LETTER TO THE EDITOR

Check for updates

ON THE STRUCTURAL VALIDITY OF BESTEST, MINI-BESTEST AND BRIEF-BESTEST: EXPLORE FIRST AND ONLY THEN CONFIRM

We read with interest the paper by Miyata et al. (1), which aimed to clarify and compare the structural validity of 3 balance scales of the Balance Evaluation Systems Tests (BESTest, Mini-BESTest, and Brief-BESTest) in older adults with femoral or vertebral fracture.

Evaluation of the internal structure of a scale is important; in fact, it is one of the first steps required to define a high-quality outcome measure (2). However, some features of this study are open to debate and there is need to carefully weigh up its conclusions, which could induce clinicians to make wrong practical decisions.

As a premise, we underline that:

- the sample size was small for factor analysis (*n*=94), as the authors acknowledge;
- participants had a history of femoral or vertebral fracture due to fall, but the fact of falling does not necessarily mean that they had a balance problem, and, conversely, such fractures can lead to biomechanical problems producing some idiosyncratic item behaviour;
- most participants seem to show a mild-to-moderate deficit of dynamic balance, as also demonstrated by the absence of variability in the responses to the item "stance on firm surface, eyes open" (which was excluded from the subsequent analysis because its score was perfect in all subjects), and by the extremely low difficulty (with very small outfit value) of the item "sit to stand";
- data were collected from 3 hospitals and their (intra- and inter-centre) reliability was not reported;
- based on the subjects' performance on the BESTest, the same therapist(s) provided a rating according to specific scoring criteria of the Mini-BESTest and Brief-BESTest (thus the 2 latter scales were not directly administered).

Bearing these points in mind, we would like to make the following comments:

About the 6 BESTest domains

Miyata et al. (1) correctly report that the BESTest was developed "from a theoretical understanding of 6 postural control systems". However, a theory must always be tested by subjecting deductive hypotheses to scientific scrutiny through empirical tests: and only if the hypotheses are experimentally confirmed can the theory from which they were deduced be considered valid. Regrettably, as far as we know, the hypothesis regarding the existence of 6 distinct domains of postural control underlying the BESTest has never received experimental confirmation (e.g. through a good structural analysis of the tool). This is not surprising, since balance is achieved by a complex integration and coordination of multiple body systems. Thus, a comprehensive clinical assessment of balance based on systemic assessments (for both diagnostic and therapeutic reasons), such as in the BESTest, may require tests covering a multidimensional construct (with implications for the related measurement).

More generally, a confirmatory factor analysis (CFA) of the BESTest would require not just "a clear hypothesis", but also some empirical evidence about its underlying structure, such as through a less restrictive exploratory factor analysis (EFA) (3). However, 2 EFA studies demonstrated multidimensionality in the BESTest, with the existence of 3 factors in adults with different balance disorders (4), 4 factors in individuals with stroke (5), and some items failing to meaningfully load in any factor (and some others with salient cross-loadings). Both 1-factor and 6-factor models were rejected. With the above demonstrations, the selection of the Brief-BESTest items according to 6 subsystems of the BESTest, and the use of 4 Mini-BESTest sub-scores, based on 4 out of the 6 claimed BESTest domains, seem scarcely justified and even misleading from a measurement point of view. On the other hand, the original Brief-BESTest failed to fit a unidimensional model in studies examining balance disorders of neurological origin (6, 7). For the above reasons, we think that the reporting of these separate sub-scores without any significant evidence of their meaningfulness represents an arbitrary decision.

Confirmatory factor analysis of BESTest, Mini-BESTest and Brief-BESTest

The 1-factor CFAs performed by the authors on the 3 scales failed to confirm their unidimensionality, while the 4-factor solution for the Mini-BESTest showed good fit indices (1). This finding was expected for the BESTest and Brief-BESTest (based on the studies cited above), but it is quite surprising for the Mini-BESTest. Hence, further analyses and comments are needed for the Mini-BESTest.

Regrettably, we do not know other specific aspects of the 1-factor CFA of the Mini-BESTest (such as model specifications, areas of localized strain, interpretability/strength of the resulting parameter estimates, modification indices, etc.) (3). On the other hand, the finding of a 4-factor model for the Mini-BESTest with a good fit does not mean that this model is the only or optimal

model for the data, if more parsimonious solutions have not been taken into consideration (as William of Occam argued, "entities should not be multiplied without necessity"). Indeed, a good fit can be obtained in 2 ways: (*i*) by a hypothesis that optimally constrains parameters of the model; or (*ii*) by estimating (too) many parameters, which necessarily contributes to good fit no matter what the data are (3). As an example, the high correlation reported (0.86) by the authors in Fig. 1 in their paper (1) between "anticipatory postural adjustments" and "stability in gait" would point to the need to examine a combination of these factors, in order to obtain a more parsimonious solution. However, the "unpacking" of the Mini-BESTest into the 4 claimed BESTest domains leads to questionable structural decisions, e.g. one factor is described by just 2 items, and inevitably affects negatively some major psychometric properties, including the measurement precision at individual level.

Overall, considering the small sample size of this study and the particular characteristics of the participants, a detailed exploratory approach would have shed more light on the structure of the Mini-BESTest data in the database. However, we would recommend a more advanced approach to verify the structural validity of the Mini-BESTest: a Rasch analysis of the full scale (including dimensionality assessment, and person statistics). This would be much more informative than an analysis of its 4 (supposed) subdomains.

Other structural analyses of Mini-BESTest

We offer here our own experimental contribution to this topic: an assessment of the structural validity of the Mini-BESTest through CFA for ordinal data, using polychoric correlations and diagonally weighted least squares as model parameters estimator (Lisrel 8.8, SSI Inc., 2007), through a secondary analysis of data from 3 different studies published by our group. In all 3 studies, EFA and/or Rasch analysis had already demonstrated the essential unidimensionality of the Mini-BESTest: (*i*) Study 1 (4): 115 patients with balance disorders due to different neurological diagnoses; median Mini-BESTest value: 15 points; (*ii*) Study 2 (8): 221 patients with a variety of neurological diseases causing balance impairment; median Mini-BESTest value: 14 points; (*iii*) Study 3 (9): 159 patients with hemiparesis due to a first-ever stroke (<4 months since onset) recruited in Slovenia, Croatia and Italy; median Mini-BESTest value: 9 points at admission; 16 points at discharge. For additional details, refer to the original papers (4, 8, 9).

Table I shows the goodness-of-fit indices related to the CFA of our 3 databases. The standards for examining the fit are the same as those reported in Miyata et

Table I. Goodness-of-fit indices related to our 3 confirmatory factor analyses of the Mini-BESTest

Model	CFI	TLI	RMSEA (95% CI)	SRMR
Study 1 (4)	0.99	0.98	0.064 (0.018–0.095)	0.098
Study 2 (8)	1.00	1.00	0.036 (0.017–0.051)	0.034
Study 3 (9)	0.99	0.99	0.078 (0.064–0.092)	0.059

CFI: comparative fit index; TLI: Tucker-Lewis index; RMSEA: root-mean square error of approximation; SRMR: standardized root-mean-square residual; 95% CI: 95% confidence interval.

al. (1): comparative fit index (CFI) and Tucker-Lewis index (TLI) values should be >0.9 for acceptable ft and >0.95 for good fit; root-mean square error of approximation (RMSEA) should be <0.10 for acceptable fit and <0.06 for good fit; and standardized root-mean-square residual (SRMR) should be <0.10 for acceptable fit and <0.08 for good fit.

The discrepancy between our results and those of Miyata et al. (1) suggests, before thinking that the measurement properties of the scale may differ across population subgroups, the authors should thoroughly check a series of potential (clinical and methodological) sources of bias. For example, a few years ago we listed some methodological points that could negatively influence the analysis of unidimensionality of the Mini-BESTest (10).

CONCLUSION

First, regarding the BESTest, we think that the structural analyses point towards the presence of multidimensionality, with a still uncertain number of distinct factors. This finding questions the use of a single summed score for this tool.

Secondly, some modifications have already been suggested for the Brief-BESTest, according to a series of psychometric analyses (7).

Thirdly, concerning the Mini-BESTest, the results of Miyata et al. (1) (coming from a relatively small sample with "peculiar" characteristics) disagree with the unidimensional model that usually fits this scale (4, 7, 8) and which our present analyses confirm. Due to the above sources of concern stemming from the study under discussion, we think that at present there is not sufficient evidence to propose the use of 4 separate Mini-BESTest sub-scores for clinical decision-making, and further discussion and high-quality research is needed in this field.

For these reasons, we wonder if Miyata et al. could kindly provide, in the light of our comments, supplementary material that might be useful as a further step in the complex ongoing discussion about the internal structure of these 3 balance scales.

The authors have no conflicts of interest to declare.

REFERENCES

1. Miyata K, Hasegawa S, Iwamoto H, Otani T, Kiazu Y, Shinohara T, et al. Comparison of the structural validity of three Balance Evaluation Systems Test in older adults with femoral or vertebral fracture. J Rehabil Med 2020; 52: jrm00079.
2. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018; 27: 1147–1157.
3. Brown TA. Confirmatory factor analysis for applied research. 2nd edn. New York: Guilford Press; 2015.
4. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using the psychometric techniques to improve the Balance Evaluation Systems Test: the Mini-BESTest. J Rehabil Med 2010; 42: 323–331.
5. Miyata K, Hasegawa S, Iwamoto H, et al. Structural validity of Balance Evaluation Systems Test assessed using factor and Rasch analyses in patients with stroke. J Phys Ther Sci 2018; 30: 1446–1454.
6. Bravini E, Nardone A, Godi M, Guglielmetti S, Franchignoni F, Giordano A. Does the Brief-BESTest meet classical test theory and Rasch analysis requirements for balance assessment in people with neurological disorders? Phys Ther 2016; 96: 1610–1619.
7. Godi M, Giardini M, Arcolin I, Ferrante S, Nardone A, Corna S, Colombo R. Is the Brief-BESTest brief enough? Suggested modifications based on structural validity and internal consistency. Phys Ther 2019; 99: 1562–1573.
8. 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; 51: 429–437.
9. Goljar N, Giordano A, Schnurrer Luke Vrbanić T, Rudolf M, Banicek-Sosa I, Albensi C, et al. Rasch validation and comparison of Slovenian, Croatian, and Italian versions of the Mini-BESTest in patients with subacute stroke. Int J Rehabil Res 2017; 40: 232–239.
10. Giordano A, Franchignoni F, Bravini E, Ferriero G. On "Benka Wallén M, Sorjonen K, Löfgren N, Franzén E. Structural validity of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in people with mild to moderate Parkinson disease". Phys Ther 2016, 96: 1843–1845.

Franco Franchignoni, MD[1] and Andrea Giordano, PhD[2]
From the [1]Istituti Clinici Scientifici Maugeri, IRCCS/ Maugeri Scientific Clinical Institutes, Institute of Care and Scientific Research, Physical and Rehabilitation Medicine Unit, Institute of Tradate, Tradate (VA), and [2]Istituti Clinici Scientifici Maugeri IRCCS, Bioengineering Unit, Institute of Veruno, Veruno (NO), Italy
E-mail: franco.franchignoni@icsmaugeri.it