Assessing balance function in patients with total knee arthroplasty

Authors:

Andy C. M. Chan[1], Marco Y. C. Pang[2]

1. Physiotherapy Department, Hong Kong Buddhist Hospital, 10 Heng Lam Street, Lok Fu Kowloon, Hong Kong.
   Email: cccmz01@ha.org.hk
   Telephone number: (852)-2339 6226

2. Department of Rehabilitation Sciences, Hong Kong Polytechnic University, Hung Hom, Hong Kong.
   Email: marco.pang@polyu.edu.hk
   Telephone number: (852)-2766 7156

Corresponding author:

Andy C. M. Chan

Physiotherapy Department, Hong Kong Buddhist Hospital, 10 Heng Lam Street, Lok Fu Kowloon, Hong Kong.

Email: cccmz01@ha.org.hk

Telephone number: (852)-2339 6226

This is a pre-copyedited, author-produced version of an article accepted for publication in Physical Therapy following peer review. The version of record Andy C.M. Chan, Marco Y.C. Pang, Assessing Balance Function in Patients With Total Knee Arthroplasty, Physical Therapy, Volume 95, Issue 10, 1 October 2015, Pages 1397–1407 is available online at: https://doi.org/10.2522/ptj.20140486.

Abstract

Background: The Balance Evaluation Systems Test (BESTest) is a relatively new balance assessment tool. Recently, the Mini-BESTest and the Brief-BESTest, which are shortened versions of the BESTest, have been developed.

Objective: To **estimate interrater and intrarater-interoccasion reliability, internal consistency, concurrent and convergent validity, and floor/ceiling effects of the three BESTests**, and other related measures, namely, Berg balance scale (BBS), Functional Gait Assessment (FGA), and Activities-Specific Balance Confidence Scale (ABC) among patients with TKA**.**

Design: This was an observational measurement study.

Methods: To establish interrater reliability, the three BESTests were administered by three independent raters to 25 participants with TKA. Intrarater-interoccasion reliability was evaluated in 46 participants with TKA (including the 25 individuals who participated in the interrater reliability experiments) by repeating the three BESTests, BBS and FGA within one week by the same rater. Internal consistency of each test was also assessed with Cronbach's alpha. Validity was assessed in another 46 patients with TKA by correlating the three BESTests with BBS, FGA and ABC. The floor and ceiling effects were also examined.

Results: The three BESTests demonstrated excellent interrater reliability ($ICC_{2,1}$=.96-.99), intrarater-interoccasion reliability ($ICC_{2,1}$=.92-.96) and internal consistency (Cronbach's alpha=0.96-0.98). These values were comparable to those for

BBS and FGA. The three BESTests also showed moderate to strong correlations with BBS, FGA, and ABC (r=.346-.811), thus demonstrating good concurrent and convergent validity. No significant floor and ceiling effects were observed, except the BBS.

Limitations: The results are generalizable only to patients with TKA due to end-stage knee osteoarthritis.

Conclusions: The three BESTests have good reliability and validity for evaluating balance in people with TKA. The Brief BESTest is least time-consuming and may be more useful clinically.

(Abstract word count – 275)

(Total word count – 4,162)

Introduction

Total Knee Arthroplasty (TKA) has become a common surgical intervention in the treatment of severe osteoarthritis (OA) of knee joint. Both scientific research and clinical observations support the use of TKA for correction of deformity, mitigation of pain, amelioration of physical function and symptoms of OA.[1-3] However, substantial functional deficits may persist long after surgery for many patients with TKA.[4-8] One important area of concern is balance impairments, which could increase fall risk in these patients.[9] Indeed, the fall rate has been reported to be as high as 7%-40% post-operatively.[10-13] Understanding balance problems in patients after TKA is important.

Previous research in balance assessment among patients with TKA mainly involved advanced technology and/or sophisticated equipment in laboratory settings such as virtual or real obstacle avoidance,[14,15] stabilogram,[16,17] kinematic and electromyographic analysis,[9,18] and computerized dynamic posturography,[19] which might not be available and feasible in real clinical situations. While the Berg Balance Scale (BBS)[20,21] is a common tool for balance assessment and can be considered as a reference standard for assessing balance in TKA patients clinically,[22,23] it is not without limitations. Firstly, it mainly assesses static balance and has been shown to have considerable ceiling effect in various patient populations.[24-26] Balance in important dynamic tasks, such as walking, is not addressed in BBS. Secondly,

maintaining body equilibrium involves many different balance control systems. The BBS has limited ability to identify what balance system (s) is/are impaired and thus direct treatment.[27]

Recently, the 36-item Balance Evaluation Systems Test (BESTest)[28] has been developed. It assesses the functioning of 6 balance control systems (i.e. I. biomechanical constraints, II. stability limits/verticality, III. anticipatory postural adjustments, IV. postural response, V. sensory orientation, and VI. stability in gait). Good interrater reliability and intrarater-interoccasion reliability (Intraclass correlation coefficient ICC≥.88) has been reported in a cohort of individuals with different diagnoses [Parkinson's disease (PD), vestibular disorders**,** total hip arthroplasty, etc.],[**28**] and people with PD.[**29,30**] However, the BESTest takes about 45 minutes to administer and might not be a practical option in daily busy clinical situations where time limitation is a serious concern. Hence, a condensed version of the BESTest, named the Mini-BESTest,[31] was derived. With only 16 items, the Mini-BESTest takes only 15 minutes to complete and has also demonstrated good interrater reliability and intrarater-interoccasion reliability (ICC≥.91) in different patient populations.[**29,32,33**] However, two of the balance systems (biomechanical constraints and stability limits/verticality) are omitted in the Mini-BESTest, thus contradicting the theoretical framework of the original BESTest and biasing towards dynamic balance.

In response to the drawbacks of the BESTest and Mini-BESTest, the 8-item Brief-BESTest[34] has been developed more recently. Unlike the Mini-BESTest, the Brief-BESTest includes items that assess all 6 balance systems. It requires less than 10 minutes to administer and could be more feasible for clinical use. Nevertheless, the psychometric properties of the BESTest, Mini-BESTest and Brief-BESTest have not been tested in individuals with TKA specifically. Using a sample of patients with TKA, the objective of this study was to **estimate** the psychometric properties **in terms of the interrater and intrarater-interoccasion reliability, internal consistency, concurrent and convergent validity, and floor/ceiling effects** of the three versions of the BESTest and other established balance and related measures, namely, BBS, Functional Gait Assessment (FGA), and Activities-Specific Balance Confidence Scale (ABC).

Methods

Study Design

An observational measurement study was undertaken to examine the reliability and validity of the three versions of the BESTest, **BBS and FGA** in people with TKA. This study consisted of two phases. In the first phase, we would like to establish the reliability of the three versions of the BESTest, **BBS and FGA** first, because reliability is a prerequisite to validity.[35] To establish intrarater-interoccasion reliability,

118 individuals with TKA participated in a two assessment sessions (within a one-week
119 interval), during which the three BESTests, BBS, and FGA were administered by the
120 same physical therapists. Some of these individuals were evaluated independently by
121 three physical therapists in the first session to establish interrater reliability. In the
122 second phase, we would like to examine the concurrent validity, convergent validity,
123 and the floor/ceiling effects of the different balance tests at different stages of
124 recovery after TKA. Another group of people with TKA were evaluated with the same
125 balance tests three times, at 2 weeks, 12 weeks and 24 weeks after the operation.
126

127 **Participants**

128 Participants admitted for TKA in Joint Replacement Centre of the Buddhist Hospital
129 in Hong Kong from September 2012 to May 2013 and referred for rehabilitation were
130 recruited. The inclusion criteria were: aged 50-85 years; having had their first TKA
131 due to a diagnosis of knee OA**;** able to follow verbal instructions and provide
132 informed consent. Exclusion criteria were: TKA due to rheumatoid arthritis (RA) of
133 knee or traumatic injury; previous history of operation on lower limbs; known
134 medical diagnoses that affect balance (e.g., stroke). **For phase one of the study (the**
135 **reliability experiments), participants were required to have undergone TKA at**
136 **least 6 months earlier.** This was important because we had to be able to assume
137 stability in the response variable (balance performance) for reliability experiments. A
138 previous study[36] showed that little improvement occurred beyond 26 weeks after TKA.

Prior to enrolling in the study, all participants signed a written informed consent that had been approved by the Human Research Ethics Subcommittee of the Hong Kong Polytechnic University and the Institutional Review Board of the Kowloon Central Cluster, Hospital Authority. All procedures were carried out according to the Declaration of Helsinki.

Sample Size Estimation

All **a** priori sample size calculations were performed using the PASS 2011 software (NCSS Statistical Software, Kaysville, Utah, USA). The sample size for interrater reliability analysis was based on the following assumptions: (1) 3 raters, (2) a null ICC of 0.75,[35] (3) an expected ICC of 0.90,[28,29] (4) a Type I error of 0.025 (1-tailed), (5) a power of 0.80.[37,38] Hence, a sample of 23 patients with TKA was required for establishing interrater reliability among 3 raters.

The sample size for intrarater-interoccasion reliability analysis was based on the following assumptions: (1) 2 occasions, (2) a null ICC of 0.75[35], (3) an expected ICC of 0.90,[29] (4) a Type I error of 0.025 (1-tailed), (5) a power of 0.80.[37,38] A minimum of **33 patients** would be needed for establishing intrarater-interoccasion reliability in 2 assessment sessions.

For establishing validity, Horak et al.[28] showed a moderate correlation between the

BESTest and Activities Balance Confidence Scale (ABC) in a mixed population (r=.64). A high correlation (r=.79) between the Mini-BESTest and BBS in patients with Parkinson's disease was demonstrated by King et al.[39] A medium to large effect size (r=0.4) was thus assumed for this study, a minimum sample size of 44 would be required.

Qualifications of raters

All 3 raters involved in this study are qualified physical therapists with at least 10 years of experience in working with people with TKA. They had prior experience using the BBS and ABC, but not the three BESTests and FGA. To ensure competency in using the BESTests and FGA, all raters were required to read the instruction manual for these tests, and viewed the training video for the BESTest. This was followed by a 2-week practice period during which the raters practiced administering the different tests used in this study among themselves. In addition, the raters were required to administer all tests on at least two patients with TKA prior to the commencement of the actual data collection period.

Measurement tools

BESTest

Each of the 36 items was scored on a 4-level ordinal scale from 0 to 3 (0: severely impaired balance or inability to complete a task; 3 no impairment of balance or able to

perform a task successfully). The BESTest provides 6 subsection scores and a total score. The six sub-sections included are section 1 – Biomechanical Constraints (5 items; score range: 0-15), Section 2 – Stability Limits/Verticality (7 items; score range: 0-21), section 3 – Anticipatory Postural Adjustment (6 items; score range: 0-18), section 4 – Postural Responses (6 items; score range: 0-18), section 5 – Sensory Orientation (5 items; score range: 0-15) and section 6 – Stability in Gait (7 items; score range: 0-21). The total score (range: 0-108) was converted to a percentage score for subsequent analysis.[28]

Mini-BESTest

The mini-BESTest consists of 16 items from the original BESTest. Each item was scored on a 3-level ordinal scale from 0 to 2 (0: severe impairment of balance; 2: no impairment in balance), yielding a maximum score of 32.[29]

Brief-BESTest

The Brief-BESTest comprises 8 items. The scoring method for each item was the same as in the full BESTest described above. The maximum possible score is 24.[34]

Berg Balance Scale

The BBS comprises a set of 14 balance tasks. Each item was scored on a 5-level ordinal scale from 0 to 4, yielding a maximum total score of 56. Higher scores

indicate better balance.[20]

Functional Gait Assessment

The FGA is a 10-item assessment used to evaluate postural stability during various walking tasks.[40-42] Each item was scored on a 4-level ordinal scale from 0 to 3 (maximum total score: 30). Higher total scores are indicative of better performance. The FGA has excellent interrater reliability (ICC=.93) in independently living individuals aged 40-89 years.[43] It also has good interrater reliability (ICC≥.86) and test-retest reliability (ICC≥.74) in individuals with PD[30] and vestibular disorders.[40]

Activity-specific Balance Confidence Scale

The ABC scale quantifies how confident a person feels that he or she will not lose balance while performing 16 activities of daily living on a scale from 0% (absolutely no confidence) to 100% (completely confident).[44] The test was self-administered and the score of the 16 items was averaged. The ABC scale had good test-retest (ICC=.99) and interrater reliability (ICC=.85).[44]

Procedures

Demographic information was obtained from medical records and **participant** interview in the first session. The **average** pain intensity experienced on the operated knee **over the past 24 hours** was measured by the Numerical Pain Rating Scale

(NPRS), which is an 11-point scale ranging from 0 (no pain) to 10 (worst imaginable pain).[45] Knee range of motion on both the operated side and non-operated side was measured with a 1-degree-increment long arm goniometer (Baseline® 180° Goniometer, NexGen Ergonomics Inc., Pointe-Claire, Quebec, Canada).

In the first assessment session, the balance performance of 25 TKA participants was assessed independently by 3 raters to establish interrater reliability. As the items for the Brief-BESTest were taken from the full BESTest and the item scoring method was exactly the same, the Brief-BESTest score was computed from the BESTest. Previous research[45] has also used a similar method to score items on reduced version in patients with total hip and knee arthroplasties. The common items for Mini-BESTest and BESTest was graded simultaneously with their respective scales (0 to 2 for Mini-BESTest; 0 to 3 for BESTest). In addition, for those items that were duplicated between the BESTest and other balance and related tests (BBS and FGA), the participants would be asked to perform it only once, and the performance was rated according to the specific scoring criteria for each test.

Each balance test was administered by any one of the 3 raters in random order, and all raters concurrently observed and rated the participant's performance. Sequence of test administration (BESTest, BBS, FGA, and ABC) and rater was randomized by a computer program (Random blocks generation by Excel 2013 by Microsoft

Corporation. One Microsoft Way, Redmond, WA, USA). The average length of the assessment session 1 was 1.5 hours. Short breaks were given between tests to avoid over-exertion as needed. The assessment session took place in the afternoon to minimize effect of morning stiffness. The raters were instructed not to discuss the scores among themselves.

A total of 46 individuals with TKA (including the 25 individuals who were involved in the interrater reliability experiments) participated in the intrarater–interoccasion reliability experiments. The procedures for the first assessment (session 1) were the same as described above. A second assessment session (session 2) was conducted within one week after session 1. No physical therapy treatment was provided during the period between sessions 1 and 2. In session 2, the 46 participants were evaluated individually with the same balance and related tests by the same rater in session 1. To minimize the confounding effect of different time of testing, assessment session 2 also took place in the afternoon.

Another 46 patients with TKA participated in the validity experiments. Each participant was assessed at 3 time points: 2 weeks, 12 weeks and 24 weeks after operation. In each session, participant was evaluated with the same six tests. The sequence of tests was also randomized as described in the reliability experiments. These data were also used to examine the floor and ceiling effects.

Statistical Analysis

IBM SPSS Statistic for Windows software program (version 19.0, IBM, Armonk, NY), was used for all statistical analyses. The level of significance was set at $p \leq 0.05$.

Reliability

Interrater and intrarater-interoccasion reliability of the BESTest, Mini-BESTest, Brief-BESTest, BBS and FGA total scores were assessed by using the intraclass correlation coefficient ($ICC_{2,1}$). Using the data from session 1 of the intrarater-interoccasion reliability tests, the internal consistency of the five balance tests was evaluated by Cronbach's alpha. Cronbach's alpha was also calculated for the subtests of the BESTest. This would allow us to examine individual items to determine how well they fit the subscales of the BESTest as well. If the Cronbach's alpha for a particular subtest is low, this may indicate that some items in the subscale may represent a different component of balance function than the other items.[35] The following criteria were used to judge the magnitude of the reliability coefficient: Poor reliability= ICC < 0.4; fair reliability= ICC≥0.4 but <0.7, good reliability= ICC≥0.7 but <0.9; and excellent reliability= ICC≥0.9.[47] The minimal detectable change at 95% confidence interval (MDC_{95}) for each balance test, which was an estimation of the smallest change in score that can be detected objectively for a participant more than measurement error, was calculated by the formula[35]: $MDC_{95} = SEM \times \sqrt{2} \times 1.96$ and

SEM = $\sqrt{}$(MSE), where MSE is the mean square error generated from the analysis of variance model based on the intrarater-interoccasion reliability data, and SEM is the standard error of measurement.[48]

Validity

The BESTest, Mini-BESTest, Brief-BESTest scores were correlated with the BBS and FGA total score (i.e., concurrent validity) and ABC score (i.e., convergent validity) using Pearson's product moment correlation coefficient (r) or Spearman's rho (ρ), depending on whether the assumptions for parametric statistics were fulfilled. The inter-correlations among the three BESTests were also examined using the same statistical methods. A correlation coefficient of .00 to .25 means little to no relationship, .25 to .50 means fair, .50 to .75 means moderate and .75 to 1.00 means high correlation.[47]

Floor and ceiling effects

The skewness (γ_1) of the score distribution at 2 weeks, 12 weeks and 24 weeks post-TKA was examined. An absolute value of greater than 1.0 indicates that the distribution is highly skewed.[49] Thus, a positive γ_1 value > +1.0 denotes substantial floor effect while a negative value < -1.0 indicates substantial ceiling effect. The floor and ceiling effects were further examined by calculating the proportion of participants attaining the lowest and highest possible scores at the three time points. A proportion

greater than 20% was considered to be significant.[33]

Source of Funding

There was no external funding source for this study.

Results

Ninety-two individuals with TKA (reliability tests n=46 and validity tests n=46) participated in the study. Characteristics of the participants are shown in Table 1. None of the participants required any mobility aids for indoor walking or during testing. One participant **did not return for the validity experiments** at 24-week follow-up because he had moved to a different city.

Reliability

Twenty-five and 46 participants were involved in the interrater and intrarater-interoccasion reliability testing respectively. The BESTest, Mini-BESTest, and Brief-BESTest demonstrated excellent interrater reliability ($ICC_{2,1}$=.96-.99, $p \leq .001$), intrarater-interoccasion reliability ($ICC_{2,1}$=.92-.96, $p \leq .001$), and internal consistency (Cronbach's alpha=.96-.98) (Table 2). Good to excellent interrater and intrarater-interoccasion reliability and internal consistency were also established for the six subtests of the BESTest (Table 2). The MDC_{95} value of the BESTest, Mini-BESTest, and Brief-BESTest was 6.22, 3.71, and 3.19 respectively.

328

329 **Validity**

330 There were moderate to high associations of the 3 BESTests with the FGA and BBS at
331 2 weeks (correlation=.73-.81, $p \leq .01$), 12 weeks (correlation=.58-.81, $p \leq .01$) and 24
332 weeks after TKA (correlation=.55-.73, $p \leq .01$), thus demonstrating good concurrent
333 validity (Table 3). The three BESTests were also significantly correlated with the
334 ABC score at 2 weeks (correlation=.34-.43, $p \leq .05$), at 12 weeks (correlation=.40-.48,
335 $p \leq .01$) and 24 weeks after TKA (correlation=.47-.50, $p \leq .01$), thus showing good
336 convergent validity. In addition, high inter-correlations were found among the three
337 BESTests (correlation=.82-.93, $p \leq .01$) at all three measurement time points.

338

339 **Score distribution, ceiling and floor effects**

340 None of the six measures had a skewness value greater than +1.0 or smaller than -1.0
341 at 2 weeks post-TKA (Table 4). At 12 and 24 weeks post-TKA, the distribution of
342 BBS and FGA showed skewness values smaller than -1.0 (i.e., ceiling effect). At 24
343 weeks post-TKA, ABC also had a skewness value lower than -1.0. The score
344 distribution of the three BESTests, BBS, FGA and ABC at 24 weeks after TKA is
345 shown in Figure 1. When examining the proportion of people obtaining the maximum
346 possible score, it was obvious that the BBS had the most severe ceiling effect, with
347 52.2% and 57.8 % of the participants attaining the maximum score at 12 weeks and
348 24 weeks after TKA respectively. The three versions of the BESTest, in contrast,

showed little ceiling effect, with only 2.2%-8.9% of the participants reaching the top score at the same time point.

Discussion

In the current study, the psychometric properties of different versions of the BESTest, **BBS and FGA** were systematically examined for the first time in people with TKA. The study showed that the three BESTests have good reliability and validity to measure balance performance in individuals with TKA due to knee OA, without significant floor and ceiling effects at 2, 12 and 24 weeks post-TKA.

Reliability

The BESTest, Mini-BESTest, Brief-BESTest had high internal consistency, indicating that the three BESTests measured the similar underlying attribute. The interrater and intrarater-interoccasion reliability of the three BESTests were excellent when administered to individuals with TKA, which were comparable to BBS and FGA. The MDC_{95} of the BESTest, Mini-BESTest, Brief-BESTest obtained in our study was 6.22, 3.71, 3.19 respectively, which represent the smallest difference that would reflect a genuine change in the total score of these tests. These values are quite comparable to those found in people with mixed neurological conditions (3.5)[32] and people with chronic stroke (3.0)[33] using the Mini-BESTest. The MDC_{95} values found here would be useful when interpreting the results of future clinical trials. A real change in

370 balance ability following intervention should exceed the MDC value.

371

372 **Validity**

373 The high correlations between the BESTest, Mini-BESTest, Brief-BESTest, and the
374 established balance (BBS, FGA) and related measures (ABC) indicate excellent
375 concurrent and convergent validity. Our results are in line with previous findings in
376 other patient populations. Strong associations of the BBS with the BESTest (r=.87)[30]
377 and Mini-BESTest (r=0.79)[39] have been reported among patients with PD. The
378 BESTest has also been shown to have strong correlations with ABC (r=0.75) and
379 FGA (r=0.88) in patients with PD.[30] In people with stroke, significant correlation also
380 existed between the Mini-BESTest and BBS (rho=.83) and ABC (rho=.50).[33] In
381 addition, the BESTest showed significant association with ABC (r=.636) in a
382 population with different balance disorders.[28] The three BESTests also showed strong
383 inter-correlations, indicating that individuals with a less optimal score in one version
384 of the BESTest also tended to have a less optimal score in the other two versions of
385 the BESTests. This finding also concurred with a previous study in patients with PD,
386 where a strong correlation was identified between the BESTest and Mini-BESTest
387 (r=0.95) [29], and also between the Mini-BESTest and Brief BESTest (r=0.94).[50]

388

389 **Score Distribution, Ceiling and Floor Effects**

390 While none of the balance tests evaluated here had significant floor effects, the three

versions of the BESTests, especially the full BESTest and Mini-BESTest had the least ceiling effect, which was shown by the low degree of skewness, and the small percentage of participants achieving the top scores at 2, 12, and 24 weeks post-TKA. A considerably higher ceiling effect of BBS was found at 12 and 24 weeks after TKA (52.2% and 57.8% respectively) compared with the three versions of the BESTest (Table 4). In addition, the FGA also demonstrated substantial skewness ($\gamma_1 > 1.0$) at 12 weeks and 24 weeks post-TKA (Table 4). Other studies also showed more severe ceiling effect of BBS and FGA than the BESTests. For example, King et al.[39] found that 1% and 13.4% of people with mild PD achieved maximum score for the Mini-BESTest and BBS respectively. Another study in PD showed that the proportion of people who received a perfect score on the BBS, FGA and BESTest was 10%, 1.3% and 0% respectively.[30] In people with chronic stroke, the score distribution for the Mini-BESTest was significantly less skewed when compared with the BBS. Only 0.9% achieved maximum score for the Mini-BESTest, compared with 32.1% for the BBS.[33]

What may explain the difference in ceiling effects between the three BESTests and BBS and FGA? The BBS consists of tasks which are relatively less challenging, for instance sitting to standing, sitting and standing without support and turn to look behind shoulder while standing. On the other hand, the FGA is an ambulatory-oriented test focusing mainly on dynamic gait balance. The majority of

412 our participants had experienced substantial recovery of their physical and functional
413 mobility, especially at 12 and 24 weeks after the operation, thus leading to a ceiling
414 effect of BBS of FGA. The BESTests, in contrast, consist of more demanding tasks
415 (e.g., hip and trunk lateral strength, reach test, and postural responses to external
416 perturbations). As such, this may have enhanced the ability of the BESTests to
417 discriminate between participants when compared with the BBS and FGA at different
418 time points.
419
420 Upon examining the ceiling effect of the three BESTests, it was found that none of the
421 participants attained the maximum score at 2 weeks post-TKA. At later stages of
422 recovery, **the proportion of people who obtained the perfect score (i.e., ceiling**
423 **effect)** was the lowest with the BESTest, followed by the Mini-BESTest, and
424 Brief-BESTest. Although the Brief-BESTest had the highest ceiling effect amongst the
425 three BESTests, merely 8.9% of participants achieved a perfect score at 24 weeks
426 post-TKA. Overall, after considering the findings on reliability, validity and ceiling
427 effect, the BESTest is the best balance assessment tool. However, the Mini-BESTest
428 and Brief-BESTest are reasonable alternatives if time constraint is an important
429 concern.
430
431 **The BESTest is a relatively new balance assessment tool**. According to the original
432 authors of the BESTest, physical therapists who were naive to the BESTest should be

able to learn how to administer it with prior review of the instructions.[28] For the sake of safety, however, it was recommended by the original authors that the push and release technique to elicit automatic postural responses by suddenly releasing the participant's leans requires observation and practice with at least video demonstration.[28] Our study confirmed that physical therapists who had no prior experience in using it can achieve excellent reliability after self-learning that involved reading the instructions manual, watching a demonstration video, and a brief practice period. The results of this study can be generalized to the physical therapists who have undergone similar training.

Limitations and Future Research Directions

The participants had first TKA (unilateral) due to knee OA. Therefore, the results can only be generalized to individuals with similar characteristics. Further investigation is warranted to confirm and expand the present results and generalizability in people with bilateral TKA or due to other conditions such as rheumatoid arthritis. Future research is warranted to evaluate the sensitivity and specificity (predictive validity) of the BESTest for predicting fallers in patients with TKA and the responsiveness of the BESTest in assessing change in balance ability in patients with TKA during recovery. The interrater reliability coefficients may have little implications in real clinical practice. While there were three raters involved in the interrater reliability experiments, only one rater actually administered the test. The other two raters simply

observed the performance of the patients and provided their own ratings independently in the same visit. Such scenario does not resemble what is typically encountered in daily clinical practice. The interrater reliability coefficients derived here do not include the patient variability that would exist if one clinician performed the measurement at a patient's initial visit and a second clinician at a reassessment, which is a situation that may be more frequently encountered in the real world. Further study is also required to use the BESTest in directing treatment regime by identifying the body balance system(s) that is/are the most impaired in people with TKA.

Conclusion

The BESTest, Mini-BESTest, and Brief-BESTest have good reliability and validity in evaluating balance in people with TKA. While the three BESTests have comparable psychometric properties, the use of the Brief-BESTest is least time-consuming and could be particularly useful for clinicians and researchers in the field.

References

1. Dickstein R, Heffes Y, Shabtai EI, Markowitz E. Total knee arthroplasty in the elderly: patients' self-appraisal 6 and 12 months postoperatively. *Gerontology*. 1998;44:204-210.
2. Hawker G, Wright J Coyte P, et al. Health-related quality of life after knee replacement. *J Bone Joint Surg Am*. 1998;80:163-173.
3. Jones AC, Voaklander DC, Williams D, Johanston C, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community-based population. *J Rhematol*. 2000;27:1745-1752.
4. Moffet H, Ouellet D, ParentE, Brisson M. Time-course of natural locomotor recovery in the first year following knee arthroplasty. *In: Arsenault AB, Mckinley P, McFadyen B, editors. Proceedings of the Twelfth ISEK Congress*. Montreal, Quebec: University of Montreal. 1998;230-231.
5. Murray MP, Gore DR, Laney WH, Gardner GM, Mollinger LA. Kinesiologic measurements of functional performance before and after double compartment Marmor knee arthroplasty. *Clin Orthop*. 1983;173:191-199.
6. Roush SE. Patient-perceived functional outcomes associated with elective hip and knee arthroplasties. *Phys Ther*. 1985;65:1496-1500.
7. Skinner HB. Pathokinesiology and total joint arthroplasty. *Clin Orthop*. 1993;288:78-86.
8. Walsh M, Woodhouse LJ, Thomas SG, Finch E. Physical impairments and

491 functional limitations: a comparison of individuals 1 year after total knee
492 arthroplasty with control subjects. *Phys Ther*. 1998;78:248-258.
493 9. Gage WH, Frank JS, Prentice SD, Stevenson P. Postural responses following a
494 rotational support surface perturbation, following knee joint replacement: Frontal
495 plane rotations. *Gait & Posture.* 2008;27:286-293.
496 10. Kearns RJ, O'Connor DP, Brinker MR. Management of falls after total knee
497 arthroplasty. *Orthropedics*. 2008;31:225.
498 11. Swinkels A, Newman JH, Allain TJ. A prospective observational study of falling
499 before and after knee replacement surgery. *Age and Aging*. 2009;38:175-181.
500 12. Matsumoto H, Okuno M, Nakamura T, Yamamoto K, Hagino H. Fall incidence
501 and risk factors in patients after total knee arthroplasty. *Arch Orthop Trauma*
502 *Surg*. 2012;132:555-563.
503 13. Levinger P, Menz HB, Morrow AD, Wee E, Feller JA, Bartlett JR, Bergman N.
504 Lower limb proprioception deficits persist following knee replacement surgery
505 despite improvements in knee extension strength. *Knee Surg Sports Traumatol*
506 *Arthrosc*. 2012;20:1097-1103.
507 14. Byrne JM, Prentice SD. Swing phase kinetics and kinematics of knee replacement
508 patients during obstacle avoidance. *Gait and Posture*. 2003;18:95-104.
509 15. Mauer AC, Draganich LF, Pandya N, Hofer J, Piotrowski GA. Bilateral total knee
510 arthroplasty increases the propensity to trip on an obstacle. *Clin Orthrop Relat*
511 *Res*. 2005;433:160-165.

16. Friden T, Zatterstrom R, Lindstrand A, Moritz U. A stabilometric technique for evaluation of lower limb instabilities. *Am J Sport Med*. 1989;17:118-122.

17. Tropp H, Odenrick P. Postural control in single-limb stance. *I Orthop Res*. 1988;6:833-839.

18. Gage WH, Frank JS, Prentice SD, Stevenson P. Organization of postural responses following a rotational support surface perturbation, after TKA: Sagittal plane rotations. *Gait & Posture*. 2007 Jan;25(1):112-20.

19. Bakirhan S, Angin S, Karatosun V, Unver B, Gunal I. A comparison of static and dynamic balance in patients with unilateral and bilateral knee arthroplasty. *Joint Dis Relat Surg*. 2009;20:93-101.

20. Berg KO, Maki BE, Williams JI, Holliday J, Wood-Dauphinee SL. Clinical and laboratory measures of postural balance in an elderly population. *Arch Phys Med Rehabil*. 1992;73:1073-1080.

21. Hess JA, Woollacott M. Effect of high-intensity strength-training on functional measures of balance ability in balance-impaired older adults. *J Manip Physio Ther.* 2005;28:582-590.

22. Lien J, Dibble L. Systems model guided balance rehabilitation in an individual with declarative memory deficits and a total knee arthroplasty: A case report. *J Neuro Phys Ther*. 2005;29:43-49.

23. Tousignant M, Moffet M, Boissy P, Corriveau H, Cabana F, Marquis F. A randomized controlled trial of home telerehabilitation for post-knee arthroplasty.

533 *J Telemed Telecare*. 2011;17:195-198.

534 24. Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke
535 rehabilitation: a systematic review. *Phys Ther*. 2008;88:559-566.

536 25. Tanji H, Gruber-Baldini AL, Anderson KE, retzer-Aboff I, Reich SG, Fishman PS,
537 et al. A comparative study of physical performance measures in Parkinson's
538 disease. *Mov Disord*. 2008;23:1897-1905.

539 26. Steffen T, Seney M. Test-retest reliability and minimal detectable change on
540 balance and ambulation tests, the 36-item short-form health survey, and the
541 unified Parkinson disease rating scale in people with Parkinsonism. *Phys Ther*.
542 2008;88:733-746.

543 27. Hinman RS, Bennell KL, Metcalf BR, Crossley KM. Balance impairments in
544 individuals with systematic knee osteoarthitic: a comparison with matched
545 controls using clinical tests. *Rheumatology*. 2002;41:1388-1394.

546 28. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest)
547 to differentiate balance deficits. *Phys Ther*. 2009;89:484-498.

548 29. Leddy AL, Crowner BE, Earhart GM. Utility of the Mini-BESTest, BESTest, and
549 BESTest sections for balance assessments in individuals with Parkinson disease.
550 *J Neurol Phys Ther*. 2011;35:90-97.

551 30. Leddy AL, Crowner BE, Earhart GM. Functional gait assessment and balance
552 evaluation system test: reliability, validity, sensitivity, and specificity for

identifying individuals with Parkinson disease who fall. *Phys Ther*. 2011;91:102–113.

31. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation System's Test: the mini-BESTest. *J Rehabil Med*. 2010;42:323-331.

32. Godi M, Franchignoni F, Caligari M, Giordano A, Turcato AM, Nardone A. 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.

33. Tsang CSL, Liao LR, Chung RCK, Pang MYC. Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. *Phys Ther*. 2013;93:1102-1115.

34. 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;92:1197-1207.

35. Portney LG, Watkins MP. *Foundations of clinical researches: Applications to practice. 3rd ed*. Upper Saddle River, NJ: Pearson/Prentice Hall; 2009.

36. Kennedy DM, Stratford PW, Riddle DL, Hanna SE, Gollish JD. Assessing recovery and establishing prognosis following total knee arthroplasty. *Phys Ther.* 2008;88:22-32.

37. Donner A, Eliasziw E. Sample size requirements for reliability studies. *Stat Med*.

1987;6:441-448.

38. Walter SD, Eliasziw E, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101-110.

39. King LA, Priest KC, Salarian A, Pierce D, Horak FB. Comparing the Mini-BESTest with the Berg Balance Scale to evaluate balance disorders in Parkinson's disease. *Parkinsons Di*s. 2012;2012:375419.

40. Wrisley DM, Marchetti GF, Kuharsky DK, Whitney SI. Reliability, internal consistency, and validity of data obtained with functional gait assessment. *Phys Ther.* 2004;84:906-918.

41. McConvey J, Bennett SE. Reliability of the dynamic gait index in individuals with multiple sclerosis. *Arch Phys Ther Rehabil*. 2005;86:130-133.

42. Jonsdottir J, Cattaneo D. Reliability and validity of the dynamic gait index in persons with chronic stroke. *Arch Phys Med Rehabil.* 2007;88:1410-1415.

43. Walker ML, Austin AG, Banke GM, et al. Reference group data for the Functional Gait Assessment. *Phys Th*er. 2007;87:1468-1477.

44. Mak MK, Lau AL, Law FS, Cheung CC, Wong IS. Validation of Chinese translated Activities-specific balance scale. *Arch Phys Med Rehabil*. 2007;88:496-503.

45. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire

595 (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale
596 (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain
597 (ICOAP). *Arthritis Care Res*. 2011;63:S240-252.
598 46. Jogi P, Spaulding SJ, Zecevic AA, Overend TJ, Kramer JF. Comparison of the
599 original and reduced versions of the Berg Balance Scale and the Western Ontario
600 and McMaster Universities Osteoarthritis Index in patients following hip and
601 knee arthroplasty. *Phys Canada*. 2011;63:107-114.
602 47. Domholdt E. *Rehabilitation research: principles and applications*. Philadelphia:
603 WB Saunders; 2005.
604 48. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of
605 interest: an applied example using elbow flexor strength data. *Phys Ther*.
606 1997;77:745-750.
607 49. Bulmer MG. *Principles of statistics*. Mineola, NY: Dover Publications, Inc.; 1979.
608 50. Duncan RP, Leddy AL, Cavanaugh JT, et al. Comparative utility of the BESTest,
609 Mini-BESTest, and Brief-BESTest for predicting falls in individuals with
610 Parkinson disease: a cohort study. *Phys Ther*. 2013;93:542-550.

Figure legend

Fig. 1

Frequency distribution of scores on the (A) Balance Evaluation Systems Test (BESTest), (B) Mini-BESTest, (C) Brief-BESTest, (D) Berg Balance Scale (BBS), (E) Functional Gait Assessment (FGA), and (F) Activities-specific Balance Confidence scale (ABC) at post-op 24 weeks (45 participants with TKA) is shown.

Table 1. Characteristics of participants

Parameters	Interrater reliability experiments (N=25)	intrarater-interoccasion reliability experiments (N=46)	Validity experiments (N=46)
Age (year)[a]	69.7 ± 6.8	69.1 ± 6.1	66.6 ± 6.1
Sex % (n)			
Men	32% (8)	26% (12)	26% (12)
Women	68% (17)	74% (34)	74% (34)
Side of knee operation % (n)			
Left	64% (16)	54% (25)	39% (18)
Right	36% (9)	46% (21)	61% (28)
Body mass index (kg/m^2) [a]	24.6 ± 3.9	25.7 ± 3.8	26.1 ± 4.0
Faller % (n)[c]	0% (0)	0% (0)	0% (0)
Numeric pain rating scale [a]	0.3 ± 0.5	0.3 ± 0.5	0.6 ± 1.1
Knee range of motion (degree)			
Flexion: operated side [a]	115.2 ± 7.6	115.7 ± 8.7	116.7 ± 9.0
Flexion: non-operated side [a]	114.4 ± 10.1	116.1 ± 12.3	117.2 ± 12.6
Extension: operated side [a]	-2.0 ± 3.2[b]	-1.7 ± 2.8	-0.5 ± 1.5
Extension: non-operated side [a]	-0.2 ± 1.0	-1.3 ± 4.0	-1.8 ± 4.1

[a]Values are mean ± SD unless otherwise indicated.

[b]Negative value in knee range of motion refers to the extension lag

[c]The participant was considered as a faller if he/she reported that he/she had experienced at least one fall in the 12-month period prior to the first assessment session.

Table 2. Interrater reliability, intrarater-interoccasion reliability, and internal consistency

	Interrater reliability (n=25)		Intrarater-interoccasion reliability (n=46)					Internal consistency (n=46)
Balance Measure	ICC[a]	95% CI[b]	ICC[a]	95% CI[b]	MDC_{95}[c]	SEM[d]	**95% CI[b]**	Cronbach's Alpha
BESTest[e]	0.99	.99-.99	0.96	.93-.98	6.22	2.24	**1.86-2.83**	0.98
I. Biomechanical constraint	0.99	.98-.99	0.96	.93-.98	10.11	3.65	**3.03-4.59**	0.98
II. Stability limits / verticality	0.99	.99-.99	0.76	.60-.86	9.08	3.28	**2.72-4.13**	0.86
III. Anticipatory postural adjustment	0.99	.98-.99	0.90	.83-.94	13.73	4.95	**4.11-6.24**	0.95
IV. Postural responses	0.99	.99-.99	0.87	.77-.92	22.71	8.19	**6.80-10.32**	0.93
V. Sensory orientation	1.00	1.00-1.00	1.00	1.00-1.00	0.00	0.00	**0.00-0.00**	1.00
VI. Stability in gait	0.98	.97-.99	0.95	.91-.97	12.54	4.52	**3.79-5.76**	0.97
Mini-BESTest	0.96	.93-.98	0.92	.87-.96	3.71	1.34	**1.11-1.68**	0.96
Brief-BESTest	0.97	.94-.98	0.94	.90-.97	3.19	1.15	**0.95-1.45**	0.97
BBS[t]	0.98	.97-.99	0.97	.94-.98	2.00	0.72	**0.60-0.91**	0.98
FGA[g]	0.98	.97-.99	0.97	.95-.98	2.59	0.94	**0.78-1.18**	0.98

[a]ICC=intraclass correlation coefficient
[b]CI=confidence interval
[c]MDC_{95}=minimal detectable change at 95% confidence interval
[d]SEM= standard error of measurement
[e]Balance Evaluation System Test
[t]Berg Balance Scale
[g]Functional Gait Assessment

Table 3. Concurrent and convergent validity

	Variables	BBS[b]	FGA[c]	BESTest[d]	Mini-BESTest	Brief BESTest
Post-op 2 weeks	1. ABC[a]	.48[**] (.23-.70)[e]	.35[*] (.08-.58)	.42[**] (.14-.64)	43[**] (.15-.65)	.34[*] (.08-.56)
	2. BBS[b]		.67[**] (.53-.80)	.78[**] (.70-.85)	.72[**] (.62-.82)	74[**] (.66-.81)
	3. FGA[c]			.81[**] (.68-.89)	.79[**] (.64-.90)	.72[**] (.54-.86)
	4. BESTest[d]				.93[**] (.89-.96)	.91[**] (.86-.94)
	5. Mini-BESTest					88[**] (.80-.93)
Post-op 12 weeks	1. ABC[a]	.32[f*] (.01-.57)	.47[**] (.21-.67)	.485[**] (.23-.66)	40[**] (.15-.60)	.40[**] (.14-.60)
	2. BBS[b]		.51[f**] (.24-.70)	.68[f**] (.46-.82)	.58[f**] (.36-.74)	.64[f**] (.39-.81)
	3. FGA[c]			.80[**] (.67-.90)	.78[**] (.63-.88)	.63[**] (.41-.80)
	4. BESTest[d]				.93[**] (.86-.96)	.91[**] (.85-.95)
	5. Mini BESTest					85[**] (.75-.92)
Post-op 24 weeks	1. ABC[a]	.33[f*] (.03-.58)	.41[**] (.22-.65)	.48[**] (.25-.71)	.47[**] (.20-.74)	.50[**] (.24-.71)
	2. BBS[b]		.43[f**] (.16-.67)	64[f**] (.43-.79)	.55[f**] (.31-.73)	.71[f**] (.53-.84)
	3. FGA[c]			.73[**] (.56-.84)	.65[**] (.53-.77)	.59[**] (.38-.73)
	4. BESTest[d]	.			.89[**] (.83-.95)	.91[**] (.86-.95)
	5. Mini-BESTest					.82[**] (.74-.88)

[a]Activities-specific Balance Confidence Scale

[b]Berg balance scale

[c]Functional Gait Assessment

[d]Balance Systems Evaluation Test

[e]Values presented are correlation coefficients (95%CI)

[f]Spearman's rho was used because BBS was not normally distributed at post-op 12 and 24 weeks. Pearson's correlation coefficients were used to generate the results otherwise.

*p<0.05, **p<0.01

Table 4. Floor and ceiling effects

	At Post-op 2 weeks			At Post-op 12 weeks			At Post-op 24 weeks		
Functional Outcome	Skewness (γ_1)	Floor Effect (%)[e]	Ceiling Effect (%)[f]	Skewness (γ_1)	Floor Effect (%)[e]	Ceiling Effect (%)[f]	Skewness (γ_1)	Floor Effect (%)[e]	Ceiling Effect (%)[f]
BESTest (0-100%)	0.56	0.0	0.0	-0.70	0.0	0.0	-0.55	0.0	2.2
Mini-BESTest (0-32)	0.29	0.0	0.0	-0.72	0.0	2.2	-0.70	0.0	4.4
Brief-BESTest (0-24)	0.69	0.0	0.0	0.04	0.0	8.7	-0.19	0.0	8.9
BBS (0-56)	-0.81	0.0	10.9	-6.26	0.0	52.2*	-1.71	0.0	57.8*
FGA (0-30)	-0.15	0.0	0.0	-1.07	0.0	8.7	-1.32	0.0	17.8
ABC (0-100%)	-0.20	0.0	0.0	-0.50	0.0	4.3	-1.62	0.0	4.4

[a]Balance Systems Evaluation Test

[b]Berg balance scale

[c]Functional Gait Assessment

[d]Activities-specific Balance Confidence Scale

[e]Floor effect: proportion of participants with the lowest possible score

[f]Ceiling effect: proportion of participants with the highest possible score

*Significant ceiling effect (>20%)

Figure 1. Score distribution at post-op 24 weeks