simulated golf and orthotic intervention on balance and proprioception in experienced golfers. *J Manipulative Physiol Ther.* 1997;20(9):590–601.
16. Berg KO, Wood-Dauphine SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. *Can J Public Health*. 1992;83(suppl 2):S7–S11.
17. Holbein-Jenny MA, McDermott K, Shaw C, Demchak J. Validity of functional stability limits as a measure of balance in adults aged 23–73 years. *Ergonomics*. 2007;50(5):631–646.
18. Ford-Smith CD, Wyman JF, Elswick RK Jr, Fernandez T, Newton RA. Test-retest reliability of the sensory organization test in noninstitutionalized older adults. *Arch Phys Med Rehabil*. 1995;76(1):77–81.
19. Brouwer B, Culham EG, Liston RA, Grant T. Normal variability of postural measures: implications for the reliability of relative balance performance outcomes. *Scand J Rehabil Med*. 1998;30(3):131–137.
20. Riemann BR, Guskiewicz KM, Shields EW. Relationship between clinical and forceplate measures of postural stability. *J Sport Rehabil*. 1999;8(2):71–82.
21. Owings TM, Pavol MJ, Foley KT, Grabiner MD. Measures of postural stability are not predictors of recovery from large postural disturbances in healthy older adults. *J Am Geriatr Soc*. 2000;48(1):42–50.
22. Hinman MR. Factors affecting reliability of the Biodex Balance System: a summary of four studies. *J Sport Rehabil*. 2000;(3):240–252.
23. Karlsson A, Frykberg G. Correlations between force plate measures for assessment of balance. *Clin Biomech (Bristol, Avon)*. 2000;15(5):365–369.
24. Liston RA, Brouwer BJ. Reliability and validity of measures obtained from stroke patients using the Balance Master. *Arch Phys Med Rehabil*. 1996;77(5):425–430.
25. Hertel J, Olmsted-Kramer LC, Challis JH. Time-to-boundary measures of postural control during single leg quiet standing. *J Appl Biomech*. 2006;22(1):67–73.
26. Ashton-Miller JA, Wojtys EM, Huston LJ, Fry-Welch D. Can proprioception really be improved by exercises? *Knee Surg Sport Traumatol Arthrosc*. 2001;9(3):128–136.
27. Safrit MJ, Wood TM, eds. *Measurement Concepts in Physical Education and Exercise Science*. Champaign, IL: Human Kinetics; 1989:21–72, 117–152.
28. Hageman PA, Leibowitz JM, Blanke D. Age and gender effects on postural control measures. *Arch Phys Med Rehabil*. 1995;76(10):961–965.
29. *Biodex Balance System Operation and Service Manual #945-300.* Shirley, NY: Biodex Medical Systems; 1999.
30. *Smart Balance Master Operator's Manual*. Version 7.0. Clackamas, OR: NeuroCom International, Inc; 1999.
31. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.
32. Baumgartner TA, Jackson AS. *Measurement for Evaluation in Physical Education and Exercise Science*. 4th ed. Dubuque, IA: Wm C Brown/McGraw-Hill; 1999:75–134.
33. Ramsey, FL, Schafer DW. *The Statistical Sleuth*. Belmont, CA: Duxbury Press; 1997:27–80.

Address correspondence to Marie L. Pickerill, PhD, ATC, Department of Kinesiology, DePauw University, 702 South College Street, Greencastle, IN 46135. Address e-mail to mariapickerill@depauw.edu.