J Rehabil Med 2005; 37: 306–311

PSYCHOMETRIC PROPERTIES OF THE MINI-MENTAL STATE EXAMINATION IN PATIENTS WITH ACQUIRED BRAIN INJURY IN TURKEY

Atilla H. Elhan,[1] Sehim Kutlay,[2] Ayse A. Küçükdeveci,[2] Çigdem Çotuk,[2] Gülsah Öztürk,[2] Luigi Tesio[3] and Alan Tennant[4]

From the Departments of [1]Biostatistics and [2]Physical Medicine & Rehabilitation, School of Medicine, Ankara University, Turkey, [3]Unit of Rehabilitation Research, Istituto Auxologico, Milan, Italy and [4]Academic Unit of Musculo-Skeletal and Rehabilitation Medicine, University of Leeds, UK

Objective: **To evaluate the psychometric properties of Mini-Mental State Examination (MMSE) in patients with acquired brain injury in Turkey.**

Methods: **A total of 207 patients with acquired brain injury were assessed. Reliability was tested by internal consistency and the person separation index; internal construct validity by Rasch analysis; external construct validity by correlation with cognitive disability; and cross-cultural validity by differential item functioning analysis compared with Italian MMSE data.**

Results: **Reliability was adequate with a Cronbach's alpha of 0.75 and person separation index of 0.76. After collapsing some categories, and adjustment for differential item functioning, internal construct validity was supported by fit of the data to Rasch model. Differential item functioning for culture was found in 2 items and after adjustment, data could be pooled between Turkey and Italy. External construct validity was supported by expected associations.**

Conclusion: **The Turkish version of the Mini-Mental State Examination can be used as a cognitive screening tool in acquired brain injury. Cross-cultural validity between Italy and Turkey is supported, given appropriate adjustment for differential item functioning. However, shortfalls in reliability at the individual level, as well as the presence of differential item functioning suggest that a better instrument should be developed to screen for cognitive deficits following acquired brain injury.**

Key words: Mini-Mental State Examination, Rasch analysis, differential item functioning, outcomes, cross-cultural, cognition, rehabilitation.

J Rehabil Med 2005; 37: 306–311

Correspondence address: Ayse A. Küçükdeveci; Gökkusagi Sitesi, Mor Blok, 12/25, Karakusunlar 06530 Ankara, Turkey.
E-mail: ayse@tepa.com.tr

Submitted May 28, 2004; accepted March 11, 2005

INTRODUCTION

Assessment of cognitive function is essential in neuro-rehabilitation practice. As cognitive impairment can limit functional gains during inpatient rehabilitation, the early and follow-up assessment of cognitive skills is crucial in the management of brain-injured patients (1, 2).

The Mini-Mental State Examination (MMSE), first introduced by Folstein et al. in 1975 (3), has been suggested as a valuable, consistent and rapid cognitive screening instrument for routine initial assessment in neurological patients (4). It comprises 6 domains of cognition: orientation, registration of new information, attention and calculation, recall, language and visuospatial construction. These domains have a variable number of items giving a total of 11 items for the scale. The items have a different number of response categories ranging from dichotomous to a 6-category response. Thus 11 items are summed to give a maximum score of 30. Several studies suggest an optimum cut-off point of 23/24 to separate patients with cognitive impairment from those who are cognitively intact (3, 5, 6).

The MMSE is one of the most widely used cognitive screening instruments in Europe (7), and has been used within different cultural and ethnic groups and translated into many languages (8–10). Factors such as language, low levels of education and ethnic origin can adversely affect performance on tests of cognitive functioning (5, 8). In order to eliminate bias, some authors provide MMSE norms specific for age and education in non-disabled adults (5, 9–11).

While the MMSE has been used in neuro-rehabilitation settings as a screening tool for research purposes (1, 12), only the reliability and validity for those with mild dementia has been established for the Turkish population (13). Thus as the validity of this instrument has not been demonstrated for those with brain injury, the aim of this study was to evaluate its psychometric properties in patients with acquired brain injury in Turkey and to decide its feasibility for clinical use. This becomes particularly important in the context of recent findings showing the lack of validity of the MMSE in a normal population in Turkey (14). In addition, its potential for use in a cross-cultural setting was also considered.

The Rasch measurement model (15) was used to evaluate the MMSE and identify implications for clinical practice. This mathematical measurement model is increasingly used to evaluate existing scales and to examine issues such as cross-cultural validity (16).

 ISSN 1650-1977
DOI 10.1080/16501970510037573

METHODS

Translation procedure

The original MMSE was translated into Turkish by 3 bilingual professionals. Two of these were medical doctors, and were regarded as "informed" translators. The third translator was an English teacher. Inconsistencies in the translations were resolved by discussion among the translators. Most of the items on the MMSE could be translated directly and used in Turkish. The standard scoring and administration procedures for the MMSE were followed (6).

Three items needed modification:

- Registration: among the words "Apple, penny, table" in the original MMSE, penny has been modified. As the Turkish equivalent has 2 words (*bozuk para*), "money" (in Turkish, *para*) is used instead.
- Language/repeat phrase: as there is no suitable Turkish counterpart for the repetition phrase "no ifs, ands, or buts", the phrase has been translated into Turkish literally as *egerler, veler veya amalar yok*.
- Language/3-step command: the third stage of 3-step command in the original version "Take this paper in your right hand, fold it in half and put the paper on the floor" has been modified to "rest the paper on your legs", since "put the paper on the floor" is inappropriate for some patients.

Once translated, field-testing for face validity was performed in a group of 30 literate patients with variable educational levels with various musculoskeletal disorders such as osteoarthritis, low back pain and radicular pain syndrome. At this stage, most patients found the "repeat phrase" item difficult to understand. However, no direct conceptual equivalent could be found and so a direct literal translation was retained.

Patients

The study was performed in the Department of Physical Medicine and Rehabilitation at the Medical School of Ankara University, Turkey, from November 2000 to September 2003. A total of 207 patients with consecutive acquired brain injury who had been admitted for rehabilitation were included in the study. Patients with significant difficulties in language expression or comprehension or with a history of previous dementia were excluded. For the cross-cultural validation, Italian data were obtained from the stroke sample of a European study of the use of rehabilitation outcome measures (17).

Reliability

The internal consistency of the adapted version of the MMSE was evaluated first. The internal consistency of an instrument is an estimate of the degree to which its constituent items are interrelated, and is assessed by Cronbach's alpha coefficient (18). Reliability was also evaluated with the person separation index from the Rasch analysis (see below). This is similar to coefficient alpha, but uses the metric latent trait in place of the summed score. A person separation index of 0.7 and above is consistent with the scale being able to differentiate at least 2 strata of patients (19) and is considered the minimum requirement for measurement.

Internal and cross-cultural validity

Internal validity is concerned with the integrity of the defined construct (20). The internal construct validity, and the cross-cultural validity of the Turkish adaptation of the MMSE were assessed using the Rasch measurement model (15, 21). The Rasch model is a unidimensional model which asserts that the easier the item (task) the more likely it will be passed, and the more able the person, the more likely they will pass an item (or be able to do a task) compared with a less able person. The Rasch model is used extensively within the medical outcomes field to test the unidimensionality of scales, the ways in which their categories function, whether or not items are "biased" for key subgroups such as age or gender (Differential Item Functioning (DIF)) and whether or not scales work in the same way across countries (17, 22–24).

The model used in the current analysis is the Partial Credit Model (25):

$$\ln\left(\frac{P_{nik}}{1-P_{nik-1}}\right)=\theta_n-b_{ik} \qquad [1]$$

where P is the probability of person n affirming category k in item i, compared with an adjacent category (k–1); θ is person ability, b is the item difficulty parameter τ_k is the difficulty of the k threshold.

A sequence of analysis establishes the internal- and cross-cultural validity of the Turkish MMSE. Initially, as some of the items in the MMSE are polytomous, an analysis is undertaken of the ordering of each category. The issue here is whether or not the transition from a lower to a higher response category within an item is consistent with an increase in the underlying trait. Where this does not occur "disordered thresholds" are said to be present, and categories are collapsed before fit to the model is examined (17). Following this, the data are then (re)fitted to the Rasch model to determine overall fit, and how well each item fits the model.

Two overall fit statistics summarize item- and person-fit and are distributed such that a mean of zero and standard deviation of 1 indicate perfect fit to the model. A chi-square interaction statistic determines the invariance of the scale across the trait (mental status), and should show a non-significant deviation from model expectation. Individual item chi-square statistics should also indicate non-significant deviations, and item residual statistics should be within the range ± 2.5 (21).

Within the framework of Rasch measurement, the scale should work in the same way, irrespective of which group is being assessed. Thus, in the case of mental status, the probability of a person affirming an item (or category), at any given level of mental state, should be the same for younger or older people, men and women, Turks and Italians, and so on. Items that do not yield the same item response function for 2 or more groups display DIF and violate the requirement of unidimensionality (26). Consequently, every item is checked for DIF by age, gender and education and, for the cross-cultural analysis, by country.

Finally, person-item deviation residuals (the residual for each person on every item) were examined with Principal Components Analysis (PCA) for associations, which may be indicative of the breach of the assumptions of local independence. Thus after the "Rasch" factor there should be no discernible pattern left in the residuals. The absence of such patterns, taken with adequate fit to the Rasch model, supports the unidimensionality of the scale.

External construct validity and responsiveness

External construct validity is concerned with whether or not the scale measures what it intends. It is assessed through convergent validity with the Functional Independence Measure (FIM) cognitive scale, previously adapted for use in Turkey (27). Although a measure of disability rather than impairment, a moderate correlation would be expected (1). Responsiveness is evaluated through the Effect Size (28).

Sample size and statistical software

For the Rasch analysis it is reported that a sample size of 150 patients will estimate item difficulty, with α of 0.01, to within ± 0.5 logits (29). This sample size is also sufficient to test for DIF where, at α of 0.01 a difference of 0.5 within the residuals can be detected for any 2 groups with β of 0.20. Bonferroni corrections are applied to both fit and DIF statistics due to the number of tests undertaken (30). A value of 0.004 is used throughout.

The Wilcoxon signed rank test was used for evaluating change over time and Spearman's rank correlation for associations between instruments. Statistical analysis was undertaken with SPSS-11, and Rasch analysis with the RUMM2020 package (31).

RESULTS

Patients

The mean age of the 207 Turkish patients was 57.9 years (standard deviation (SD) 15.3), 64% were men, and patients had a mean time since injury of 152.3 days. The cause of the acquired brain injury was predominately stroke (84%) and almost 1-in-8 patients were illiterate. In addition, the data from 105 stroke patients in Italy with a mean age of 71.9 years (SD 12.0) were included for cross-cultural analysis, 61% of whom were women. There was a lack of gender equivalence between the 2 samples.

Table I. *Fit of Mini-Mental State Examination to Rasch model (after rescoring)*

Item	Location	SE	Residual	Chi-square	*p*
Orientation time	0.24	0.07	−1.90	3.23	0.357
Orientation place	−0.03	0.12	−1.56	6.80	0.079
Registration	−3.95	0.54	−0.27	0.88	0.829
Attention and calculation	1.42	0.16	−1.36	12.32	0.006
Recall	1.42	0.09	3.06	17.08	0.001*
Language naming	−3.53	0.45	−0.01	2.29	0.514
Language repeat phrase	2.91	0.20	1.53	3.40	0.325
Language 3-step command	−1.44	0.22	−0.64	1.89	0.594
Language comprehension	−0.49	0.18	−1.00	3.42	0.330
Language sentence construction	1.39	0.16	−1.13	11.75	0.008
Copy figure	2.06	0.17	−0.71	13.15	0.004

* Misfitting item. SE = standard error.

Reliability

Internal consistency of the Turkish version of MMSE was adequate at the group level with Cronbach's alphas of 0.75 and 0.70 at admission and discharge, respectively. The person separation index of 0.76 on admission indicates that the scale had the ability to differentiate 2 groups of patients (19).

Internal construct validity

Initially, many items displayed disordered thresholds, necessitating collapsing of categories. For example, while the item "orientation time" displayed ordered thresholds, in that each threshold demonstrates an increasing level of the trait of mental state, the item "attention" failed to display ordering such that it was necessary to dichotomize the item. In total 6 items displayed disordered thresholds and were re-scored. Following this, all items apart from "recall" were found to fit the model (given a Bonferroni adjustment fit level of 0.004) (Table I). Overall mean item fit was −0.362 (SD 1.465) and mean person fit was −0.224 (SD 0.515). However, the item-trait interaction was significant, suggesting that cumulatively, across the scale, there was a lack of invariance of items (chi-square (df = 22) 52.5, $p = 0.00026$). The person separation index was satisfactory (0.764) indicating the ability of the scale to differentiate more than 2 groups of patients (19). Overall, the scale was well targeted in that the measurement, expressed through the distribution of the thresholds, covered all levels of patients across the trait (Fig. 1). With a mean person score of 0.915, patients in this study displayed a slightly higher average mental status than the average item difficulty.

The "attention and calculation" item displayed DIF by both age and education. The items "recall", "language/comprehension" and "language/sentence construction" displayed substantive DIF for educational level, with the last 2 showing that the illiterate group were scoring below expectation (Fig. 2). Due to the low number of illiterate patients, these 3 items were first grouped into 1 item (that is their responses were added together to avoid extreme scores), and the analysis re-run with 9 items (8 originals, plus 1 item made up of the sum of scores from the 3). Furthermore, the new grouped item was split such that 1 new item related to just those who were illiterate, and the other items to the other education groups, making 10 items in all. Following this modification, good fit to the Rasch model was attained. Overall mean item fit was −0.437 (SD 0.912) and person fit of −0.252 (SD 0.481). The item-trait interaction was now non-significant, supporting the invariance of items across the scale (chi-square (df = 19) 37.8, $p = 0.007$). The person separation index was 0.77, indicating the ability of the scale to differentiate at least 2 groups of patients (19).

Finally, PCA analysis of residuals showed no discernible pattern, with the first factor taking 18% of the variation amongst the residuals, so supporting the assumption of unidimensionality.

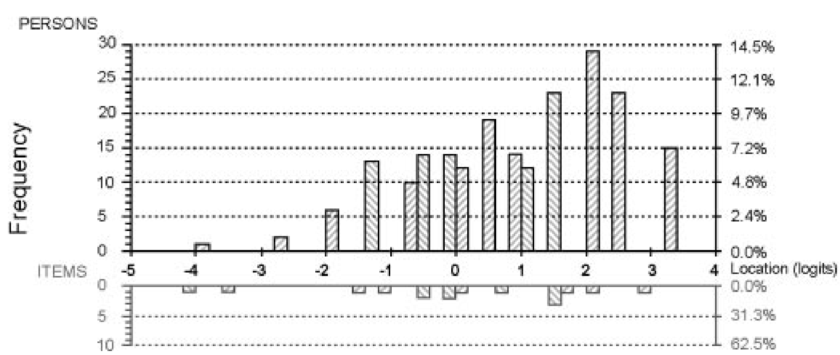

Fig. 1. Targeting of scale to patient ability (after re-scoring). Person-Item Threshold Distribution (grouping set to interval length of 0.20 making 45 groups).

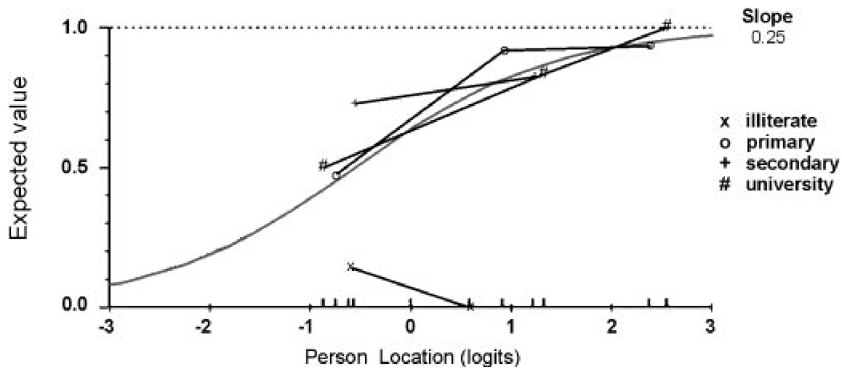

Fig. 2. Differential Item Functioning for item [10009] "language/comprehension" by education level.

Cross-cultural validity

The Italian and Turkish stroke data were pooled to test for invariance across cultures. Due to the potential confounding effect of education (for example all illiterates were within the stroke group) only those completing at least middle education in Turkey were included in the analysis. Likewise, because of gender differences between the 2 samples, the absence of DIF by gender was first evaluated for all items. After re-scoring overall mean item fit was -0.510 (SD 1.030) and person fit of -0.200 (SD 0.424). The item-trait interaction was significant, indicating a lack of invariance of items across the scale for the pooled data (chi-square (df = 33) 66.0, $p = 0.0006$). The person separation index was 0.895, indicating the ability to distinguish 4 groups of patients across the construct (19). Two items "language/repeat phrase" and "attention and calculation" were found to display DIF across the 2 countries. Thus, for example, at the same level of cognitive impairment, Italian patients were more likely to correctly repeat the phrase item than Turkish patients (Fig. 3). The items were split for country and following this, overall mean item fit was -0.561 (SD 0.860) and person fit of -0.229 (SD 0.454), indicating good fit to model expectation. The item-trait interaction improved although still just significant (chi-square (df = 39) 68.4, $p = 0.003$). Nevertheless no single item displayed significant misfit to the model ($p > 0.004$) and thus the scales can be considered equivalent and, given the split items above, data can be pooled.

External construct validity and responsiveness

The mean MMSE score of the 207 patients at admission was 19.3 (SD 6.1, median: 19, min–max: 4–29). The Spearman correlation between the FIM cognitive scale and the MMSE was 0.60 at admission and 0.53 at discharge, confirming convergent validity with an expected moderate association between cognitive impairment and cognitive disability. However, as expected, a weaker correlation was found between MMSE and FIM motor scale ($r = 0.29$ at admission, 0.30 at discharge). The effect size for the MMSE was 0.60 compared with 0.19 for the FIM cognitive scale. The effect size for the FIM motor scale was 0.74 showing that, as expected, the change in physical function was the greatest.

DISCUSSION

The MMSE is the most widely used screening measure of cognitive impairment. It has the advantages of brevity, ease of administration and high inter-rater reliability. It can easily be incorporated into routine clinical practice and provides a rough and ready evaluation of cognitive function. It has some

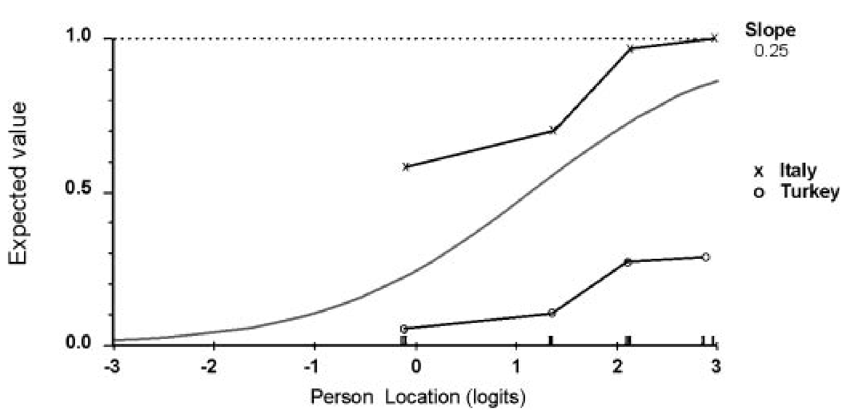

Fig. 3. Differential Item Functioning by country for item [10007] "language/repeat phrase".

limitations, for example in detecting focal deficits and insensitivity to frontal lobe disorders (6, 32). The present study investigates the reliability and validity of the MMSE as a brief cognitive screening test amongst acquired brain injury patients in Turkey. The reliability of the present version was supported with internal consistency at group levels (0.75 and 0.70 at admission and discharge), comparable to other cognitive screening instruments in acquired brain injury (33). However, this level of reliability is disappointing for a screening tool, which is routinely used at the individual patient level.

Internal construct validity of the scale was evaluated by fit to the Rasch model. There was DIF for age and education in item attention/calculation. Moreover, 3 items showed DIF by education. These findings suggested that some of the items were affected by illiteracy and that the scale works differently in the cognitive assessment amongst illiterate people. This influence of age and education on the MMSE has been well documented in other studies (5, 9, 11, 34). Although a cut-off score of 23/24 was initially suggested for distinguishing between impaired and normal subjects, later, it has been suggested that cut-off scores should be adjusted according to age and educational level. Furthermore, population-based norms by age and education have been defined in some countries (5, 10), but the scale has been found to be invalid in a normal population in Turkey (14).

Comparison of the Turkish MMSE with the Italian counterpart found 2 items presenting DIF. DIF in the item "language/repeat phrase" might be due to the fact that the phrase used in Italy was a more common item in that culture compared with the phrase used in Turkey. Thus the cross-cultural adaptation has delivered a scale, which is shown to be similar to that used in Italy. However, this comparison is only available after complex analysis, which would not normally be available outside research groups with the necessary resources. Substantive work on large samples would need to be undertaken to provide the information necessary for comparisons in the absence of such technologies. This experience of this adaptation process is such that the rehabilitation outcome measurement cannot be taken for granted, and that instruments, even when using state-of-the-art adaptation procedures, cannot be considered equivalent without appropriate analysis to test for invariance across cultures. Furthermore, cumulative deficits in the quality of the instrument with respect to its reliability and internal construct validity, including its scoring functions, suggest that the scale is less than ideal when judged against modern psychometric standards. This would suggest that those with data from other countries should re-examine the scale using the same methodology to assess whether or not the same limitations apply elsewhere.

Given this, the external construct validity of the scale was demonstrated by expected associations with physical and cognitive disability. Similar associations have been reported elsewhere (1, 2, 12). The scale has also been shown to have a respectable effect size compared with the FIM cognitive scale.

In conclusion, the Turkish version of the MMSE can be used as a cognitive screening tool in acquired brain injury, given the limitations mentioned above. Its cross-cultural validity has also been established between Italy and Turkey, but only after adjustments for cultural differences. If data are to be pooled across countries then formal tests of invariance must be undertaken before any confidence can be made in the findings. Furthermore, insufficient reliability of the MMSE at the individual level, the lack of discrimination across categories for some items, and the presence of DIF for some items, as well as acknowledged shortfall in the range of cognitive areas covered, suggest that in the medium-term a more comprehensive scale with better psychometric properties should be sought.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support of the Scientific and Technical Research Council of Turkey for this project. The European data for the cross-cultural analysis was funded by the European Commission within its BIOMED 2 programme under contract BMH4-CT98-3642. This study was conducted in the Department of Physical Medicine & Rehabilitation, School of Medicine, University of Ankara, Ankara, Turkey.

REFERENCES

1. Zwecker M, Levenkrohn S, Fleising Y, Zeiling G, Ohry A, Adunsky A. Mini-Mental State Examination, cognitive FIM instrument, and the Loewenstein Occupational Therapy Cognitive Assessment: relation to functional outcome of stroke patients. Arch Phys Med Rehabil 2002; 83: 342–345.
2. Hajek VE, Gagnon S, Ruderman JE. Cognitive and functional assessments of stroke patients: an analysis of their relation. Arch Phys Med Rehabil 1997; 78: 1331–1337.
3. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975; 12: 189–198.
4. Dick JPR, Guiloff RJ, Stewart A, Blackstock J, Bielawska C, Paul EA. Mini-Mental State Examination in neurological patients. J Neurol Neurosurg Psychiatry 1984; 47: 496–499.
5. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population based norms for the Mini-Mental State Examination by age and educational level. JAMA 1993; 269: 2386–2391.
6. Hodges JR. Cognitive assessment for clinicians. Oxford University Press: Oxford; 1994.
7. Haigh R, Tennant A, Biering-Sorenson F, Grimby G, Marincek C, Philips S, et al. The use of outcome measures in physical medicine and rehabilitation within Europe. J Rehabil Med 2001; 33: 273–278.
8. Mungas D, Marshall SC, Weldon M, Haan M, Reed BR. Age and education correction of the Mini-Mental State Examination for the English and Spanish speaking elderly. Neurology 1996; 46: 700–706.
9. Ostrosky-Solis F, Lopez-Arango G, Ardilla A. Sensitivity and specificity of the Mini-Mental State Examination in a Spanish speaking population. Appl Neuropsychol 2000; 7: 25–31.
10. Grigoletto F, Zappala G, Anderson DW, Lebowitz BD. Norms for the Mini-Mental State Examination in a healthy population. Neurology 1999; 2: 315–320.
11. Jorm AF, Scott R, Henderson AS, Kay DWK. Educational level differences on the Mini-Mental State: the role test bias. Psychol Med 1988; 18: 727–731.
12. Diamond PT, Felsenthal G, Macciocchi SN, Butlet DH, Lally-Cassady D. Effect of cognitive impairment on rehabilitation outcome. Am J Phys Med Rehabil 1996; 75: 40–43.
13. Gungen C, Ertan T, Eker E, Yasar R, Engin F. Reliability and validity of the standardized Mini-Mental State Examination in the diagnosis of mild dementia in Turkish population. Turk Psikiyatri Derg 2002; 13: 273–281.
14. Küçükdeveci AA, Kutlay S, Elhan AH, Tennant A. Preliminary study to evaluate the validity of the Mini-Mental State Examination

(MMSE) in a normal population in Turkey. Int J Rehabil Res 2005; 28: 77–79.
15. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960 (reprinted 1980).
16. Lundgren-Nilsson Å, Grimby G , Ring H, Tesio L, Lawton G, Slade A, et al. Cross-cultural validity of FIM items in stroke in six European countries: a Rasch perspective. J Rehabil Med 2005; 37: 23–31.
17. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model. The PRO-ESOR project. Med Care 2004: 42: I-37–I-48.
18. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16: 297–334.
19. Fisher WP. Reliability statistics. Rasch Measure Trans 1992; 6: 238.
20. Nunally JC Jr. Psychometric theory (2nd edn). New York: McGraw-Hill; 1978.
21. Andrich D. Rasch Models for Measurement. London: SAGE Publications; 1988.
22. Prieto L, Alonso J, Lamarca R, Wright BD. Rasch measurement for reducing the Items of the Nottingham Health Profile. J Outcome Measure 1998; 2: 285–301.
23. Hagquist C, Andrich D. Is the Sense of Coherence-instrument applicable on adolescents? – a latent trait analysis using Rasch-modeling. Pers Indiv Differ 2004; 36: 955–968.
24. Küçükdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A. Issues in cross-cultural validity: example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. Arthritis Rheum 2004; 51: 14–19.
25. Master GN. A rasch model for partial credit scoring. Psychometrika 1982; 47: 149–174.
26. Dorans NJ, Holland PW. DIF detection and description: Mantel-Haenszel and standardisation. In: Holland PW, Wainer H, eds. Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993, p. 36–66.
27. Küçükdeveci AA, Yavuzer G, Elhan AH, Sonel B, Tennant A. Adaptation of the Functional Independence Measure for use in Turkey. Clin Rehabil 2001; 15: 311–319.
28. Kaziz L, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989; 27: S178–S189.
29. Linacre JM. Sample size and item calibration stability. Rasch Measure Trans 1994; 7: 28.
30. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995; 310: 170.
31. Andrich D, Lyne A, Sheridon B, Luo G. RUMM 2010. Perth: RUMM Laboratory; 2000.
32. Feher EP, Mahurin RK, Doody RS, Cooke N, Sims J, Pirozzolo FJ. Establishing the limits of the Mini-Mental State. Examination of subtests. Arch Neurol 1992; 49: 87–92.
33. Alderson AL, Novack TA. Reliable serial measurement of cognitive processes in rehabilitation: the Cognitive Log. Arch Phys Med Rehabil 2003; 84: 668–672.
34. Marshall SC, Mungas D, Weldon M, Reed B, Haan M. Differential item functioning in the mini-mental state examination in English- and Spanish-Speaking older adults. Psychol Aging 1997; 12: 718–725.