Alternate-Form Reliability of the Montreal Cognitive Assessment Screening Test in a Clinical Setting

Ana S. Costa\textsuperscript{a} Bruno Fimm\textsuperscript{a} Paul Friesen\textsuperscript{b} Herve Soundjock\textsuperscript{a}
Claudia Rottschy\textsuperscript{a,c,d,e} Theresa Gross\textsuperscript{b} Frank Eitner\textsuperscript{b} Arno Reich\textsuperscript{a}
Jörg B. Schulz\textsuperscript{a,c} Ziad S. Nasreddine\textsuperscript{f} Kathrin Reetz\textsuperscript{a,c,d}

\textsuperscript{a}Department of Neurology and \textsuperscript{b}Division of Nephrology and Clinical Immunology, RWTH Aachen University Hospital and \textsuperscript{c}JARA – Translational Brain Medicine, Aachen, and \textsuperscript{d}Institute of Neuroscience and Medicine, Research Center Jülich GmbH, Jülich, Germany; \textsuperscript{e}Research Imaging Institute, University of Texas Health Science Center, San Antonio, Tex., USA; \textsuperscript{f}Neuro Rive-Sud/CEDRA: Centre Diagnostique et Recherche sur la Maladie d'Alzheimer, McGill University and Sherbrooke University, Montreal, Que., Canada

Key Words
Screening test • Alzheimer's disease • MCI • Alternate-Form • Reliability

Abstract

Aims: The Montreal Cognitive Assessment (MoCA) has gained recognition for its validity in detecting cognitive impairment in several clinical populations. For serial assessments, alternate forms are needed to overcome possible practice effects. Our objective was to investigate the reliability of two German MoCA alternate forms for longitudinal assessment applications.

Methods: The original and one of two alternate forms of the MoCA were administered within a 60-min interval of a clinical interview in a counterbalanced order to 100 healthy elderly controls, 30 patients with mild cognitive impairment (MCI) and 30 patients with Alzheimer’s disease (AD). The diagnosis of the majority of patients was supported by in vivo AD pathology biomarkers.

Results: There was a strong correlation between the alternate forms and the original MoCA in all groups, but particularly in the clinical samples. Total mean scores did not differ significantly between the MoCA versions, even taking into account the presentation order. As in previous studies, age and education influenced performance in the MoCA. The same pattern of group differences (controls > MCI > AD) was observed for each of the versions.

Conclusion: All three forms can be reliably and interchangeably used in serial cognitive assessment, confirming the MoCA’s applicability in research and clinical longitudinal approaches.

Introduction

The Montreal Cognitive Assessment (MoCA) is a brief cognitive screening instrument originally developed to detect mild cognitive impairment (MCI) as a possible prodromal stage of Alzheimer’s disease (AD) [1]. It assesses several cognitive domains, including attention, executive functions, language, memory and orientation, with a maximum score of 30 points. The MoCA
has now been translated into numerous languages and validated as a reliable screening tool for cognitive impairment in several other neurodegenerative diseases such as Huntington’s disease [2] and Parkinson’s disease [3, 4], as well as in other neurological and medical conditions, for example stroke [5] and cardiovascular disease [6].

Given its good suitability and broad clinical application, the MoCA is widely used in primary and secondary care [7]. Although concerns about its lack of specificity [8, 9] and necessity of population-based norms [10] have arisen in recent years, the MoCA has also gained recognition as being superior to other commonly used screening tests [11]. For example, a recent study showed that the MoCA was superior in both sensitivity and specificity to the Mini-Mental State Examination (MMSE) in predicting cognitive impairment in a heterogeneous clinical sample that included, among others, amnestic and nonamnestic MCI, AD, Parkinson’s disease dementia and vascular dementia [12]. Other studies have also emphasized the particular suitability of the MoCA to identify cognitive impairment in clinical settings compared to nonclinical samples [13]. Nevertheless, despite the MoCA’s well-established psychometric properties, there are still limitations concerning its use in longitudinal studies or clinical follow-up, in which serial assessments are required and several confounders such as practice effects may occur. To address this issue, Phillips et al. (unpub. data) developed alternate forms for the original English and French MoCA (the French alternate versions are still undergoing validation) by replacing original items with new but similar elements. Their preliminary data on the English alternate forms revealed that the three English MoCAs had equivalent total scores and also effectively discriminated MCI patients from controls and AD patients.

The aim of the present study was to investigate the alternate-form reliability and the utility of two German alternate forms of the MoCA (MoCA 2 and MoCA 3) in a memory clinic setting. We expected that reliability analyses between the alternate forms and the original MoCA (MoCA 1) would confirm their equivalence. Furthermore, we expected that all test forms would be able to equally discriminate between the clinical (MCI and AD) and control groups and determine convergent validity with the MMSE. We assumed there would be no presentation order effects, which would support the consistency of equivalence between the MoCA forms. We predicted that demographic variables such as age and education would partially explain performance in all three forms of the MoCA.

Methods

Participants

Patients with MCI or AD were prospectively recruited in the memory clinic at the Department of Neurology, RWTH Aachen University Hospital. Diagnostic procedures were based on national guidelines [14] and included clinical history (together with participant interview/questionnaire), neurological examination, routine blood tests, clinical imaging of the brain (CT or MRI) and a comprehensive neuropsychological assessment. The neuropsychological assessment battery covered several domains and included the following: (1) the CERAD-PLUS [15] (psychomotor speed, executive functions – cognitive flexibility, semantic and phonemic word fluency –, naming, verbal and nonverbal memory and constructional praxis), (2) the Wechsler Memory Scale [16] (verbal and nonverbal memory and digit span forwards and backwards), (3) block design test (visuospatial construction) [17], (4) subtests of the Visual Object and Space Perception Battery (object and spatial perception: incomplete letters, dot counting and cubes analysis) [18] and (5) the Stroop test (inhibitory control) [19]. Data on specific cerebrospinal fluid biomarkers of AD (amyloid β, total tau, and phospho-tau, amyloid β/tau ratio) were also available for the majority of patients. According to criteria proposed by Petersen [20], we included 30 patients (50% female, 93.4% Caucasian) with a single- or multidomain amnestic MCI (mean age = 67.8 years, SD = 8.13). Based on neuropsychological assessment, all MCI patients presented with an objective memory impairment, 46% (n = 14) of whom also showed deficits in other cognitive domains. None of the patients showed significant interference in activities of daily living and, therefore, did not fulfill the criteria for a dementia diagnosis. At the time of inclusion, 19 of our MCI patients (about 60%) had biomarker data available and showed evidence of typical structural changes on MRI or cerebrospinal fluid of AD pathology, thus fulfilling the criteria for prodromal AD as proposed by Dubois et al. [21]. Patients included in the AD group (n = 30, 50% female, all Caucasian) presented with a significant episodic memory impairment, associated with changes in at least one other cognitive domain, sufficiently severe to interfere with activities of daily living [22]. At the moment of inclusion, information on MRI and cerebrospinal fluid biomarkers of Alzheimer’s pathology was available for 24 of the 30 AD patients and, in all of them, findings were supportive of AD diagnosis [21, 23]. None of the patients showed evidence of substantial concomitant cerebrovascular disease. The mean age for the AD group was 71.07 years (SD = 8.57). As in the original MoCA validation study [1], we excluded patients presenting with moderate or severe depressive symptoms, according to their score on the short version of the Geriatric Depression Scale (GDS) [24, 25].

In the control group, we included 100 healthy subjects (40% female, all Caucasian, mean age = 65.4 years, SD = 9.26) recruited as volunteers in the community and at the RWTH Aachen University Hospital. For inclusion, we considered the absence of subjective cognitive impairment, current or past history of neurological or psychiatric disease, drug or alcohol abuse, major systemic disease and medication which could sig-
nificantly impair cognition; German as a first language and an MMSE score above 25 were also necessary for inclusion [26]. The study was approved by the Ethics committee of the Medical Faculty of the RWTH Aachen University and conducted in agreement with the Declaration of Helsinki. Written informed consent was given by all participants.

**Materials and Procedure**

The two alternate forms of the German MoCA were adapted from the English versions available online (http://www.mocatest.org). Each English alternate form was translated and back-translated into German by a neurologist and a psychologist, assisted independently by an experienced native-speaking translator. The majority of the items remained identical to the English versions except for some linguistic adaptations made to the following subtest: (a) Language – Verbal Fluency: we selected the letters ‘K’ and ‘M’ as the phonemic letter fluency criteria to assure a similar letter frequency [27] in all three German versions of the MoCA and, therefore, prevent possible differences in the task difficulty level between the alternate forms.

According to the online available instructions, a neurologist, a psychologist or a trained PhD/MD student tested the participants individually. Based on previous work on the alternate-form reliability of the Dementia Rating Scale [28], one of the two alternate forms of the MoCA and the original MoCA were administrated in a counterbalanced order within a 60-min interval and by the same examiner. To minimize possible interferences of other cognitive tasks and potential differences in the administration procedure between the groups, we conducted a structured clinical interview and a neurological examination with all participants during the 60-min interval. The structured clinical interview included a full medical history, family history, demographics and the short version of the GDS [24, 25]. We predicted that the clinical interview and neurological examination would generally take less time with the healthy participants than with the patients and, therefore, the former also completed the Quality of Life Questionnaire SF-36 [29]. After the second MoCA test was administrated, all participants were assessed with the MMSE [30].

**Statistical Analysis**

We calculated the association between the uncorrected total (raw score) and domain scores of the three MoCAs as well as their association with other measures (Pearson’s correlation, r). We also controlled for presentation order effects both at the level of mean differences and correlation measures. We looked to performance in the MMSE and its relation to the MoCA regarding group differences and correlation patterns for establishing concurrent validity. We also calculated Cronbach’s α for internal consistency for each of the versions. The interpretation of the magnitude of correlation coefficients followed the guidelines proposed by Cohen [31]: small (0.10–0.29), medium (0.30–0.49), and large (>0.50).

For group difference analyses on the uncorrected total scores we used t tests, Wilcoxon signed-rank or the Kruskal–Wallis tests (Mann-Whitney U test for post hoc analysis), depending on the comparison. Age differences between groups were calculated with one-way ANOVA. Multiple regression with the uncorrected total scores was used to assess the effects of age, education and sex on the total score. These statistical analyses were performed using SPSS version 20 (IBM, 2011) with an alpha value set at 0.05 as the statistical threshold for significance.

The diagnostic accuracy of each MoCA form, using the total score corrected for education, for the prediction of MCI or AD was assessed through a receiver-operating characteristic (ROC) curve calculated with MedCalc (version 12.1.4.; MedCalc Software, Mariakerke, Belgium). The optimal cutoff value for each group was selected based on maximal sensitivity and specificity.

**Results**

Regarding our sample, as presented in Table 1, one-way ANOVA revealed a significant age difference between the groups [F(2,157) = 4.697, p < 0.05]. A Kruskal–Wallis test showed a significant effect of education [H(2) = 7.399, p < 0.01]. Planned comparisons showed that AD patients were older and less educated than MCI patients and healthy participants. MCI patients and healthy controls did not differ regarding mean age and education level. Both patient groups showed a significantly higher score on the GDS compared to healthy controls [H(2) = 18.28, p < 0.01]. Nevertheless, the mean value for the GDS in each group lies under the clinical relevant threshold [24]. The MMSE score differed significantly between the groups [H(2) = 77.228, p < 0.01]. Further analysis showed that the AD group performed worse in the MMSE than the MCI (U = 101.0, z = -5.196, p < 0.0001) and control groups (U = 77.500, z = -8.020, p < 0.0001), and that the MCI group had a lower score than healthy participants (U = 703.500, z = -4.528, p < 0.0001).

The performance of the healthy subjects on the original version of the MoCA correlated significantly with

| Table 1. Demographic and clinical characteristics of the participants |
|------------------|------------------|-----------------|------------------|---|
|                  | Controls (n = 100) | MCI (n = 30)    | AD (n = 30)      | p  |
| Age, years       | 65.4±9.26         | 67.80±8.13      | 71.07±8.57       | 0.01|
| Education, years | 11.94±2.79        | 11.47±2.85      | 10.60±2.42       | 0.02|
| Female, n        | 40 (40%)          | 15 (50%)        | 15 (50%)         | n.s.|
| GDS              | 1.90±1.89         | 3.59±2.10       | 3.39±3.04        | 0.001|
| MMSE             | 28.75±1.27        | 27.00±3.59      | 21.37±5.20       | 0.001|

All data are shown as means ± SD, except where noted. n.s. = non-significant.

Alternate-Form Reliability of the MoCA
their performance on the alternate forms [MoCA 2: \( r = 0.69, p < 0.01 \) (one-tailed); MoCA 3: \( r = 0.52, p < 0.01 \) (one-tailed)]. This association was even stronger in the patient groups [for the MCI group, MoCA 2: \( r = 0.83, p < 0.01 \) (one-tailed); MoCA 3: \( r = 0.92, p < 0.01 \) (one-tailed), and for the AD group, MoCA 2: \( r = 0.95, p < 0.01 \) (one-tailed); MoCA 3: \( r = 0.94, p < 0.01 \) (one-tailed)]. The same association pattern was found between the cognitive domains of each version of the test (see online suppl. table for correlations; for all suppl. material, see www.karger.com/doi/10.1159/000340006).

When controlling for presentation order effects, we found no significant differences between the mean raw total scores when the MoCA 1 was presented first, or after any of the alternate versions [MoCA 1: \( t(98) = 0.645, \text{NS} \); MoCA 2: \( t(50) = -1.299, \text{NS} \); MoCA 3: \( t(46) = -1.367, \text{NS} \)]. The association between the original version of the MoCA and each of the two alternate forms remained significant when controlling for the order of presentation [when MoCA 1 presented first, MoCA 2: \( r = 0.55, p < 0.01 \) (one-tailed); MoCA 3, \( r = 0.46, p < 0.01 \) (one-tailed), and when MoCA 1 presented second, MoCA 2: \( r = 0.81, p < 0.01 \) (one-tailed); MoCA 3: \( r = 0.64, p < 0.01 \) (one-tailed)].

There were no significant differences between the mean raw total scores of each of the versions. A Kruskal-Wallis test revealed that there was a significant difference between the groups regarding all three test versions [MoCA 1: \( H(2) = 85.8, p < 0.0001 \); MoCA 2: \( H(2) = 47.80, p < 0.0001 \); MoCA 3: \( H(2) = 35.85, p < 0.0001 \)]. Post hoc comparisons showed that on all MoCA versions the AD group achieved the lowest mean raw total score, followed by the MCI group, whereas healthy controls had the highest score (see online suppl. figure for total and cognitive domain scores per group and per MoCA version).

We calculated a multiple regression analysis to check the influence of age, education and sex on the score of each of the versions. The regression model included all variables and was able to explain 11, 14 and 13% of the total variance, respective of each version (MoCA 1, 2 and 3). There was a negative effect of age (MoCA 1: \( \beta = -0.159, p < 0.001 \); MoCA 2: \( \beta = -0.124, p < 0.05 \); MoCA 3: \( \beta = -0.168, p < 0.05 \)) and a positive effect of education (MoCA 1: \( \beta = 0.437, p < 0.05 \); MoCA 2: \( \beta = 0.522, p < 0.05 \); MoCA 3: \( \beta = 0.302, p < 0.05 \)). Sex did not show a significant influence on the total score of any of the forms.

There was a strong positive association between each of the MoCA versions and the MMSE (including all groups; MoCA 1: \( r = 0.86, p < 0.001 \); MoCA 2: \( r = 0.88, p < 0.001 \); MoCA 3: \( r = 0.86, p < 0.001 \)). Cronbach’s α for internal consistency between the 8 domain subscores was 0.84 for the original MoCA, 0.82 for MoCA 2, and 0.76 for MoCA 3. A ROC curve analysis was carried out for determining the diagnostic accuracy of each of the MoCA versions. The discriminative power of all forms for MCI was high, with an AUC of 0.85 (95% CI, 0.780–0.909) for MoCA 1, 0.882 (95% CI, 0.776–0.950) for MoCA 2 and 0.772 (95% CI, 0.654–0.865) for MoCA 3. The discriminative power for AD was also high in all three forms, with an AUC of 0.996 (95% CI, 0.963–1.000) for MoCA 1, 0.998 (95% CI, 0.945–1.000) for MoCA 2 and 0.974 (95% CI, 0.897–0.998) for MoCA 3. For all MoCA versions discriminating MCI from controls the optimal cut-off point was ≤26 (MoCA 1: sensitivity 93%, specificity 62%; MoCA 2: sensitivity 100%, specificity 60%; MoCA 3: sensitivity 90%, specificity 57%). Discriminating AD patients from controls was best achieved using a cutoff value of ≤22 (MoCA 1: sensitivity 97%, specificity 97%; MoCA 2: sensitivity 100%, specificity 96%; MoCA 3: sensitivity 92%, specificity 96%). Since we used an MMSE value above 25 for the inclusion of healthy participants we were not able, for comparison, to conduct the ROC curve analysis for the MMSE.

**Discussion**

The main objective of this study was to assess the alternate-form reliability of the MoCA in a memory clinical setting. Our results showed that all versions of the MoCA produced equivalent mean total scores. The correlations between the original (MoCA 1) and the two alternate forms of MoCA (MoCA 2 and 3) were positive, strong and significant, which reflects a good alternate-form reliability. The association measures between the cognitive domains of each new version and the original MoCA also showed moderate-to-good correlations. In clinical and research contexts, longitudinal assessments are crucial to verify disease progression, help the differential diagnosis and measure possible treatment effects. Nevertheless, serial assessment also presents with several challenges, especially regarding practice effects [32]. Although not eliminating all of the carryover effects of the test-retest method – e.g. performance can still improve because the subject learns how to effectively approach a task – the alternate-form method reduces important confounders related to repeated exposure to a specific task [32–35].

All correlations were higher when the patient groups were included in the analysis. Previous research with the
MoCA also acknowledged a higher reliability in clinical groups compared to a sample of healthy controls, which may be due to the greater variability in clinical populations [13]. Our current results fit with the observation that the MoCA may be especially suitable for the detection of cognitive impairment in clinical populations.

Since we found no effect of presentation order in the total mean scores and correlation patterns, we can assume that all versions of the MoCA can be used interchangeably. Although in principle the use of alternate forms already prevents substantial practice effects, our analyses show that there was no significant improvement in performance when the MoCA was administered for a second time, therefore suggesting the lack of significant practice effects.

As in previous validation studies for the original MoCA, in different languages and populations [36, 37], there was a significant effect of age and education on the total score for all forms. We could determine convergent validity through the high correlations between each of the MoCAs and the MMSE. All MoCA versions also showed a high internal consistency.

Larger samples are needed to conclusively determine the sensitivity and specificity of the alternate forms. Nevertheless, our results already suggest that all MoCA versions show a comparable discriminate power for MCI and AD, which is in line with previous studies [12, 36, 37]. However, the appropriate cutoff value seems to be population-specific [38], and therefore the discriminative validity and other psychometric measures, such as test-retest reliability, as well as the utility of population-based norms, should be further addressed in future studies using the MoCA alternate forms, particularly when studying other populations. Due to its indisputable clinical usefulness, the sensitivity of the MoCA and its alternate forms to detect changes across time should be one of the next steps in the further validation of this cognitive measure.

A limitation of the present study is the lack of a comprehensive neuropsychological assessment and information on AD-specific biomarkers of the control group, therefore not controlling for subtle cognitive impairment in an overall elderly population. Nevertheless, we only included healthy participants without subjective complaints and with an MMSE score above 25 [26]. Their mean scores in the MoCA are also comparable to results from previous studies with healthy participants, either using a cutoff score [1, 38] or normative values [10, 39]. Given our focus on the utility of the MoCA alternate forms in a clinical context, one of the strengths of this study lies in the homogeneity of the clinical groups, whose diagnosis in the majority of cases was supported by in vivo AD pathology-specific biomarkers.

In summary, our results substantiate that the alternate forms of MoCA are valid and reliable tools for the repeated cognitive screening of patients with MCI and AD, consequently providing evidence for their analogous applicability in longitudinal studies and clinical follow-up assessments.

Acknowledgments

A.S. Costa was supported by a PhD fellowship (SFRH/BD/65743/2009) from the Fundação para a Ciência e Tecnologia (Portugal), financed by the POPH-QREN Program. K. Reetz was funded by the DFG JARA BRAIN Translational Brain Research in Psychiatry and Neurology (DFG ZUK32/1).

Disclosure Statement

The authors declare no conflict of interests.

References


