The death knoll for the MMSE

Mark, Ruth

Published in:
Journal of Geriatric Psychiatry and Neurology

Document version:
Publisher's PDF, also known as Version of record

Publication date:
2010

Citation for published version (APA):
The Death Knoll for the MMSE: Has It Outlived Its Purpose?

Ruth E. Nieuwenhuis-Mark, PhD

Abstract
The Mini-Mental State Examination (MMSE) is arguably the best-known cognitive screen in the world. Originally designed to assess cognitive impairment in elderly populations, it has become one of the first steps toward a dementia diagnosis. Routinely used in the clinic and in research internationally, the MMSE, despite its flaws, has managed to retain its popularity for more than 30 years. This review explores when and how the test is used, lists its advantages and disadvantages, and ultimately questions its value. The specific issue that is addressed here is whether the test has outlived its original purpose. The conclusion is that although the MMSE may be a useful tool in many circumstances where a cognitive screen is required, practitioners should be wary of using MMSE total scores as a shortcut toward a dementia diagnosis.

Keywords
Mini-Mental State Examination, MMSE, aging, dementia, cognitive screening

The Original Purpose of the MMSE
The test was developed in the 1970s by 2 young physicians (Marshall Folstein and his wife Susan) working under the mentorship of Paul McHugh. Susan would report daily on the mental state of the geriatric patients at Cornell Medical Center’s Westchester Hospital and Marshall would typically ask how she would report whether someone was doing “better” or “worse.” Susan finally asked him to write down the questions he wanted her to ask the patients. He did and the MMSE was born, at the time a truly innovative tool for assessing cognitive functioning in an elderly hospitalized population. The 5- to 10-minute screen was given the name “Mini” because it only assessed aspects of cognitive functioning, ignoring among other things, mood. The authors believed that it thoroughly measured cognition, that it could be used for repeated testing, and that it could differentiate between patients with and without cognitive disturbance. They made these claims after assessing a total of only 206 patients with a wide range of disorders compared to a group of 63 healthy, age-matched controls.

Item Structure of the Original MMSE
The original test is untimed and includes 11 questions divided into 2 sections. The first measures orientation (to time and place), memory and attention and verbal responses are required. The second section measures the ability to name, to follow commands (verbal and written), the ability to spontaneously write a sentence and, finally, to copy a relatively complex figure (two interlocking pentagons). The total score possible for the test is 30. Ramirez et al found that the Flesch Reading Ease Index was 86.4 for the English MMSE and the Flesch-Kincaid grade level was 2.9. This suggests that it

1 Department of Medical Psychology and Neuropsychology, University of Tilburg, Netherlands

Corresponding Author:
Ruth Elaine Nieuwenhuis-Mark, Department of Medical and Neuropsychology, University of Tilburg, Postbox 90153, 5000 LE Tilburg, The Netherlands.
Email: R.E.Mark@uvt.nl
is easy to understand by everyone at US second grade reading level and above making its use among various populations widespread.

**Who Uses the MMSE?**

**The Professionals Who Use It**

Many different professionals use the test. Most primary care physicians believe that cognitive and/or dementia screening is important but that time constraints often prevent it. The MMSE has been recommended as a dementia screen by both the US Preventive Services Task Force and in a report by the American Academy of Neurology among others. The National Institute of Health and Clinical Excellence use MMSE cutoffs to determine Alzheimer Disease (AD) treatment eligibility.

**Populations It Has Been Used in**

Originally intended to be used with psychogeriatric patients, the test has been used in diverse populations including those with a variety of neurological disorders.

**Dementia.** The MMSE has been most often used to assess dementia an obvious development considering the fact that it was designed as a cognitive screen for the older population and the incidence of dementia typically increases with age. While the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for a dementia diagnosis focus on memory loss in the early stages, there are many different types of dementia and memory problems may/may not be the first "sign." The question is whether the MMSE can discriminate between the dementias.

Due to the fact that the MMSE is memory- and language-oriented, it has a bias toward temporal and parietal lobe functioning and as such should be more sensitive in patients with degeneration and/or lesions in these areas. Indeed, patients with AD have been found to perform more poorly than other demented patients on the items "recall of 3 words" and "orientation to time" while those with vascular dementia did more poorly on items of attention. However, Brown et al found that the MMSE could only detect a low 52% of their patients with AD. Furthermore, frontal patients typically suffer more from problems with executive functioning (EF), an aspect of cognition that is underrepresented in the test. Indeed, the MMSE has been found to be poorly sensitive to frontal lobe dysfunction and those with frontotemporal dementia have typically higher total scores and a slower decline than is the case in other forms of dementia. The MMSE may therefore not be the best instrument to use if clinicians want to determine dementia type.

**Parkinson disease dementia (PDD).** The MMSE has also been used to detect dementia in the Parkinson population. Zadikoff et al concluded that the test lacks sensitivity in early PDD. This is perhaps not surprising given the finding that it also lacks sensitivity in differentiating between mild cognitive impairment (MCI) and the early stages of any form of dementia.

**Delirium.** Arsène and Lassaunière compared the MMSE with the Blessed Orientation-Memory-Concentration test (BOMC) in delirious patients who also had cancer and found the BOMC to be better for screening delirium in these patients. McManus et al found a strong correlation between MMSE scores less than 10 and delirium in patients who had just suffered a stroke and concluded that low scores may enhance the identification of patients at risk of delirium in the stroke population. Improvements in MMSE scores may be observed in the short term after a delirious episode has abated, but this may not hold over the long term. The fluctuating nature of delirious episodes makes it particularly difficult to test these patients while such episodes have also been linked with poorer long-term (tested at 12-month follow-up) memory even in those without underlying dementia. O’Keeffe et al found, however, that abrupt declines and improvements in serial MMSE scores could be useful when monitoring the development and resolution of delirium respectively in a population of elderly hospital patients.

**Depression.** The MMSE has been used to assess cognitive functioning in patients with various mood disorders. The problem is how to interpret the scores. Low scores could mean cognitive impairment and possible comorbidity with dementia or simply that patients are not motivated to perform well on the test (as a function of their underlying mood state). The MMSE has however been usefully employed to monitor cognitive change over time and pharmacological treatment in patients with clinical depression.

**Stroke.** The MMSE is routinely used in stroke assessment. Studies show, however, that it may only be adequate in detecting cognitive impairment in left- but not right-hemispheric stroke especially if the extended version (the 3MS) is used. Grace et al also found that the test misclassifies those with aphasia, a common problem in stroke, as it is chiefly language dependent. Nys et al found that the accuracy of detecting cognitive impairment with the MMSE in acute stroke was no better than chance and could find no optimum cutoff score that they could use in this population. These authors concluded that the test was insensitive to impairments in EF, abstract reasoning, and visual perception/construction, all of which are often impaired in the early stages after a stroke.

**Multiple sclerosis (MS).** Beatty and Goodkin reviewed the use of the test in patients with MS. Their conclusion was that it is not sensitive enough to identify (diagnosed) dementia in these patients but that it can still be useful as a predictor of focal cognitive impairments in (mildly physically disabled) relapsing-remitting patients.
Table 1. The Advantages and Disadvantages of the MMSE

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Internationally used and widely translated</td>
<td>Translations not always good. Scores in other cultures may not be reliable</td>
</tr>
<tr>
<td>2. Can be administered by experts and nonexperts</td>
<td>Intercrater reliability may be unstable. Testers &amp; settings can influence scores</td>
</tr>
<tr>
<td>3. Comparison between studies possible</td>
<td>Heavy reliance on total scores while differing cutoff scores are used</td>
</tr>
<tr>
<td>4. Use of a single, total score to measure cognitive impairment</td>
<td>Item and pattern analysis might be the way forward but requires more time &amp; effort</td>
</tr>
<tr>
<td>5. Measures general cognitive ability</td>
<td>Heavy reliance on language and memory but not EF makes its use for differential diagnostics limited</td>
</tr>
<tr>
<td>6. Norms (age and education) available</td>
<td>More variables than just age and education affect scores and norms not always available in all countries</td>
</tr>
<tr>
<td>7. Total maximum score = 30 making scoring quick and easy</td>
<td>Score range limited. Floor and ceiling effects common</td>
</tr>
<tr>
<td>8. Psychometric properties of the MMSE believed to be good</td>
<td>Not all studies find stable interrater reliability</td>
</tr>
<tr>
<td>9. Quick to administer/portable</td>
<td>Not quick enough for General Practitioners (~ 10 minutes)</td>
</tr>
<tr>
<td>10. Was free and widely available</td>
<td>Copyrighted since 2001. May be too costly for many users especially in developing countries</td>
</tr>
</tbody>
</table>

General Medical Populations

Cukierman et al.\(^3\) used the MMSE to assess cognitive change in diabetic patients but found that it had limited ability to detect changes in attention and processing speed and therefore questioned how useful it was in this population. The test has also been used to assess cognitive change in a wide variety of patients with cancer, but again its use has been questioned.\(^29\)

Settings It Has Been Used in

The MMSE has been used in a wide variety of settings over the years, but whether setting influences the scores obtained has not been comprehensively analyzed, at least until recently. Mitchell\(^8\) identified 3 chief settings where the test has been used: memory clinics, mixed specialist hospitals, and nonclinic community settings. The case-finding ability of the test was found to be best in specialist settings and the least successful in memory clinics. The differences were however relatively small and could be arguably attributed to ceiling effects. Considering the fact that memory clinics may be the first “port-of-call” for older people concerned about their memory this finding deserves more research attention.

Advantages and Disadvantages of the MMSE

The MMSE has retained its popularity over the years suggesting that it has its advantages as a cognitive screening instrument. However, for every advantage a disadvantage can also be named. A list of the pros and cons of the test can be found in Table 1.

1) \textit{International use, translations of the MMSE.} While the test is internationally used, the question remains as to whether the translations are always true to the original intentions of Folstein et al.\(^1\) The literature\(^37\) suggests that some words and phrases are not easily translated (eg, “no ifs, ands, or buts”) and some of the concepts may not be relevant to people from other, non-English speaking cultures (eg, spell WORLD backward). Caution should therefore be used when interpreting MMSE scores in different countries.

2) \textit{Tester/clinician characteristics.} The test can be administered by experts and nonexperts alike and is widely taught in many higher education establishments. Many researchers believe that scoring should not fluctuate widely between testers but there are some indications\(^23\) that interrater reliability is not stable suggesting that the test may be administered and/or scored differently by different testers.

3) \textit{Comparison between studies: the use of standard cutoff scores.} One of the most often quoted advantages of the MMSE is that it makes comparison between studies possible and as such is often used in the inclusion/exclusion criteria in research. The problem is that different cutoff scores have been used and it is difficult to know which are the best to use when assigning patients into groups for comparison. A cutoff of 23 or less with 8 years or more of education has generally been taken as evidence of cognitive impairment.\(^38\) However, some authors\(^35,39\) suggest that this relatively low cutoff needs to be raised to improve the test’s sensitivity and specificity, while still others say it needs to be lowered to identify more cognitive impairment.\(^42\) Which cutoff one uses will obviously have an impact not only on participant inclusion in research but also on whether a person is labeled as having cognitive impairments or not.

4) \textit{Total scores versus item analysis.} A related question as to which cutoff is the best one is the issue of using a single, total score to measure cognitive impairment. This has its advantages in that it is a kind of barometer or objective measure of cognitive ability that can easily be compared across studies and populations. While Folstein et al.\(^1\) suggested that there were 11 categories covered in the MMSE, Tombaugh and McIntyre\(^7\) said that
the items could be grouped into 7 cognitive domains. Lopez et al. carried out a factor analysis and found 3 main factors (Orientation, Language, and Attention and Calculation), while others have suggested 5 factors. The point here is that the MMSE is not unidimensional, which a total score seems to suggest. Furthermore, Ramirez et al. found using a differential item functioning (DIF) analysis that different items performed differently for different groups tested. Lopez et al. agreed many items were classified as “too easy” in their item analysis. Focusing on how points are won and lost may therefore provide more information about a patient than their total score.

5) **General versus specific cognitive assessment.** The test was initially developed as a screen to detect general cognitive impairment and not, as it has become, a diagnostic tool, which implies assessment of specific cognitive domains. Indeed, the MMSE has been criticized for its heavy reliance on verbal tests, especially those concerned with orientation (total of 10 points) and language (total of 9 points), a problem when using the test in those with poor language skills and/or low education. Ultimately, the test may be more useful in measuring the severity of cognitive problems as it is less able to discriminate between healthy older people, those with MCI, and those suffering from early dementia.

6 & 7) **Norms for age and education and other variables that can affect scores.** MMSE total scores are sensitive to education, age, race, socioeconomic status and culture among other variables. In addition, gender may or may not influence scores. Generally, as age increases total score decreases and 12% of the variance of MMSE scores has been attributed to age and education alone. Lower scores are also found for those with lower education and lower socioeconomic status. As a result of these studies, age- and education-based norms have been proposed and indeed the test (since 1993) comes with the Crum norms attached. Jones and Gallo questioned the usefulness of these norms, however, due to the fact that the original sample sizes used to formulate the values were low—less than 100 and as small as 17. Other researchers have found that these norms may not actually be routinely used by either clinicians or researchers. Those with a high level of education tend to do well on the MMSE despite impending dementia (the cognitive reserve hypothesis may explain this phenomenon). Le Carret et al. found that those with higher education tend to have more problems with abstract thinking in the early stages of dementia whereas those with less education tend to do more poorly on memory and attention tests. As the MMSE focuses on the latter 2 cognitive domains and does not address abstract thinking, it may bias the results in favor of those with more education. Indeed, although rare, individuals can score 30/30 on the MMSE and still have been just diagnosed with early-stage dementia.

Other factors that might influence scores include vision and hearing impairment, anxiety, depression, musculoskeletal disorders, dysarthria and dysphasia and the effects of medication on scores should be considered. For example, cholinesterase inhibitors, used widely in the treatment of early AD, have been found in at least 1 study to stabilize MMSE scores (see for a review). Ceiling and floor effects are common with the test.

8) **Psychometric properties of the MMSE.** Internal consistency and test–retest reliability (0.80 to 0.95) are generally good while interrater reliability is variable. Reliability tends to be high but can drop with repeated testing and seems to depend on when the retesting is carried out (it seems to be better if the period between baseline and follow-up is short or at 1 year). The validity for a cutoff score of 24 has been quoted to be 87% for sensitivity and 82% for specificity. According to Mitchell, there have been 70 MMSE validation studies conducted to date but most are underpowered giving a misleading impression of accuracy. Possible learning/practice effects have also been found throughout the literature and indeed patients may talk to each other and discuss tests they have been given. They may also be more likely than was the case in the past to attempt to self-diagnose: there is nothing stopping individuals from self-testing online. Clinicians can change the words for the recall item, but there are no parallel forms for other components of the test. Recently, Mitchell stated that at the very least specificity, sensitivity, positive likelihood ratio (LR), and negative LR should be calculated for tests that claim to be cognitive screens. Cherbuin et al. added misclassification rate to this list and gave this to be 15% for the MMSE.

9) **Ease of administration.** A selling point for the MMSE has always been its portability: the fact that all that is needed is paper and pencil and the clinician can test a patient at their bedside. This and the fact that it is relatively quick to administer (<10 minutes) make it a favorite among many testers. General practitioners are, however, less enthusiastic. In a recent survey conducted in the United Kingdom, Milne et al. found that the average doctor’s consultation lasts just 7.5 minutes, while Deveugele et al. stated that the average length of a consultation in Europe was 10.7 minutes. The MMSE (10 minutes to complete) is therefore not as widely used among general practitioners as it is assumed to be.

10) **Availability vs copyright.** The test has been placed under copyright since 2001, a fact that few users are aware of. This entails paying a fee ($1 per test) every time it is administered in its entirety, making it too costly and prohibitive for many. Martin and O’Neill have called for the removal of the copyright fees.

**Discussion**

The MMSE remains the most used test for dementia screening and is regularly used in research as a benchmark for severity of
dementia. The question as to whether it should stay in this position is addressed in this review. Few cognitive screening instruments have had such a successful history and as widespread acceptance as the MMSE. This popularity could be due to a number of things including: its relative brevity, its portability, its focus on cognitive status as a single total score (making it easy for people to understand and compare), its possible ability to track decline over time, that you do not need to be an expert to administer it, and ultimately due to the real need for quick cognitive assessment, especially in an aging world where limiting assessment duration is often a priority. Cutoff scores are also easier to interpret than differing patterns of cognitive dysfunction which item analysis can highlight but which requires more effort/work/time to achieve, and time is something doctors and other professionals do not have. Many users of the MMSE are, however, becoming more aware of the need to pay attention to how points are lost on the test and in turn to focus less on total scores.

Accurate detection of early stage dementia is crucial as treatments begin to evolve (currently, only those able to slow down progressive disease are available). A cognitive screen that is used as a first step in dementia diagnosis clearly needs to perform well in those with early dementia. This ability to differentially diagnose, especially in the early stages of dementia, is one of the major difficulties of the MMSE and calls into question its usefulness as a diagnostic tool. Age and education are not the only variables which can affect test scores: race, socioeconomic status and (some researchers find gender) can also affect scores making translated MMSEs difficult to interpret in heterogeneous, culturally diverse populations. Despite these flaws, the test has been used to quantify the severity of cognitive impairment while it clearly suffers from ceiling effects, especially in the highly educated.

A central question addressed at the beginning of this review was: Does the MMSE fulfill its original purpose? To recap, the MMSE had 3 main original aims: to screen for cognitive impairment in elderly patients under psychiatric care; to assess the severity of cognitive impairment; and to monitor change via serial testing. The question is: has it fulfilled these aims? The short answer has to be yes, the MMSE has fulfilled its original purpose, certainly in the early years (1970s-1980s). However, as with many tools, its weaknesses have since been exposed and competitors developed. The test may still be useful in certain circumstances with the caveat that monitoring change may be more problematic than was originally thought (practice effects, patients studying the test, low interrater scores, differing medications, comorbidities, life events, settings and more have all been shown to affect follow-up scores). While the MMSE may not have completely outlived its usefulness, other short cognitive screens (for recent reviews see ) should be comprehensively and systematically assessed over a variety of populations and settings instead of simply relying on the MMSE to screen (mainly verbal aspects of) cognition.

This review has also highlighted the fact that clinicians and researchers may place too much emphasis on MMSE cutoff scores and focus not enough on which items are passed and failed. While a total score on any screen (not just the MMSE) may be taken as a first indication that something is wrong, it may hide too much about what the person can or cannot do. The focus should therefore be on what mistakes are made followed up with much more detailed assessment. The MMSE still has its uses and should be considered before adopting a new, perhaps untested, screening instrument. Above all, a diagnosis of dementia should not rely chiefly on a MMSE total score. The focus should be on the individual, their history, their strengths, and weaknesses. Only then will clinicians be able to understand and hopefully assist patients in the way these older people deserve to be treated.

Declaration of Conflicting Interests
The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding
The authors received no financial support for the research and/or authorship of this article.

References


