

Assessing the Reliability, Correlation, and Validity of Commonly Used Rehab Tools to Measure Balance

Darcy Dzurino, SPT; Nicole Dawson, PT, PhD, GCS; Melissa Karleskint, SPT; Jennifer Tucker PT, DPT, PCS
University of Central Florida, Department of Health Professions, Doctor of Physical Therapy Program

PURPOSE

Balance and postural control are essential to ensuring not only safe activities of daily living for individuals, but for the performance of safe locomotion in general.¹ There is a lack of reliability and validity data supporting the utilization of any one method as the best objective tool to capture a comprehensive balance component of a musculoskeletal and neuromuscular examination. Functional performance measures such as the Four Square Step Test (FSST)² and the Timed Up and Go (TUG)³ have been found to have clinical utility when assessing various parameters, such as balance and postural control, in different planes of motion, while also assessing fall risk. These measures have easily been included as part of the routine musculoskeletal examination by physical therapists worldwide. Although these measures have been used consistently in clinical practice, new methods of measuring these same variables have been introduced into rehabilitation settings, such as the Biodex Balance System. The Biodex SD Stability System has been shown to be a reliable assessment tool for postural stability. However, its ability to provide an accurate representation of balance has not been compared to functional performance measures, such as FSST and TUG. Therefore, the purpose of this study is to investigate reliability, internal consistency, and construct validity of FSST, TUG, and Biodex SD (LOS and m-CTSIB).

METHODS

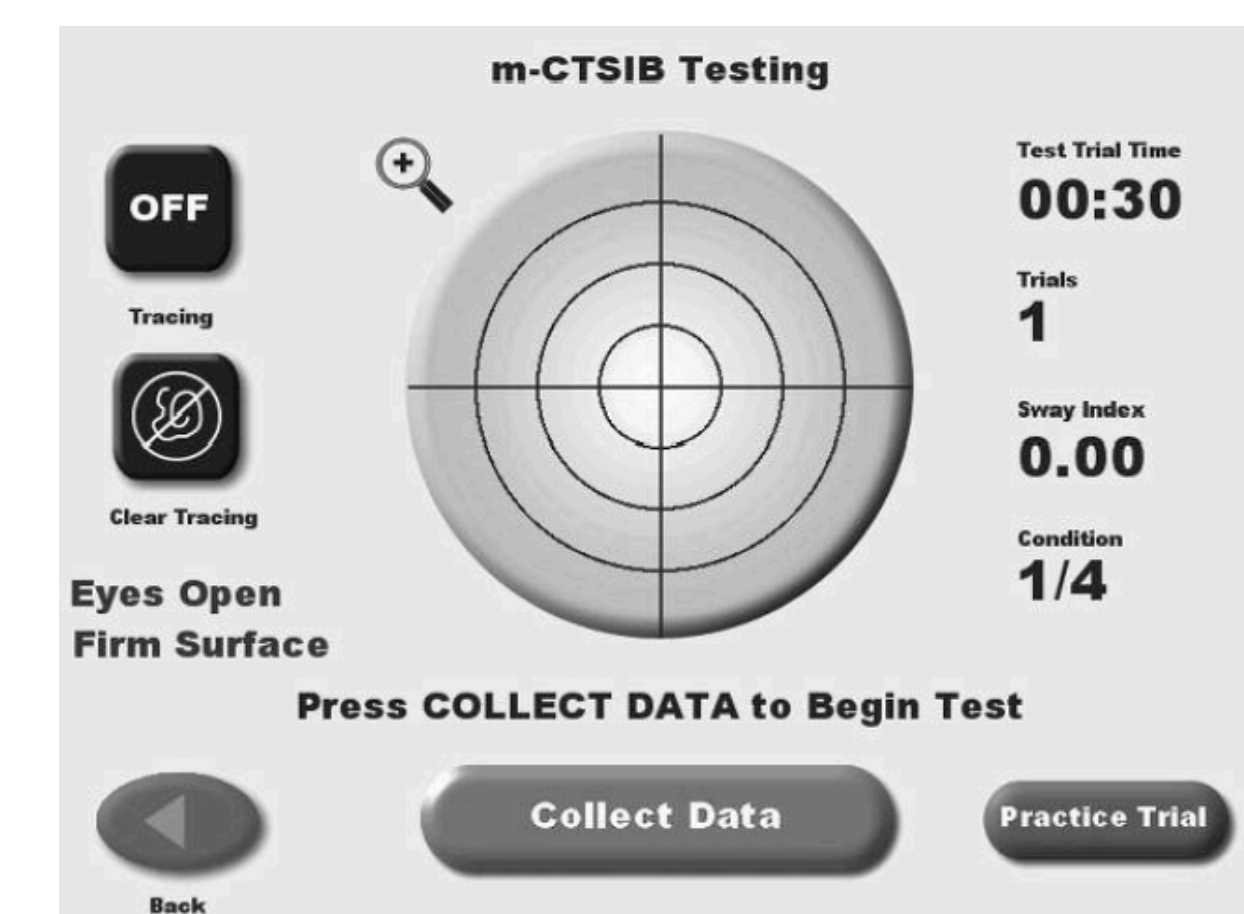
An observational reliability and validity study was conducted in which a convenience sample of 105 healthy adults, 77 females and 28 males, mean age 24.5 years old (\pm 4.66 SD) performed balance assessments including the FSST, TUG, Biodex SD Limits of Stability (LOS) and modified Clinical Test of Sensory Organization and Balance (m-CTSIB). For LOS, the overall percentage and test duration were recorded. For m-CTSIB, the overall Sway Index (SI) was recorded. Condition 1 of the m-CTSIB represented simple postural stability.

ANALYSIS

Data analysis was performed with SPSS version 22. Descriptive statistics were determined for age, gender, education level, and number of LE injuries in the past. Intraclass correlation coefficients (ICC[3,1]) determined test-retest reliability for each of the 6 variables. Values below 0.50 demonstrated poor reliability, with values between 0.50-0.75 represented moderate reliability, values above 0.75-1.00 demonstrated good reliability. Pearson's product, r , was used to determine construct validity for each measure to identify whether they assessed similar or unique components of balance.



Biodex SD LOS Testing Screen Display



Biodex SD m-CTSIB Testing Screen Display



Biodex Balance System SD (Biodex Medical Systems, Shirley NY)

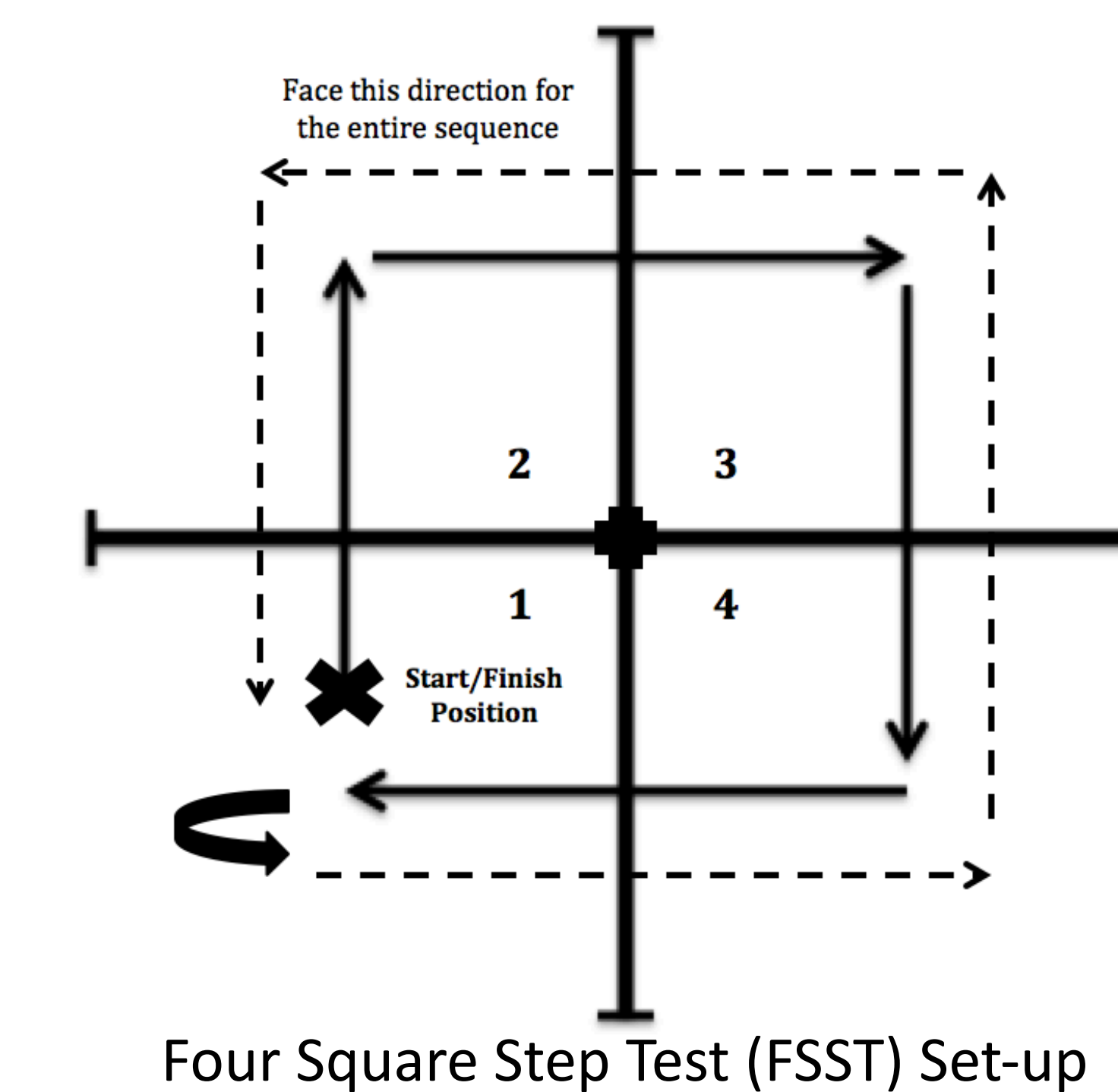
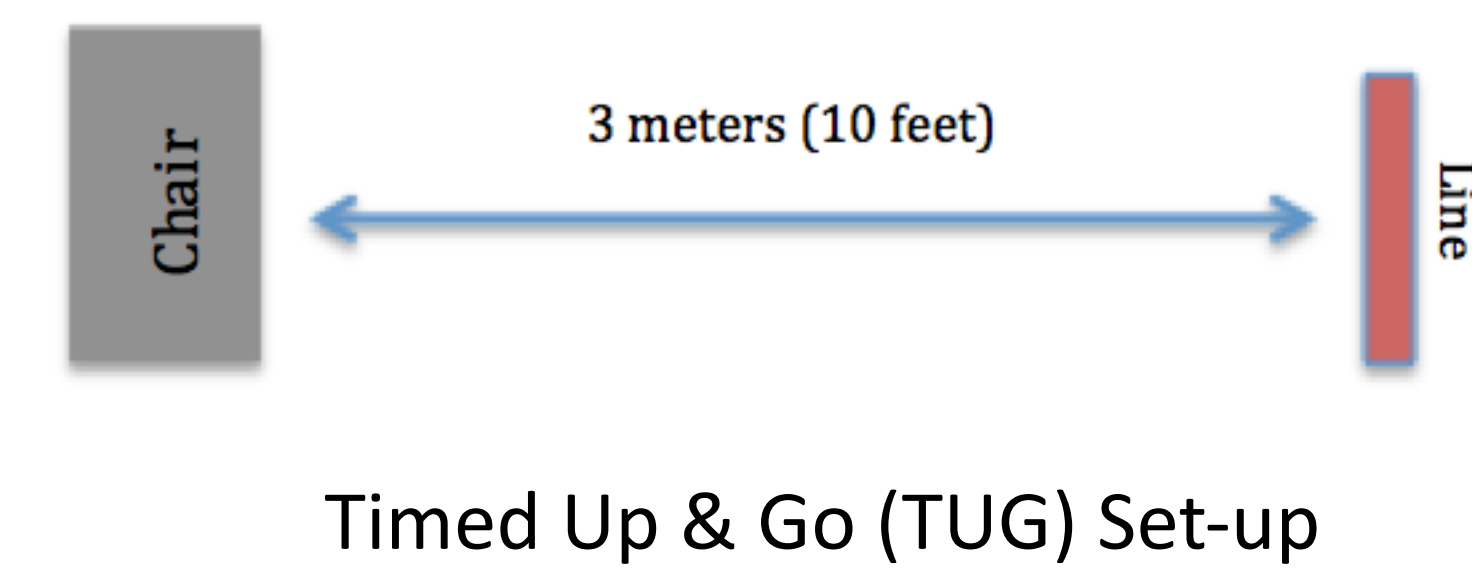


Table 1. Test-retest reliability of balance measures. (N=105)

Balance Assessment	ICC (3,1)	Descriptor (Portney & Watkins, 2009)
TUG	0.88	Good
FSST	0.92	Good
Biodex SD LOS – Overall Percentage	0.83	Good
Biodex SD LOS – Test Duration	0.58	Moderate
Biodex SD m-CTSIB – Stability Index	0.75	Good
Biodex SD m-CTSIB – Condition 1	0.24	Poor

Note. *ICC = intraclass correlation coefficient

Table 2. Intercorrelations of balance measures examining validity. (N=105)

	TUG	FSST	Biodex LOS	Biodex m-CTSIB (Stability Index)	Biodex m-CTSIB (Condition 1)
TUG	1				
FSST	.14	1			
Biodex SD LOS – Overall Percentage	-.07	-.15	1		
Biodex SD m-CTSIB – Stability Index	.22*	.14	-.09	1	
Biodex SD m-CTSIB – Condition 1	.09	.21*	-.01	.41**	1

Note. TUG = Timed Up & Go; FSST = Four Square Step Test; LOS = Limits of Stability; m-CTSIB = Modified Clinical Test of Sensory Organization and Balance. * $p < .05$; ** $p < .001$.

RESULTS

Biodex SD LOS overall percentage, TUG, and FSST showed excellent test-retest reliability (ICC [3,1] = .83, .88, .92 respectively), while the Biodex SD m-CTSIB SI demonstrated strong test-retest reliability (ICC [3,1] = .75) indicating that these assessments can be repeated on separate occasions and still have agreeable results. The LOS test duration showed moderate test-retest reliability (ICC [3,1] = .58) and the m-CTSIB condition 1, showed poor test-retest reliability (ICC [3,1] = .24) indicating that these measures do not demonstrate good repeatability when it comes to using these assessments to score the same subject on different occasions. Intercorrelations (using Pearson's r) between measures ranged from .15 to .22 indicating poor construct validity among all measures indicating that these assessments are measuring completely different aspects of balance and are not able to be used interchangeably to determine a patient's balance assessment.

DISCUSSION

As new assessment methods are introduced into clinical practice, there is a clinician demand to establish a single evaluate construct that defines balance. Theoretically, objective tools to measure balance should have variables that correlate highly, or at least moderately, with one another to demonstrate construct validity of an assessment of an individual patient's balance. However, the commonly used assessment measures that were examined in the current study revealed poor construct validity indicating that each tool assessed completely different components of postural stability and balance in this sample of participants. Additionally, reliability of these balance tests should demonstrate a level of repeatability that promotes confidence in the utilization of these tools for the clinician or researcher wanting to perform a complete dynamic postural stability assessment.⁴ Findings from the current study found each measure to be reliable, however, left uncertainty regarding the particular construct being measured. These data contribute significantly to the current body of literature identifying excellent test-retest reliability for FSST and TUG while strengthening reliability data of Biodex SD LOS and m-CTSIB tests. However, more interestingly, the study revealed poor construct validity between measures indicating the selected methods of balance assessment examine unrelated constructs.

CLINICAL RELEVANCE

Findings from this study have clinical implications warranting further discussion in efforts to assist clinicians and researchers in selecting the most appropriate tools. Clinicians and researchers should be deliberate when choosing a balance assessment tool. It is important to take into account that while all of these outcome measures do look at components of balance, not one of them can serve as the single evaluate construct of balance itself. Balance is complicated and it is recommended that clinicians understand this as we encourage the utilization of multiple balance assessment tools to capture the entire picture of their patient's balance.

REFERENCES

- Arifin N, Osman NAA, Abas WABW. Intrarater Test-Retest Reliability of Static and Dynamic Stability Indexes Measurement Using the Biodex Stability System During Unilateral Stance. *Journal of Applied Biomechanics*. 2014;30:300-304.
- Dite, W. and Temple, V. A. (2002). "A clinical test of stepping and change of direction to identify multiple falling older adults." *Arch Phys Med Rehabil* 83(11): 1566-1571.
- Podsiadlo D, Richardson S. The timed "Up & Go": A test of basal functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*. 1991;39(2):142-148.
- Pickering ML, Harter RA. Validity and reliability of limits-of-stability testing: A comparison of 2 postural stability evaluation devices. *Journal of Athletic Training*. 2011;46(6):600-606.