A Transcultural perspective on the Mini Mental State Examination

Steven Klimidis & Aslihan Tokgoz

Victorian Transcultural Psychiatry Unit & Centre for International Mental Health

The Mini Mental State Examination (MMSE) is a brief instrument designed to assess cognitive function and developed within Anglophone neuropsychiatry. It has grown in popularity among clinicians and researchers, and particularly within geriatric psychiatry and neurology settings. The brevity of the MMSE and its ease of administration and scoring probably explain its wide popularity among clinicians and researchers alike. While it is important to identify brief, cost-effective, diagnostic tools for clinical settings, a principal concern is the injudicious application of tools such as the MMSE to patients. In neuropsychiatry settings many of the professional 'users' of the MMSE are from disciplines (e.g., nursing, social work, etc.) that, typically, are not trained in the development and administration of psychological tests and often may not appreciate the difficulties deriving from misapplication of the MMSE. Of particular concern here is the application of the MMSE in groups that differ on cultural and linguistic grounds from the original development and standardisation groups. The consequence of this, in immigrant patients, is to attribute low scores to pathological processes rather than to other factors such as education level, literacy, cultural differences in cognitive and perceptual information processing, and so on. Such application can result in the diagnostic error of over-diagnosis (in particular) of significant morbidity, and for treatment to be over-prescribed.

The MMSE

The MMSE was developed by Folstein (Folstein, Folstein & McHugh, 1975). The authors argued that there was a need for a short battery to examine the mental state of psychiatric patients, particularly those with delirium or dementia syndromes (Folstein, Folstein & McHugh, 1975). The MMSE requires only 5-10 minutes to administer (Folstein, Folstein & McHugh, 1975). The MMSE is divided into two sections, the first of which requires only verbal responses and covers orientation (time and place), memory, and attention. The second part assesses the ability to name, follow written and verbal commands, write a sentence spontaneously, and copy a complex polygon. The score maximum is 30, with the cut off score 20 or less reported to discriminate between patients with dementia, delirium or psychoses from healthy elderly and those with primary diagnoses of 'neurosis' or personality disorder. Twenty-four hours retest by single or multiple examiners revealed high retest reliability (r=.887-single examiner and r= .827-multiple examiner). Twenty-eight days retest in elderly depressed and demented patients also revealed high retest reliability (correlation between test 1 and test 2 was .98) (Folstein, Folstein & McHugh, 1975).

The nature of test performance & scores

Scores derived from psychological tests such as the MMSE are affected by a number of factors. These include general factors in the test situation. Within the assessment process deviations in the recommended 'standard' administration procedure can introduce score variation. For example, the memory registration component of the MMSE requires that three objects be presented by the examiner at one second intervals. Deviation from this procedure with faster or slower presentation will, respectively, limit or increase the opportunity for rehearsal of presented material in short term memory, thus leading to weaker or stronger registration of the input. This, in turn, would result in lower or higher performance in the recall test. The examiner's selection of objects to be remembered by the patient may also affect scores. This is to do with 'built-in' biases in human information processing. Three related objects (e.g., desk, pencil and book) may be more readily remembered due to semantic cross-cueing between these objects compared to the condition where unrelated objects are presented. Similarly selecting two related objects and a third which is remarkably different may improve remembering by both semantic cross-cueing and by the effect of novelty on the encoding or registration process. Novelty has an alerting effect, attracting attention to the presented material. In the memory part of the MMSE, recall is delayed by the presentation of an 'attention' task (serial 7's or spelling the word 'world' backwards). The degree of delay and experienced distraction may be quite different between those able and less able to do the distraction task, introducing greater disadvantage to the less able in the succeeding task of delayed recall. Additional general factors affecting test scores include test-taking attitude, motivation and strategy. Test attitude has to do with the patient's approach to the examination and, indeed, the examiner. Is the patient co-operative, is he involved with the examination, does he appear motivated, does he persist in the face of failure? These are important considerations in evaluating test attitude. Test attitude may be affected also by the examiner's behaviour during the examination. In cases where there is repeated failure on the tasks this often diminishes a patient's motivation. Experienced examiners will use social reinforcement for participation as one means of maintaining task motivation under repeated failure; obviously scores across a range of testing situations may therefore be affected by the examiner's behaviour. The patient's test taking strategy is another important consideration. This relates to 'how', or by what pathways, the patient manages the task. To take an example, in organising information for the purposes of recall patients will vary in the way that they 'memorise' materials. Some may, for example, use 'imagery' (that is, covertly visualise the objects presented) whereas others may rely on the phonetic elements of the words presented (how they sound). Such strategy variations produce differing recall performances but are often unknown to the examiner. Strategy selection and use is not necessarily under conscious control or conscious decision-making. That is to say, the patient may not have made a decision to use one rather than another strategy or to emphasise one than another in cases were multiple strategies are engaged in simultaneously. Instead, his selection and use may be highly automatic. To understand the automatic nature of information processing consider the difference in conscious decision-making between a learner driver and a highly experienced veteran. The learner reacts to every emerging situation with variable degrees of an 'alarm' response and makes conscious decisions about his physical handling and guiding of the vehicle. An experienced driver, on the other hand, negotiates hundreds of decisions and actions but may in fact be concentrating to her conversation with her passenger.

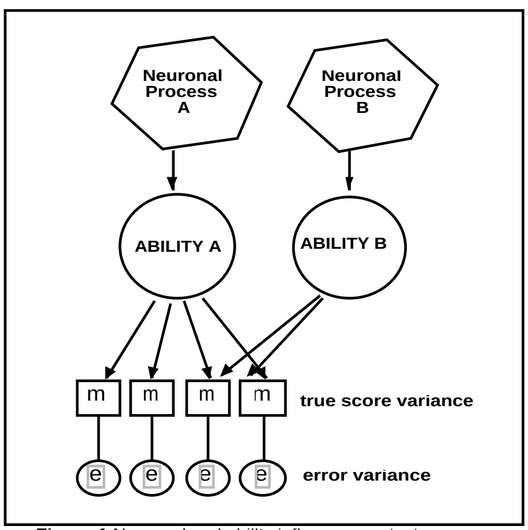

Figure 1 Neuronal and abillty influences on test scores.

The issue of strategy and pathways to performance raises the issue of the complex constitution of tasks. Since the 1940's psychologists have been concerned with identifying fundamental cognitive processes involved in 'intelligence'. A single task used to examine a patient will be a measure of a complex composite of such fundamental processes. Some examples, include 'freedom from distractibility', memory for linguistic materials, figural organisation of two-dimensional representations, etc. Figure 1 shows a simplified schematic of two fundamental cognitive processes, labelled Ability A and Ability B, which may be thought to be 'driven' by underlying neuronal networks, both centrally and in the peripheral nervous system. The overall score on the test shown in the example of Figure 1 is determined by performance on four measurement items (or four opportunities for observation) and is jointly determined by the underlying cognitive processes. The balance of influence of the two cognitive processes on performance may differ between items of the same test. (Scores will also vary according to error of measurement, which we need not go into here.) The complex constitution of test scores, like those from the MMSE, has certain implications with regard to the validity of translation or transfer of a test from one cultural or linguistic group to another, as we discuss next.

Cognitive and perceptual processing across cultures

The extent of potential distortion to the psychometric properties of a test when it is translated or is applied from one cultural group to another is not well appreciated. Psychometric theory indicates that tests can be described with reference to several properties. One form of psychometric theory termed item response theory (IRT) posits that properties of tests can be described against the variation of abilities in the population. When we say for example that a test is easy or hard to pass the reference for stating the difficulty of the test is the proportion of people from the population who are likely to pass that test. Similarly, if an individual is said to be skilled or unskilled by passing a certain test this is a statement that incorporates the position of this test within the overall range of that ability in the population. According to IRT, there are three psychometric properties of tests that may be affected by language translation and by transfer across cultures. These include the difficulty of the test, the range abilities that it is able to discriminate, and the ability to guess the correct answer. Each of these may be altered in two directions (e.g. lowering or increasing test difficulty). So for a simple yes-no response test with a single question there are potentially 16 distortions (across language or culture X increase or decrease difficulty X increase or decrease discriminability X increase or decrease guessing potential) that are possible to the psychometric properties of the test. Returning to the example of Figure 1, altering the test properties could result in changing what underlying abilities the test is measuring.

The issue is complicated further by the different cognitive and perceptual development occurring in people raised in vastly different cultures and physical environments. Members of different cultures are exposed to different types of perceptual and cognitive problems in day-to-day transactions. As mentioned, practice means that certain cognitive processes may produce their outputs more automatically, with little need for conscious monitoring. This implies that across cultures different groups are 'tuned' to more efficient processing of different problems. Early work with Aborigines living in Australian bush settings, for example, indicated their superior ability in spatial localisation under minimal visual cues compared with their European counterparts. Closer to home, the task of oral spelling (e.g. spell the word 'world') in some groups is very difficult because in some languages there is less need to conduct oral spelling due to the clear phonetic nature of the written form. In some groups written language does not have a syllabic structure but relies on figural recognition and rote learning of the correspondence between the spoken word and the figural symbol (pictogram). Thus, test materials that are not validated with respect to a particular cultural group are quite likely to carry different psychometric properties than the original.

Table 1 provides an example of how much the MMSE varies in its ability to discriminate between those with cognitive decline from others when it has been applied in different ethnic, language or cultural groups. The table summarises the sensitivity (the proportion of actual cases that are said by the test to be cases) and the specificity (the complement of the proportion of actual non-cases that are said to be cases by the test) of the MMSE in different samples. As Table 1 indicates, MMSE criteria for the detection of cognitive decline range between 20 and 26. That is to say, 23 percent of possible scores on this test fall in a region of diagnostic uncertainty. So if a patient scored 26 on this test the clinician would not be confident that the patient was cognitively incapacitated or not. Similarly so if the patient scored 20. In addition, the table indicates the best cut-off (i.e. maximising sensitivity and specificity) found in each study. Even so some studies reveal sensitivity values as low as .57 and specificity values as low as 63. In other words, the error rate for the purposes of clinical decisions can be quite high with approximately up to 40 percent of those having cognitive decline being missed by the test and a similar proportion of those without decline are suggested as having decline. In addition, most of the work summarised in Table 1 has noted the effects of additional factors on test scores including, primarily, level of education and the presence of psychopathology such as depression. Other noted factors include: age, literacy, marital status, income level, urban versus rural location of the patient's residence, and, important to the present discussion, immigrant status, and for immigrants the language in which the interview was conducted. Of note the interdependence and replicability of these effects

Table 1 : Summary of recent work with the MMSE in different ethno-linguistic groups

Study (Country or Ethnicity or Language of the sample)	Cut-off criterion	Discriminant validity results
O'Connor et al (1989) (Britain)	23/24	Sensitivity in detecting elderly general practice patients with 'organic mental disorder' from those without was .86 and specificity was .92. (55% of those scoring 23 or less were classified as suffering from mild dementia or delirium and several well educated persons with dementia scored above 23.)
Gagnon et al (1990) (France)	23/24	Contrasting between community-dwelling subjects meeting criteria for dementia according to the DSM III and those not meeting such criteria, the sensitivity was 1.00 and specificity .78.
Ritchie & Fuhrer (1992) (France)	24/25 23/24	In discriminating between a group patients with mild to moderate dementia (based on DSM III criteria) versus non-demented subjects (mainly community dwelling), for the 24/25 criterion the sensitivity was .91 and specificity was .80. For the recommended 23/24 criterion the sensitivity was .84 and specificity .90.
Feher et al (1992) (USA)	23/24 26/27	The standard 23/24 criterion resulted in a sensitivity value of .69 and a value of specificity of .90 in distinguishing between discharged patients with dementia (as defined in DSM-IIIR) and those with non-dementia disorders. However the optimal cut-off in this study was 26/27. For this criterion sensitivity was .78 and specificity was .84.
van der Cammen et al (1992) (Netherlands)	24/25	Sensitivity was .88 and specificity was .82 in discriminating between patients referred to a geriatric outpatient clinic with and without dementia.
Grace et al (1995) (USA)	23/24	Tested against diagnosis made by use of a wider neuropsychological test battery ('intact' versus 'impaired') the sensitivity and specificity values were, respectively, .44 and .84 in discriminating between 'intact' and 'impaired'groups.
Morales et al (1995) (Spain)	21/22	In discriminating between demented versus non-demented (based a neurologist diagnosis using DSMIII-R) subjects the sensitivity was.57 and specificity was .84.
Fuh et al (1995) (Taiwan)	20/21	Using MMSE estimated scores from the Cognitive Abilities Screening Instrument, the sensitivity was.72 and specificity was .81 in discriminating between a group of clinically diagnosed patients with dementia versus non-demented controls drawn from the community.
Mulligan et al (1996) (French, in Switzerland)	23/24 26/27	Sensitivity was.76 and specificity was .93 in discriminating between demented versus non-demented subjects using the 23/24 criterion. The corresponding values for the 26/27 criterion were .91 and .63.
Del-Ser et al (1997) (Spain)	22/23	In discriminating between demented (as defined in DSM-IIIR) versus non-demented neurology outpatients the sensitivity was.89 and specificity was .80.
Morales et al (1997) (Spain)	Urban sample 21/22 Rural sample 20/21	In their urban sample the discrimination of mild dementia (based on neurologist diagnosis using DSM-IIIR criteria) from non-dementia controls revealed a sensitivity value of .73 and specificity value of .78. The corresponding values for their rural sample were, respectively, .83 and .74.
Cossa et al (1997) (Italy)	23/24	Against a neurologist's diagnosis the sensitivity was .86 and the specificity was .90. (Notably those not meeting criteria for a diagnosis of dementia but who suffered from other neurological problems were excluded from calculations of sensitivity & specificity.)

Ortiz et al (1997) (Spanish or English, Mexican-Americans)	23/24	Lowly educated people (those with less than high school education) were 1.39 times more likely to be classified into the impaired range than those with higher education.
Ferrucci et al (1998) (Italy)	Approx. 20/21	From their Figure 3 it can be discerned that the MMSE discriminates between clinically diagnosed cases of dementia (based on DSM IV criteria) from healthy controls with a sensitivity of above .90 and a specificity of approximately .80.
Werner et al (1999) (Israel)	23/24	Discriminates between cases with various types of dementia and healthy controls (Sensitivity= .89 and Specificity= .68; Positive predictive Value = .82); discriminates between cases with Alzheimer's dementia and healthy controls (Sensitivity= .90 and Specificity= .89; Positive predictive Value = .89).
Black et al (1999) (Spanish or English, Mexican-Americans)	23/24	Among other findings, logistic regression indicated independent contributions to membership in the 'cognitively impaired' range by the following factors: no education (odds ratio = 8.0), illiteracy (odds ratio = 14.3), presence of depressive illness (odds ratio = 1.4), and experience of stroke (odds ratio = 3.2).
Erzigkeit et al (2000) (UK, Germany, Spain)	23/24	Discriminates between cases with severe and moderate cognitive decline, and cases with mild or no cognitive decline based on the Global Deterioration Scale; Across three samples (UK, Germany and Spain) the area under the ROC curve for this discrimination ranged between .87 to .94.

in a range of samples are both unclear given the relatively few studies available in the literature. However, the dependence of test scores on factors such as these implies that the clinician needs to consider what alternative hypotheses may explain a patient's poor performance, other than a dementing process.

Recommendations for clinicians

It is unlikely that our brief analysis will convince clinicians to avoid the use of the MMSE in cultural and linguistic groups for which it was not designed or validated so we make some suggestions regarding its use. The following cautions may need to be exercised in the clinical administration and interpretation of the MMSE:

1. Those using the MMSE should receive training in its administration and score interpretation, including a general understanding of issues of psychometric distortion; this also means that when non-English versions of the MMSE are administered through an interpreter that the interpreter should be given similar instruction.
2. Little weight should be given to scores from the MMSE in clinical decisions unless scores are extreme (either very high or very low); the clinician should recognise the wide 'window' of diagnostic uncertainty in scores ranging between 20 and 26 as indicators of deficit; this makes the test especially problematic in cases of borderline cognitive decline.
3. Where MMSE scores are low, corroborative evidence for cognitive decline should be sought by use of other tests, referral to relevantly qualified clinicians (and in the case of immigrants to clinicians with appropriate language and cultural skills), and by use of functional self-report or informant-report measures.
4. Cultural, language, social, and co-morbid conditions may contribute to score distortions including lowering of scores so clinicians need to consider a range of factors that may have led to the score outcome of a particular test administration.
5. Those administering the MMSE in a language other than English ought to avoid spontaneous translations and should use pretranslated versions; where such are available an attempt should be made to discern whether these are validated and that norms are available for the target group; it should be recognised that one's own pre-translation is no better than an unvalidated version of the test.

Acknowledgements

We appreciate Ms Di Gabb’s advise and proof-reading of this work.

References

Black SA, Espino DV, Mahurin R, Lichtenstein MJ, Hazuda HP, Fabrizio D, Ray LA, Markides KS. The influence of noncognitive factors on the Mini-Mental State Examination in older Mexican-Americans: findings from the Hispanic EPESE. Established Population for the Epidemiologic Study of the Elderly. *J Clin Epidemiol 1999* 52:1095-102.

Cossa FM, Sala SD, Musicco M, Spinnler H, Ubezio MC. Comparasion of two scoring of the Mini-Mental State Examination as a screening test for dementia. *J Clin Epidemiol 1997* 50:961-965.

Del-Ser, T, Morales, J-M, Barquero, M-S., Canton, R, & Bermejo, F. Application of a Spanish version of the “Informant Questionnaire on Cognitive Decline in the Elderly” in the clinical assessment of dementia. *Alzheimer Dis Assoc Disord 1997* 11:3-8.

Erzigkeit H, Lehfeld H, Pena-Casanova J, Bieber F, Yekrangi-Hartmann C, Rupp M, Rappard, F, Arnold, K, Hindmarch, I. The Bayer-activities of daily living scale (B-ADL): results from a validation study in three European countries. *Dement Geriatr Cogn Disord 2001* 12:348-58.

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.

Ferrucci L, Del Lungo I, Guralnik JM, Bandinelli S, Benvenuti E, Salani B, Lamponi M, Ubezio C, Benvenuti F, Baroni A. Is the telephone interview for cognitive status a valid alternative in persons who cannot be evaluated by the Mini Mental State Examination? *Aging (Milano) 1998* 10:332-8.

Folstein M, Folstein S, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res. 1975* 12:189-98.

Fuh JL, Teng EL, Lin KN, Larson EB, Wang SJ, Liu CY, Chou P, Kuo BI, Liu HC. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) as a screening tool for dementia for a predominantly illiterate Chinese population. *Neurology 1995* 45:92-6.

Gagnon M, Letenneur L, Dartigues JF, Commenges D, Orgogozo JM, Barberger-Gateau P, Alperovitch A, Decamps A, Salamon R. Validity of the Mini-Mental State examination as a screening instrument for cognitive impairment and dementia in French elderly community residents. *Neuroepidemiology. 1990* 9:143-50.

Grace J, Nadler JD, White DA, Guilmette TJ, Giuliano AJ, Monsch AU, Snow MG. Folstein vs modified Mini-Mental State Examination in geriatric stroke. *Arch Neurol 1995* 52:477-484.

Morales JM, Bermejo F, Romero M, Del-Ser T. Screening of dementia in community-dwelling elderly through informant report. *Int J Geriatr Psychiatry 1997* 12:808-16.

Morales JM, Gonzalez-Montalvo JI, Bermejo F, Del-Ser T. The screening of mild dementia with a shortened Spanish version of the"Informant Questionnaire on Cognitive Decline in the Elderly". *Alzheimer Dis Assoc Disord 1995* 9:105-11.

Mulligan R, Mackinnon A, Jorm AF, Giannakopoulos P, Michel JP. A comparison of alternative methods of screening for dementia in clinical settings. *Arch Neurol 1996* 53:532-536.

O'Connor DW, Pollitt PA, Hyde JB, Fellows JL, Miller ND, Brook CP, Reiss BB. The reliability and validity of the Mini-Mental State in a British community survey. *J Psychiatr Res. 1989* 23:87-96.

Ortiz IE, LaRue A, Romero LJ, Sassaman MF, Lindeman RD. Comparison of cultural bias in two cognitive screening instruments in elderly Hispanic patients in New Mexico. *Am J Geriatr Psychiatry 1997* 5:333-8.

Ritchie K, Fuhrer R. A comparative study of the performance of screening tests for senile dementia using receiver operating characteristics analysis. *J Clin Epidemiol 1992* 45:627-637.

van der Cammen TJ, van Harskamp F, Stronks DL, Passchier J, Schudel WJ. Value of the Mini-Mental State Examination and informants' data for the detection of dementia in geriatric outpatients. *Psychol Rep. 1992* 71:1003-9.

Werner P, Heinik J, Mendel A, Reicher B, Bleich A. Examining the reliability and validity of the Hebrew version of the Mini Mental Sate Examination. *Aging (Milano) 1999 Oct*;11(5):329-34.