OBSERVATIONAL MEASUREMENT STUDY

Novice Versus Experienced Rater Reliability of the Mini Balance Evaluation Systems Test (Mini-BESTest) in Patients with Acquired Brain Injury (ABI)

Charity E Maqueda, PT, DPT, ATC, LAT[1,2,3]; Rupal Patel, PT, MS[1]
[1]Texas Woman's University, Houston, TX
[2]TIRR Memorial Hermann, Houston, TX
[3]University of Tennessee Medical Center, Knoxville, TN

ABSTRACT

Objectives: The purpose of this study was to examine novice versus experienced rater's inter-rater and test retest reliability when scoring the Mini-BESTest in patients with acquired brain injury (ABI); and to provide clinically relevant information regarding the reliability and clinical feasibility of using the Mini-BESTest in a population with ABI. *Background:* Individuals who sustain an ABI often have many functional deficits including impaired balance. Currently there is no gold standard for assessing balance in individuals with ABI. Many balance measures have been shown to have ceiling effects and inconsistent reliability results when measured by novice physical therapists verses experienced physical therapists. *Methods and Measures:* Participants were selected using convenience sampling meeting specific inclusion and exclusion criteria. Each was video recorded while completing the Mini-BESTest, administered by one test-administrator who was not a rater. Inter-rater reliability was tested using four raters (two novice, two experienced), whom simultaneously and independently observed and rated performance from the same angle. All raters were blinded to others' scores. Test retest reliability was tested using three raters (two novice, one experienced), whom independently viewed video recording in random order approximately four weeks later. All raters were blinded to each other's scores and their own scores. *Results:* There was a significant intra-class correlation coefficient (p < .001) demonstrating excellent inter-rater reliability and a significant Pearson's correlation (p < .001) demonstrating high test retest reliability among all raters. Novice raters demonstrated slightly higher test retest reliability (r=.953) than the experienced raters (r=.933). *Conclusion:* Both novice and experienced raters demonstrated excellent inter-rater and test retest reliability when rating individuals with ABI using the Mini-BESTest. The Mini-BESTest is a reliable tool, which uses multiple constructs to assess balance deficits in 4 specific constructs. It can be used by both novice and experienced physical therapists to guide interventions for the impaired systems. Increased time to administer and modifications to verbal instructions may be appropriate for use in the ABI population. Follow-up studies are necessary to determine the extent and validity of instruction modifications.

Background

Clinical assessment of functional mobility following an acquired brain injury (ABI) is standard protocol in rehabilitation. ABI includes any brain injury attained after birth due to traumatic event, stroke, tumors or brain illness. It does *not* include congenital or degenerative diseases.[1] Brain injury can result in both acute and long-term effects on functional balance. Impaired balance control is the most common functional deficit in individuals with stroke and is associated with activity limitations, reduced functional gait, and decreased participation in society.[2] Balance is also a significant predictor of falls and need for long term care.[2] Balance is multifactorial, comprised of various constructs including static and dynamic balance, transitions, anticipatory responses,

reactive postural control, sensory orientation and dynamic stability in stance and gait.[3,4] Improving balance is an essential piece of the post-injury rehabilitation process and is continuously evolving throughout the course of recovery. Measuring balance constructs is clinically useful to assess progress and to guide treatment.

The Neurology Section of the American Physical Therapy Association (APTA) developed the StrokEdge and TBI Edge tools to provide therapeutic guidelines for clinicians;[5] yet currently no gold standard measure has been established to assess balance in individuals with ABI. Common functional balance measures[5-14] currently used in higher functioning ABI populations are: Balance Error Scoring System (BESS),[6,7] Berg Balance Scale (BBS),[8,9] Community Balance and Mobility Scale (CB&M),[10-12] and the High-Level Mobility Assessment Tool (HiMAT).[13,14]

However, there are limitations to each measure such as variability in balance constructs tested[15] and ceiling effect when used with people with traumatic brain injury (TBI). [8-14] Furthermore, not all measures systematically evaluate all balance constructs and multiple measures must be used to gain an accurate clinical picture.[9]

The 36-item BESTest was introduced by Horak and colleagues to assess six constructs (Biomechanical constraints, stability limits and verticality, anticipatory postural adjustments, postural response, sensory orientation, stability in gait) underlying balance deficits.[16] Scoring is based on a four point ordinal scale (0 -3) with higher score (max score of 108) indicating poorer balance. Sub-scores for each construct can be used to target interventions to specific deficits. The BESTest has shown excellent interrater reliability and validity in community dwelling individuals with and without

balance disorders (ICC = .91),[16] Parkinson's Disease (ICC = .91),[17] and most recently in patients with subacute stroke (ICC = .99).[18] The time it takes to administer (30 – 60 minutes), is a major drawback to using the BESTest clinically.[16, 19-21]

Franchignoni and colleagues developed a shorter fourteen-item version, the Mini-BESTest,[19, 22] eliminating ten items and two constructs (biomechanical constraints and stability limits) using factor and Rasch analysis. The scoring was modified to a three-point ordinal scale (0-2), with maximum score of twenty-eight and time to administer was shortened to ten to fifteen minutes. [21] The Mini-BESTest demonstrated excellent psychometric properties in a sample of 115 adults with diverse neurological conditions in an inpatient setting.[19] Furthermore, a recent Rasch validation study by Franchignoni, further supports the Mini-BESTest as a clinically useful tool for measuring dynamic balance for individuals with stroke and other neurological conditions.[22]

Since then, several studies have examined the utility, concurrent validity, responsiveness and reliability of the Mini-BESTest on individuals with balance disorders.[2, 4, 23-25] The Mini-BEST was found to have excellent interrater reliability (ICC = .91) and test retest reliability (ICC = .88) in individuals with Parkinson's Disease;[23] and excellent interrater reliability (ICC = .96-.99) in individuals with stroke.[2, 24] The Mini-BESTest also showed very good concurrent validity with the Berg Balance Scale (BBS, r=.85).[4] There is conflicting evidence to support the use of the Mini-BESTest as an indicator of fall risk.[2, 25] Although literature on the Mini-BESTest shows high reliability, support for its use for people with ABI that have cognitive impairments is limited.[24]

There is variation in the amount of experience or training of clinicians that use

balance measures such as the Mini-BESTest. Previous studies have used student or novice raters and clinicians with varying levels of experience. [2,16,17,23,24] Clinical experience may influence ratings of the Mini-BESTest regardless of previous use of measure. Yet no other research has differentiated reliability of novice versus experienced raters, when neither has previous experience in the use of this measure in evaluating balance in individuals with ABI. "Novice" has previously been defined as an individual new to using the MiniBESTest,[24] yet no study to date has defined novice as current entry-level physical therapy students.

Previous reliability studies on the MiniBESTest primarily utilized live testing one to sixteen days apart to determine inter-rater or intra-rater reliability.[2,4,23, 24] In this case, reliability could be skewed, because of the learning effect of both participants and raters. Increasing the time frame between tests could decrease the stability of the response variable, and the possibility of participant recovery may also skew the data. To minimize this, Dahl and Jorgensen used video recordings to assess intra- and interrater reliability of the Mini-BESTest in people with stroke, resulting in excellent reliability for both. However, raters were given the ability to watch the video clips several times prior to finalizing their score. [24]

Purpose

The purpose of this study was twofold: (1) to examine novice versus experienced rater's interrater and test retest reliability of the Mini-BESTest in people with ABI, and (2) to provide modification to instructions of the Mini-BESTest that could be utilized in a population with ABI and cognitive deficits.

Methods

Setting

The study was conducted from April – May 2013 at a post-acute community re-entry rehabilitation facility in a large urban city in the USA.

Participants

Current adult patients or patients discharged in the past twelve months were identified as potential participants by clinicians based on study criteria. Inclusion criteria included: >18 years of age, diagnosis of acquired brain injury, able to follow two-step commands, walk at least twenty-five feet, and endurance to perform balance tasks for twenty to twenty-five minutes. Participants were excluded if they had: any recent medical procedures, conditions that might prevent them from performing test related functional tasks or if they had cognitive or communication impairments that limited their ability to understand English instructions and execute the tasks specified for the test. Thus, individuals with severe aphasia were excluded. For participants with TBI, cognitive impairment was operationally defined as individuals determined to have equal or greater function of Level VI on the Rancho Los Amigos Level of Cognitive Functioning Scale.[27]

Using consecutive convenience sampling of those meeting inclusion criteria, twenty-five participants were recruited, and scheduled for testing by facility staff. This study was approved by the Institutional Review Board (IRB) of Texas Woman's University, Houston, TX, and all participants gave written informed consent. Demographic information is provided in Table 1.

Raters

There were four raters; two experienced and two novice raters. The experienced raters were licensed physical therapists, one with twenty-three years of clinical experience and the other with more than four years of clinical experience. The novice raters were second year Doctor of Physical Therapy (DPT) students with seven weeks of clinical experience. Neither experienced raters nor novice raters had previously used the Mini-BESTest clinically. Raters were blinded to each other's data.

Test Administrator

One person, who was not a rater, administered the Mini-BESTest to every participant using standardized instructions. After initial practice, it was evident that reactive postural control and dynamic gait tasks required demonstration. Therefore, the test administrator demonstrated tasks for each participant to maximize comprehension of the task. Clinical reasoning was used to determine if further demonstration was necessary and was allowed at any time upon request of participant.

Mini-BESTest Modifications

The Mini-BESTest scoring was modified to match the published clarifications by King and Horak,[26] and the instructions were standardized to two-step commands by adding pauses and breaking down the original phrases to accommodate for cognitive deficits in this population. The phrase "Please listen first" was added at the beginning of each task item to limit impulsivity of some participants with ABI, who may begin the task as the test administrator began speaking versus waiting until full instructions were given (Appendix A).

Materials needed to perform the Mini-BEST test included: a quiet room with a 25ft walkway, stop-watches, a gait belt, one firm chair available within reach of the test administrator (for participant rest breaks as needed), a second firm chair with arm rests for item fourteen, tape to measure a 3-meter distance on the floor from the chair, cones to mark every 5 feet, two stacked shoe boxes, 4-inch thick, medium-density Temper® foam, and a 10 degree incline ramp (at least 2 x 2 ft.) (foam and incline ramp fromhttp://bestest/).[28]

Procedures

Prior to data collection, all raters and the test administrator watched a training video of the original BESTest to familiarize themselves with scoring the test items.[28] The test administrator practiced using the modified instructions on individuals without ABI and the raters practiced rating these individuals. During live testing the raters were positioned at the same level and distance as the camera, approximately 15ft. from the test administrator and participant. Raters viewed the testers from the front for Anticipatory, Reactive Postural Control, and Sensory Orientation items except "Rise to Toes," which was viewed laterally. All Dynamic Gait test items were viewed laterally. One rater served as a second pair of hands for safe guarding participants with severe balance impairments, particularly for "Single Leg Stance" (Item 3) and all Reactive Postural Control items (Items 4-6). For item fourteen, all participants first performed the timed up and go (TUG), their time was recorded by the raters, then each attempted to count backward by three's. If participants were unable to accurately count backward by three's, they were scored as a zero for this item.[26] Participants were then asked to count backward by two's or if unable by one's, then performed a dual-task TUG to assess their ability to perform a cognitive dual-task. This score was documented, however, it was not used as

part of the total score on the Mini-BESTest. This would have significantly modified the instructions and scoring.

Interrater reliability was measured among raters during the live testing. Four raters (two novice student raters, two experienced clinicians) simultaneously and independently observed and rated performance from the same viewing angle. Test retest reliability was measured using video playback approximately four weeks after initial participant performance. Three raters (two novice student raters, one experienced clinician) independently viewed video recordings in random order. Raters were blinded to their own previous scores and others' scores throughout.

Data Analysis

All statistical analyses were performed using SPSS 19.0 software *(SPSS Inc, Chicago, Illinois). A priori* power analysis was conducted using G*Power (version 3.1.7), the minimal sample size needed was twenty-three participants for test retest correlation. Interrater reliability of the sum total score on the Mini-BESTest was determined using an intraclass correlation coefficient (ICC [3,1]) and 95% CI. Item agreement was analyzed using Cohen's kappa (κ) and percent agreement. Both were calculated among all combinations of raters (e.g., $Novice_1$ to $Novice_2$, $Experienced_1$ to $Novice_1$, etc.) as well as the average percent agreement for each given pair.

Test-retest reliability was assessed using ICC values after two novice raters and one experienced rater reviewed the video recording (at *Time 2*) of the same participants completing the same trial (*Time 1*) of the Mini-BESTest. Pearson's *r* was used to compare each rater's score from time one to time two (live vs. video scores).

Results

Inter-Rater Reliability

- **Total Score Interrater Reliability**

Each rater scored each participant's performance, and total scores were compared between each pair of raters' to determine interrater reliability. Interrater reliability of the sum total score on the Mini-BESTest was excellent, ICC .90 - .97 (Table 2).

- **Item Agreement & Internal Consistency**

The degree of interrater reliability for each Mini-BESTest item was examined further for scoring similarities and differences among raters. Cohen's kappa (κ) and percent agreement were calculated among all combinations of raters. Novice raters had slightly higher agreement ($\kappa = .697$) compared to experienced raters ($\kappa = .687$). Percent agreement of individual items and subscale scores was fair to moderate, 84.6% and 84.0% for novice and experienced raters respectively.

Furthermore, individual item comparisons provided clinically important information about which items were scored more consistently than others (Table 3). Figure 1 provides a visual of every test item's inter-rater reliability. The items with the highest reliability are 7: Standing with eyes open on a firm and flat surface and 8: Standing with eyes closed on foam surface. Two items which illustrated the lowest reliability are 11: Walking with horizontal head turns and 12: walking with pivot turns.

Internal consistency of the Mini-BESTest was assessed using Cronbach's Alpha (α) and standard error of measurement (SEM). All four raters have good internal consistency (Table 4), suggesting that their

assessments of each item on the Mini-BESTest were similar. Among raters, novice raters showed slightly less internal consistency than experienced raters.

Test-Retest Reliability: Total Score

Test-retest reliability was high, ICC .88-.97, among all raters from Time one viewing (live) to Time two viewing four weeks later via video recording (Table 5). Figure 2 shows the test-retest reliability using the Pearson's (r) product moment correlation for the two novice raters, the one experienced rater, and all the raters combined. All the raters had similar ratings using the Mini-BESTest, $r = .95$ ($p < .001$), novice raters had higher agreement than the experienced rater, $r=.953$, $r=.933$, respectively.

Summary of Results

In summary, there was excellent percent agreement of rating items, 83.6% and interrater reliability in novice versus experienced raters, ICC = .909 - .970 ($p < .001$). A significant Pearson's r correlation was found for all raters, $r = .946$ ($p < .001$) demonstrating high test retest reliability among all raters. In addition, novice raters demonstrated slightly higher test retest reliability $r = .953$ than the experienced rater $r = .933$.

Discussion

Interrater, Test Retest Reliability & Item Agreement

Based on these results, there was not a significant difference between raters for interrater reliability and test retest reliability. Our findings are consistent with findings reported by previous studies (Table 4). Our test retest procedure included the use of video recording after a four-week interval to prevent maturation effect; leading to

stability in the response variable. Additionally, raters were blinded to their previous scores, minimizing rater recall; which strengthens the internal validity.

The item agreement found in this study compared to other studies indicates[2,24] there was perfect agreement among all raters for items seven[2] and eight.[24] Both items are low to mid-level tasks requiring a narrow base of support on firm surface with eyes open (7), then on foam surface with eyes closed (8). Clinically one might consider the ceiling effect of each single item, and whether it is necessary to assess within this set of tasks. Conversely, item eleven: "walking with horizontal head turns," had the lowest agreement in our study and other studies.[2, 24] One conjecture may be that this item is used in other outcome measures (Dynamic Gait Index,[29] BESTest[16]) with a different scale for scoring (0 - 3). Raters may have been biased if they had previously used the DGI or BESTest scoring, as the Mini-BESTest scale (0-2) is newer. There was fair to adequate agreement for item twelve: "walking with pivot turns," $\kappa = .450$. This may be attributed to the test administrator's demonstration error, causing participants to perform the task as demonstrated compared to performing per test instructions, which may have led to decreased agreement in scoring by the raters.

Clinical Implications

This study used the recently updated Mini-BESTest scoring,[26] and modified the instructions to accommodate cognitive deficits of participants with ABI. While novice raters had been used in previous studies of the Mini-BESTest,[2,23,24] this study is the first to differentiate results between novice/student raters and experienced/clinician raters, for each test item and overall score. These findings illustrate that the Mini-BESTest is reliable, can be administered independently from

theBESTest, and can be used by both novice student raters and experienced clinicians. While previous studies included multiple diagnoses,[2,4,19] this study only included individuals with ABI and used the Rancho Scale to quantify cognitive abilities of the subset of individuals with TBI.

Previously, test administration time was reported to be between ten to fifteen minutes.[19,21] Yet, with instructions, set up, and modifications required for this sample of people with ABI, average time to administer was twenty minutes per participant. Modifications to accommodate for cognitive deficits included pauses, breaking the original instructions into two-step commands, and demonstration of test items by a single test administrator.

Limitations

A video training resource for the Mini-BESTest was not yet available at the commencement of this study. Therefore, the original BEStest training video was used to train raters to use the Mini-BESTest. However, the BESTest uses a four-point scoring scale (0-3) vs. the three-point scale (0-2) of the Mini-BESTest, which could have affected the ratings. Now, training videos are available for free through www.bestest.us to assist with administering and scoring the Mini-BESTest.[28] Due to convenient sampling, our sample included mostly individuals with stroke; thus, findings cannot be generalized to all individuals with ABI at this time. Additional limitations included: learning effect by the raters, potential instrumentation effect due to test administrator demonstration error (e.g. Item twelve), and use of an additional person guarding during certain test items. Reactive measurement effect may be due to the presence of a video camera, causing the participants' performance to vary from their norms. A limitation to this study is that there were no baseline measurements or age

normative values for comparison. Since the completion of data collection and analysis, a recent study reports normative values for the Mini-BESTest among individuals fifty years or greater.[30] Lastly, due to long periods of testing, rater and test-administrator fatigue could have influenced test demonstration and ratings.

Future Considerations

Upon completion of this study there is an evident need to standardize instructions and administration of the Mini-BESTest for use with cognitively impaired adults with ABI. Potential factors contributing to weak agreement of test items between novice and experienced raters should be examined. Additionally, many participants in our sample were challenged by item fourteen, the cognitive dual-task TUG. The relationship between dual-tasks and fall risk or fall history in people with ABI should be further examined. Also, additional studies are needed to further establish age and gender based normative values for the Mini-BESTest so that clinicians can compare their patient's performance to established norms.[30]

Conclusion

The Mini-BESTest demonstrates excellent interrater and test retest reliability when rating individuals with ABIs, proving to be a reliable tool to assess balance deficits for multiple constructs. Novice (student) raters were found to have slightly higher agreement among scores than experienced (clinician) raters. This study shows that the Mini-BESTest can be used by both novice and experienced physical therapists to guide balance rehabilitation. Additional research is needed to further support the Mini-BESTest's use with ABI populations, specifically in individuals with cognitive deficits and/or traumatic brain injury (TBI).

References

1. Acquired Brain Injury: A web portal for and by brain injury survivors. Retrieved from http://www.acquiredbraininjury.com/abi_manual. Updated 2015. Accessed April 13, 2015.
2. Tsang C, Liao L, Chung R, Pang M. Psychometric Properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in Community-Dwelling Individuals With Chronic Stroke Phys Ther. April 2013;98:1102-1115. doi: 10.2522/ptj.20120454. PMID: 23559522.
3. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med. 2010 Apr;42(4):323-31. doi: 10.2340/16501977-0537. PMID: 20461334.
4. Godi M, Franchignoni F, Caligari M, et al. Comparison of reliability, validity, and responsiveness of the Mini-BESTest and Berg Balance Scale in patients with balance disorders. Phys Ther. 2013;93:158– 167. doi: 10.2522/ptj.20120171. PMID: 23023812
5. APTA Neurology Section. Neurology section outcome measures recommendations. Retrieved from http://www.neuropt.org/professional-resources/neurology-section-outcome-measures-recommendations. Updated 2011. Accessed April 13, 2015.
6. Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error scoring system. Sports Health. 2011 May;3(3):287-95. PMID: 23016020.
7. Guskiewicz KM. Balance Error Scoring System. University of North Carolina's Sports Medicine Research Laboratory, Chapel Hill, NC.
8. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health. 1992 Jul-Aug;83 Suppl 2:S7-11. PubMed PMID: 1468055.
9. Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. Phys Ther. 2008 May;88(5):559-66. doi: 10.2522/ptj.20070205. Epub 2008 Feb 21. Review. PubMed PMID: 18292215.
10. Inness EL, Howe JA, Niechwiej-Szwedo E, Jaglal SB, McIlroy WE, Verrier MC. Measuring Balance and Mobility after Traumatic Brain Injury: Validation of the Community Balance and Mobility Scale (CB&M). Physiother Can. 2011 Spring;63(2):199-208. doi: 10.3138/ptc.2009-45. PMID: 22379260.
11. Howe JA, Inness EL, Venturini A, Williams JI, Verrier MC. The Community Balance and Mobility Scale--a balance measure for individuals with traumatic brain injury. Clin Rehabil. 2006 Oct;20(10):885-95. PMID: 17008340.
12. Community Balance &Mobility Scale: Administration and Scoring. Toronto Rehabilitation Institute. 1998. Revised May 2011.
13. Williams GP, Robertson V, Greenwood KM, Goldie PA, Morris ME. The high-level mobility assessment tool (HiMAT) for traumatic brain injury. Part 2: content validity and discriminability. Brain Inj. 2005 Sep;19(10):833-43. PMID: 16175843.
14. Williams GP, Greenwood KM, Robertson VJ, Goldie PA, Morris ME. High-Level Mobility Assessment Tool (HiMAT): interrater reliability, retest reliability, and internal consistency. Phys Ther. 2006 Mar;86(3):395-400. PMID: 16506875.
15. Horak, FB.; Macpherson, JM. Postural orientation and equilibrium. In: Shepard, J.; Rowell, L., editors. Regulation and Integration of Multiple Systems Handbook of Physiology: Section 12, Exercise. New York: Oxford University Press; 1996. p. 255-292.
16. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther. 2009 May;89(5):484-98. doi: 10.2522/ptj.20080071. PMID: 19329772
17. Leddy AL, Crowner BE, Earhart GM. Functional Gait Assessment and Balance Evaluation System Test: Reliability, Validity, Sensitivity, and Specificity for Identifying Individuals With Parkinson Disease Who Fall. Physical Therapy. 2011;91(1):102-113. doi:10.2522/ptj.20100113.
18. Chinsongkram B, Chaikeeree N, Saengsirisuwan V, Viriyatharakij N, Horak FB, Boonsinsukh R. Reliability and validity of the Balance Evaluation Systems Test (BESTest) in people with subacute stroke. Phys Ther. 2014 Nov;94(11):1632-43. doi: 10.2522/ptj.20130558. Epub 2014 Jun 12. PubMed PMID: 24925073.
19. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med. 2010 Apr;42(4):323-31. doi: 10.2340/16501977-0537. PMID: 20461334.
20. Padgett PK, Jacobs JV, Kasser SL. Is the BESTest at its best? A suggested brief version based on interrater reliability, validity, internal consistency, and theoretical construct. Phys Ther. 2012 Sep;92(9):1197-207. doi: 10.2522/ptj.20120056. Epub 2012 Jun 7. PubMed PMID: 22677295.
21. Rehab Measures Database. Instruments: Balance Evaluation Systems Test and Mini Balance Evaluation Systems Test. Retrieved from http://www.rehabmeasures.org/rehabweb/allmeasures.aspx?PageView=Shared. Updated August 2014. Accessed April 13, 2015.
22. Franchignoni F, Godi M, Guglielmetti S, Nardone A, Giordano A. Enhancing the usefulness of the Mini-BESTest for measuring dynamic balance: a Rasch validation study. Eur J Phys Rehabil Med. 2015 Feb 5. [Epub ahead of print] PubMed PMID: 25653078.
23. Leddy AL, Crowner BE, EarhartM. Utility of the Mini-BESTest, BESTest, and BESTest sections for balance assessments in individuals with Parkinson disease. J Neurol Phys Ther. 2011 Jun;35(2):90-7. doi: 10.1097/NPT.0b013e31821a620c. PMID: 21934364.
24. Dahl SSH, Jørgensen L (2014) Intra- and Inter-Rater Reliability of the Mini-Balance Evaluation Systems Test in Individuals with Stroke. Int J Phys Med Rehabil 2:177. doi: 10.4172/2329-9096.1000177.
25. Yingyongyudha A, Saengsirisuwan V, Panichaporn W, Boonsinsukh R. The Mini-Balance Evaluation Systems Test (Mini-BESTest) Demonstrates Higher Accuracy in Identifying Older Adult Participants With History of Falls Than Do the BESTest, Berg Balance Scale, or Timed Up and Go Test. J Geriatr Phys Ther. 2015 Mar 19. [Epub ahead of print] PubMed PMID: 25794308.
26. King L, Horak F. On the mini-BESTest: scoring and the reporting of total scores. Phys Ther. 2013 Apr;93(4):571-5. doi: 10.2522/ptj.2013.93.4.571. PMID: 23547173.
27. Hagen C. Malkmus D. Durham E. Rehabilitation of the Head Injured Adult: Comprehensive Physical Management. Professional Staff of Rancho Los Amigos Hospital; Downey, CA: 1979. Levels of cognitive functioning.
28. BESTest: Balance Evaluation Systems Test. Learn how to administer the BESTest. Retrieved from http://www.bestest.us/training/. Accessed April 14, 2015.
29. Jonsdottir J, Cattaneo D. Reliability and validity of the dynamic gait index in persons with chronic stroke. Arch Phys Med Rehabil. 2007 Nov;88(11):1410-5. PubMed PMID: 17964880.
30. O'Hoski S, Winship B, Herridge L, Agha T, Brooks D, Beauchamp MK, Sibley KM. Increasing the clinical utility of the BESTest, mini-BESTest, and brief-BESTest: normative values in Canadian adults who are healthy and aged 50 years or older. Phys Ther. 2014 Mar;94(3):334-42. doi: 10.2522/ptj.20130104. Epub 2013 Oct 3. PubMed PMID: 24092900.

Acknowledgements

Table 1: Depicts the characteristics of all participants.

Demographics	Participants (n=25)
Age, y	50.0 (22-72)
Gender (Male:Female), n	16:9
Type of ABI (Stroke:TBI), n	18:7
Time since onset, mo.	7.44 (2-19)
Rancho Level (n = 6)*	N = 3 Rancho VII N = 3 Rancho VIII
Use of Assistive Device (Y:N)	2:23

*1 subject with TBI did not report Rancho score

Table 2: Shows Mini-BESTest Total Score Interrater Reliability among every pair of raters.

Inter-rater Reliability	ICC	95% CI	p
N1:N2	0.94	.869 - .973	< .001
N1:E1	0.909	.806 - .959	< .001
E1:N2	0.97	.933 - .987	< .001
E1:E2	0.939	.867 - .973	< .001
E2:N2	0.962	.915 - .983	< .001

N1=Novice Rater 1   E1 = Experienced Rater 1

N2 = Novice Rater 2   E2 = Experienced Rater 2

Table 3: Shows each rater's scores on every test item for all participants & the kappa (κ) correlation among all raters for each item.

Mini-BESTest item score	n* (N1) 0	1	2	n* (N2) 0	1	2	n* (E1) 0	1	2	n* (E2) 0	1	2	κ
Anticipatory Response													
1. Sit to stand	0	3	22	1	2	22	1	3	21	2	2	21	.786
2. Rise to toes	6	6	13	5	9	11	6	12	7	6	14	15	.528
3. Stand on one leg	9	9	7	5	15	5	8	11	6	5	14	6	.756
Reactive Postural Response													
4. Compensatory stepping correction in forward direction	2	11	12	3	10	12	2	10	13	4	6	15	.736
5. Compensatory stepping correction in backward direction	6	14	5	6	12	7	3	14	8	8	10	7	.658
6. Displacement, lateral stepping	3	10	12	4	8	13	3	7	15	9	6	10	.594
Sensory Organization													
7. Stance, eyes open on firm and flat surface	0	0	25	0	0	25	0	0	25	0	0	25	1.000
8. Stance, eyes closed on foam surface	0	6	19	0	6	19	0	6	19	0	6	19	1.000
9. Stance, eyes closed on firm and inclined surface	1	1	23	0	2	23	0	3	22	0	2	23	.758
Dynamic Gait													
10. Change in gait speed	0	6	19	2	4	19	1	5	19	1	6	18	.575
11. Walk with horizontal head turns	0	8	17	1	8	16	0	4	21	2	3	20	.381
12. Walk with pivot turns	3	10	12	4	6	15	1	6	18	2	9	14	.450
13. Step over obstacle	2	7	16	3	3	19	4	0	21	4	0	21	.591
14. TUG and TUG with dual task	14	5	6	13	8	4	12	10	3	12	8	5	.662
	N1			**N2**			**E1**			**E2**			
Chronbach's Alpha†	.838			.870			.885			.880			
(SEM)‡	(.393)			(.360)			(.312)			(.350)			

n*- the number of participants who received a score of 0, 1, or 2 by that rater for each item

κ - Cohen's kappa - proportion of agreement expected by chance, represents interrater reliability of individual test items.

†Chronbach's alpha illustrates the internal consistency of each raters' scores on every item

‡SEM - Standard error of measurement - reflects the amount of potential error that could have occurred

N1=Novice Rater 1   E1 = Experienced Rater 1

N2 = Novice Rater 2   E2 = Experienced Rater 2

Table 4: Is a comparison of Total Score Reliability & Item Agreement with previous studies.

	Total Score Interrater Reliability	Total Score Test Retest Reliability	Individual Item Agreement	Individual Item Agreement 2
	ICC	ICC	High (1.000)	Low (.360 -.499)
Results	.909 - .970	.844 - .967	Items: 7, 8	Items: 11, 12
Tsang et. al 2013	.960	NR*	Items: 1, 7, 9	Items: 6, 8, 11, 13,10, 3
Godi et. al 2013	.980	.960	NR	NR
Leddy et. al 2011	.910	.920	NR	NR
Dahl & Jorgensen 2014	.940	NR	Items: 1, 3, 8	Item: 11

* NR = Not Reported

Updated 4/14/15

**Table 5:** Shows test retest reliability of total scores between time 1 & time 2 among novice vs. experienced raters.

Test Retest Reliability	Time 1, mean total score (SD)	Time 2, mean total score (SD)	ICC	95% CI	*p*
N1	20.72 (4.65)	20.60 (4.28)	0.935	.858 - .971	< .001
N2	20.52 (5.35)	20.52 (5.25)	0.967	.927 - .985	< .001
E1	21.08 (5.20)	22.00 (3.73)	0.884	.755 - .947	< .001

SD = Standard Deviation
N1=Novice Rater 1
N2 = Novice Rater 2
E1 = Experienced Rater 1

Figure 1: Shows Interrater Reliability of Individual Test Items using Cohen's Kappa.
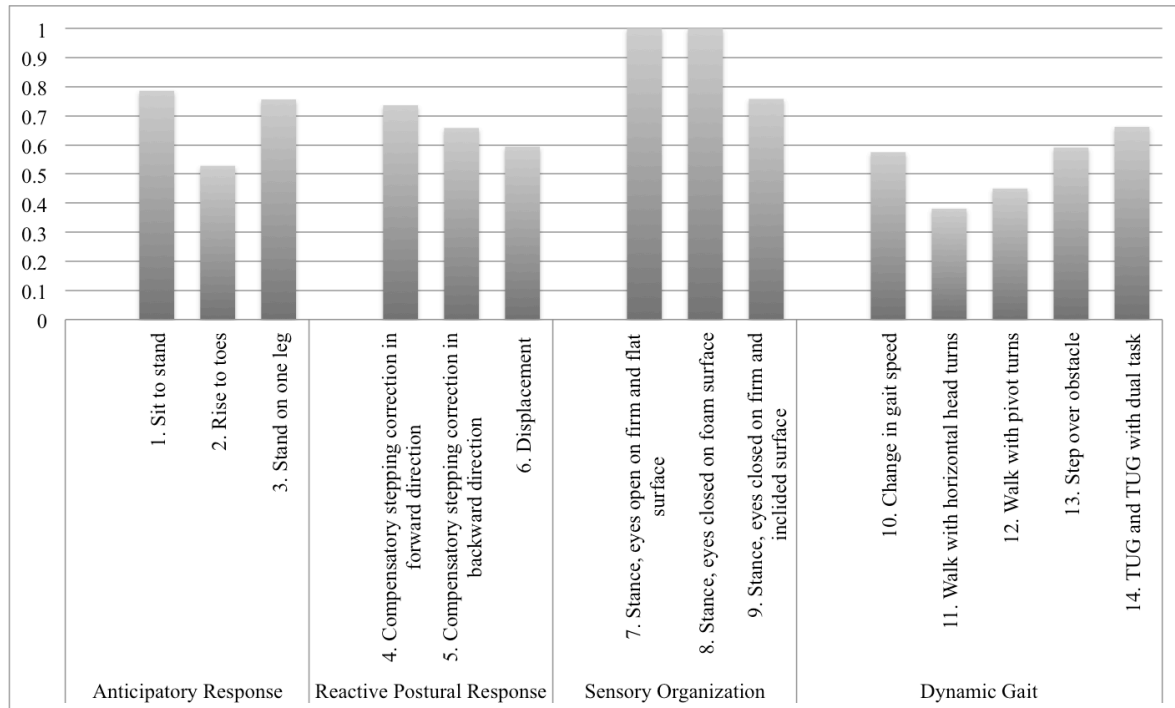
Figure 2: Illustrates Pearson's product moment correlation (r) for test retest reliability between novice & experienced raters.
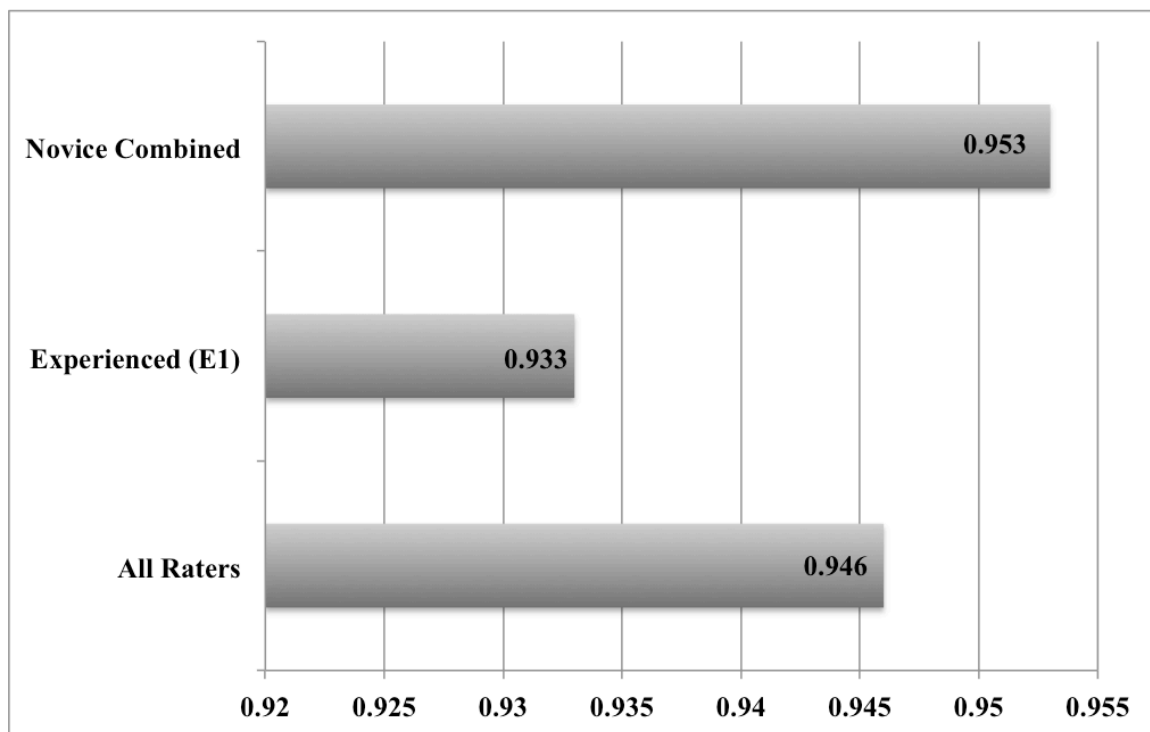
Appendix

MINI BESTest of DYNAMIC BALANCE
Balance Evaluation Systems Test Copyright 2005-2013
Instructions were standardized to 2 step commands with pauses to accommodate cognitive deficits.

Subjects should be tested with flat-heeled shoes OR shoes and socks off.
If subject must use an assistive device for an item, score that item one category lower. If subject requires physical assistance to perform an item, score "0" for that item.

ANTICIPATORY   **SUBSCORE:**   **/6**

==**ITEM 1: SIT TO STAND: ** *SIDE VIEW*==→ **Watch for legs touching chair**
(2) Normal: Comes to stand without use of hands and stabilizes independently.
(1) Moderate: Comes to stand WITH use of hands on first attempt.
(0) Severe: Unable to stand up from chair without assistance OR needs several attempts with use of hands.

Instruction: Please listen first, "Cross arms across your chest. # Try not to use your hands unless you must. # Do not let your legs lean against the back of the chair (*hand behind legs) # when you stand. ## Please stand up now."

Examiner Instructions: Note the initiation of the movement and the use of the subject's hands on the seat of the chair, the thighs, or the thrusting of the arms forward.

==**ITEM 2: RISE TO TOES: ***SIDE VIEW (BROWN TILES)*==
(2) Normal: Stable for 3 s with maximum height.
(1) Moderate: Heels up, but not full range (smaller than when holding hands) OR noticeable instability for 3 s.
(0) Severe: $\leq$ 3 s.

Instruction: Please listen first, "Place your feet shoulder width apart. # Place your hands on your hips. # Try to rise as high as you can onto your toes. # I will count out loud to 3 seconds. # Try to hold this pose for at least 3 seconds. # Look straight ahead. ## Rise now."

Examiner Instructions: Allow the subject **two attempts**. Score the best attempt. (If you suspect that subject is using less than full height, ask the **subject to rise up while holding the examiners' hands**.) Make sure the subject looks at a **nonmoving target 4-12 feet away**.

==**ITEM 3: STAND ON ONE LEG: ***FRONT VIEW (BROWN TILES)*==
Left: *Time in sec Trial 1:* _____ *Trial 2:* _____   **Right:** *Time in sec Trial 1:* _____ *Trial 2:* _____
(2) Normal: 20 s.   (2) Normal: 20 s.
(1) Moderate: < 20 s.   (1) Moderate: < 20 s.
(0) Severe: Unable.   (0) Severe: Unable.

INSTRUCTIONS THEN DEMO
Instruction: Please listen first, "Look straight ahead. # Keep your hands on your hips. # Lift your leg off of the ground behind you without touching # or resting your raised leg upon your other standing leg. # Stay standing on one leg as long as you can. # Look straight ahead. ## Lift now."

Examiner Instructions: Allow the subject **two attempts** and record the times. Record the number of seconds the subject can hold up to a **maximum of 20 seconds**. ** Stop timing when the subject moves hands off of hips or puts a foot dow**n**. Make sure the subject looks at a **nonmoving target 4-12 feet ahead**. Repeat on other side.

LET THE SUBJECT TAKE THE STEPS NECESSARY TO REGAIN BALANCE FOR ITEMS 4-6

REACTIVE POSTURAL CONTROL   **SUBSCORE: /6**

ITEM 4: COMPENSATORY STEPPING CORRECTION- FORWARD: ***SIDE VIEW (BROWN TILES)***

- (2)   Normal: Recovers independently a single, large step (second realignment step is allowed).
- (1)   Moderate: More than one step used to recover equilibrium.
- (0)   Severe: No step, OR would fall if not caught, OR falls spontaneously.

Instruction: Please listen first, # wait until I say 'Go' to begin # "Stand with your feet shoulder width apart, # arms at your sides. # Lean forward against my hands # beyond your forward limits. # When I let go, # do whatever is necessary #, including taking a step, # to avoid a fall." ## "Go"

Examiner Instructions: **Stand in front of the subject** with **one hand on each shoulder** and ask the subject to lean forward (make sure there is room for them to step forward). Require the subject to ****lean until the subject's shoulders and hips are in front of toes****. After you feel the subject's body weight in your hands, very suddenly release your support. **The test must elicit a step**. NOTE: Be prepared to catch subject.

ITEM 5: COMPENSATORY STEPPING CORRECTION- BACKWARD: ***SIDE VIEW (BROWN TILES)***

- (2)   Normal: Recovers independently with a single, large step.
- (1)   Moderate: More than one step used to recover equilibrium.
- (0)   Severe: No step, OR would fall if not caught, OR falls spontaneously.

Instruction: Please listen first, # wait until I say 'Go' to begin # "Stand with your feet shoulder width apart #, arms down at your sides. # Lean backward against my hands # beyond your backward limits. # When I let go, # do whatever is necessary, # including taking a step, # to avoid a fall." ## "Go"

Examiner Instructions: ****Stand behind the subject with one hand on each scapula**** and ask the subject to lean backward (make sure there is room for the subject to step backward.) Require the subject to ****lean until the shoulders and hips are in back of the heels****. After you feel the subject's body weight in your hands, very suddenly release your support. **Test must elicit a step**. NOTE: Be prepared to catch subject.

ITEM 6: COMPENSATORY STEPPING CORRECTION- LATERAL: ** *FRONT VIEW (BROWN TILES)*

Left
- (2) Normal: Recovers independently with 1 step (crossover or lateral OK).
- (1) Moderate: Several steps to recover equilibrium.
- (0) Severe: Falls, or cannot step.

Right
- (2) Normal: Recovers independently with 1 step (crossover or lateral OK).
- (1) Moderate: Several steps to recover equilibrium.
- (0) Severe: Falls, or cannot step.

Use the side with the lowest score to calculate subscore and total score.

Instruction: Please listen first, # wait until I say 'Go' to begin # "Stand with your feet together #, arms down at your sides. # Lean into my hand beyond your sideways limit. # When I let go, # do whatever is necessary, # including taking a step, # to avoid a fall." ## "Go"

Examiner Instructions: **Stand to the side of the subject**, place ****one hand on the side of the subject's pelvis****, and have the subject lean the whole body into your hands. Require the subject to ****lean until the midline of the pelvis is over the right (or left) foot**** and then suddenly release your hold. NOTE: Be prepared to catch subject.

SENSORY ORIENTATION   **SUBSCORE: /6**

ITEM 7: STANCE (FEET TOGETHER); EYES OPEN, FIRM SURFACE: **_FRONT VIEW (BROWN TILES)_

Time in seconds: _____

(2) Normal: 30 s.
(1) Moderate: < 30 s.
(0) Severe: Unable.

Instruction: "Place your hands on your hips. # Place your feet together until almost touching. # Look straight ahead. # Be as stable and still as possible, # until I say stop." ## "Begin" …. "Stop"

Examiner Instructions: Record the time the subject was able to stand with feet together up to a **maximum of 30 seconds**. Make sure subject looks at a **nonmoving target 4-12 feet away**.

ITEM 8: STANCE (FEET TOGETHER); EYES CLOSED, FOAM SURFACE: **_FRONT VIEW (DYCEM & FOAM)_

Time in seconds: _____

(2) Normal: 30 s.
(1) Moderate: < 30 s.
(0) Severe: Unable.

Instruction: "Step onto the foam. # Place your hands on your hips. # Place your feet together until almost touching. # Be as stable and still as possible, # until I say stop. # I will start timing you when you close your eyes" ## "Close your eyes"

Examiner Instructions: Use medium density Temper® foam, 4 inches thick**. **Assist subject in stepping onto foam****. Record the time the subject was able to stand in each condition to a **maximum of the 30 seconds**. Have the subject step off of the foam between trials. Flip the foam over between each trial to ensure the foam has retained its shape.

ITEM 9: INCLINE- EYES CLOSED*: **SIDE VIEW (DYCEM & INCLINE BOARD)_*

Time in sec: _____

(2) Normal: Stands independently 30 s and aligns with gravity.
(1) Moderate: Stands independently <30 s OR aligns with surface.
(0) Severe: Unable.

Instruction*:* "Step onto the incline ramp. # Please stand on the incline ramp with your toes toward the top. # Place your feet shoulder width apart # and have your arms down at your sides. ## I will start timing when you close your eyes." ## "Close your eyes"

Examiner Instructions: ****Aid the subject onto the ramp****. Once the subject closes eyes, begin timing and record time. Note if there is excessive sway.

DYNAMIC GAIT   **SUBSCORE:**   **/6**

**ITEM 10: CHANGE IN GAIT SPEED*: ** SIDE VIEW (GUARD ALONG WALL)*

    (2) Normal: Significantly changes walking speed without imbalance.

    (1) Moderate: Unable to change walking speed or signs of imbalance.

    (0) Severe: Unable to achieve significant change in walking speed AND signs of imbalance.

Instruction: Please listen first # and watch me # "Begin walking at your normal speed, # when I tell you 'fast,' # walk as fast as you can. # When I say 'slow,' # walk very slowly." # "Now I'll show you" **DEMO** # "When I say 'Go', begin" ## "Go"

Examiner Instructions: Allow the subject to take **3-5 steps at normal speed, and then say "fast."** After **3-5 steps, say "slow**." Allow 3-5 slow steps before the subject stops walking.

**ITEM 11: WALK WITH HEAD TURNS – HORIZONTAL*: **GUARD ALONG WALL*

    (2) Normal: performs head turns with no change in gait speed and good balance.

    (1) Moderate: performs head turns with reduction in gait speed.

    (0) Severe: performs head turns with imbalance.

Instruction: Please listen first # and watch me # "Begin walking at your normal speed, # when I say "right", # turn your head and look to the right. ## When I say "left" # turn your head and look to the left. # Try to keep yourself walking in a straight line." # "Now I'll show you" **DEMO** # "When I say 'Go', begin" ## "Go"

Examiner Instructions: **Allow the subject to reach normal speed, and give the commands "right, left" every 3-5 steps.** Score if you see a problem in either direction. If subject has severe cervical restrictions, allow combined head and trunk movements.

**ITEM 12: WALK WITH PIVOT TURNS*: **GUARD WALKING BEHIND SUBJECT*

    (2) Normal: Turns with feet close, FAST ($\leq$ 3 steps) with good balance.

    (1) Moderate: Turns with feet close SLOW ($\geq$4 steps) with good balance.

    (0) Severe: Cannot turn with feet close at any speed without imbalance.

Instruction: Please listen first # and watch me # "Begin walking at your normal speed. # When I tell you to 'turn and stop,' # turn as quickly as you can, # face the opposite direction, # and stop. # After the turn, your feet should be close together." # "Now I'll show you" **DEMO** # "When I say 'Go', begin" ## "Go"

Examiner Instructions: Demonstrate a pivot turn. **Once the subject is walking at normal speed, say "turn and stop."** Count the number of steps from "turn" until the subject is stable. Imbalance may be indicated by wide stance, extra stepping, or trunk motion.

**ITEM 13: STEP OVER OBSTACLES*: **PLACE SHOE BOX HORIZONTAL; GUARD WALKING BEHIND SUBJECT*

    (2) Normal: Able to step over box with minimal change of gait speed and with good balance.

    (1) Moderate: Steps over box but touches box OR displays cautious behavior by slowing gait.

    (0) Severe: Unable to step over box OR steps around box.

Instruction: Please listen first # and watch me # "Begin walking at your normal speed. # When you get to the box, # step over it, # not around it and keep walking." # "Now I'll show you" **DEMO** # "When I say 'Go', begin" ## "Go"

Examiner Instructions: ** Place the box (9 inches or 23 cm height) 10 feet away from where the subject will begin walking. Two shoeboxes taped together works well to create this apparatus.

ITEM 14: TIMED UP & GO (ITUG) WITH DUAL TASK: ***PLACE CONE @ 10 FT; GUARD WALKING BEHIND SUBJECT*

> (2) Normal: No noticeable change in sitting, standing or walking while backward counting when compared to TUG without Dual Task.
>
> (1) Moderate: Dual task affects either counting OR walking (>10%) when compared to the TUG without Dual Task.
>
> (0) Severe: Stops counting while walking OR stops walking while counting.

TUG:

Instruction: "When I say "Go," # stand up from the chair, # walk at your normal speed <u>around the cone</u> on the floor,# turn around, # and come back to sit in the chair." ## "Go"

Examiner Instructions: *Use the TUG time to determine the effects of dual tasking. The subject should walk a 3-meter distance.* Have the subject sitting with the subject's back against the chair. **The subject will be timed from the moment you say "Go" until the subject returns to sitting.** Stop timing when the subject's buttocks hit the chair bottom and the subject's back is against the chair. **The chair should be firm without arms.**

TUG with Dual Task:

Instruction: Practice counting backwards by threes starting at **75** # Now listen, # "Count backwards by threes starting at__**96**__. # When I say 'Go,' # stand up from chair, # walk at your normal speed around the cone on the floor, # turn around, # and come back to sit in the chair. # Continue counting backwards the entire time." ## "Start counting" # "Go"

Examiner Instructions*:* While sitting, determine how fast and accurately the subject can count backwards by threes starting from a number between 90 and 100. **Then, ask the subject to count from a different number and after a few numbers say "Go."** Time the subject from the moment you say "Go" until the subject returns to the sitting position. Score dual task as affecting counting or walking if speed slows (> 10%) from TUG and or new signs of imbalance.

TOTAL SCORE:_____**/28**

Mini-BESTest Instructions

Subject Conditions: Subject should be tested with flat-heeled shoes OR shoes and socks off.

Equipment: Temper® foam (also called T-foam™, 4 inches thick, medium density T41 firmness rating), chair without arm rests or wheels, incline ramp, stopwatch, a box (9 inches high) and a 3-meter distance measured out and marked on the floor with tape (from chair).

Scoring: The test has a maximum score of 28 points from 14 items that are each scored from 0 to 2. "0" indicates the lowest level of function and "2" the highest level of function.

If a subject must use an assistive device for an item, score that item one category lower.

If a subject requires physical assistance to perform an item, score "0" for that item.

For **Item 3** (stand on one leg) and **Item 6** (compensatory stepping-lateral), only include the score for one side (the worse score).

For **Item 3** (stand on one leg), select the best time of the 2 trials (from a given side) for score.

For **Item 14** (timed up & go with dual task), if a person's gait slows greater than 10% between the TUG without and with a dual task, then the score should be decreased by a point.

- indicates pauses