International Psychogeriatrics (2011), 23:7, 1116–1123 © International Psychogeriatric Association 2011
doi:10.1017/S1041610211000615

Development of a method for quantifying cognitive ability in the elderly through adaptive test administration

Susanna Konsztowicz,[1] Haiqun Xie,[2] Johanne Higgins,[3] Nancy Mayo[4] and Lisa Koski[4,5]

[1] *Integrated Program in Neuroscience, Department of Neurology and Neurosurgery, McGill University*
[2] *Division of Experimental Medicine, Department of Medicine, McGill University*
[3] *School of Rehabilitation, Faculty of Medicine, University of Montreal, Canada*
[4] *Division of Clinical Epidemiology, Research Institute of the McGill University Health Centre*
[5] *Division of Geriatrics, Research Institute of the McGill University Health Centre*

ABSTRACT

Background: The field of geriatric medicine has identified a need for an evaluative tool that can rapidly quantify global cognitive ability and accurately monitor change over time in patients with a wide range of impairments. We hypothesized that the development of an adaptive test approach to cognitive measurement would help to meet that need. This study aimed to provide evidence for the interpretability of scores obtained from a novel, adaptive approach to cognitive assessment, called the Geriatric Rapid Adaptive Cognitive Estimate (GRACE) method.

Methods: An adaptive method for cognitive assessment was developed using data from 185 patients referred for geriatric cognitive assessment, and pilot tested in an additional 137 patients. Correlations between test scores and between rank orders of patients were computed to examine the reliability and validity of cognitive ability scores obtained by (1) administering test questions out of their usual order, (2) administering only a subset of questions, and (3) administering questions adaptively using simplified selection rules based on the most difficult question passed.

Results: Cognitive ability scores obtained with the GRACE method correlated highly with the Montreal Cognitive Assessment (MoCA) scores ($r = 0.93$) and ranked patients similarly in order of ability ($r > 0.87$). A simplified adaptive testing algorithm for pencil-and-paper assessment demonstrated moderately high correlations with scores obtained from administering the full set of MMSE and MoCA items as well as the MoCA items alone.

Conclusions: Scores from the GRACE method can be obtained easily in 5–10 minutes, reducing test burden. The resulting numeric score quantifies cognitive ability, allowing clinicians to compare patients and monitor change in global cognition over time. The adaptive nature of this method allows for evaluation of persons across a broader range of cognitive ability levels than currently available tests.

Key words: GRACE method, MoCA, MMSE, Rasch analysis, assessment, geriatric, aging

Introduction

The aging of the Canadian population and its concomitant increase in cognitive disorders, such as dementia, heightens the importance of early detection and monitoring of cognitive impairments. Currently, rapid assessments of cognitive status are conducted with screening tests, such as the Mini-Mental State Examination (MMSE; Folstein *et al.*, 1975), the Clock Drawing Test (CDT; Tuokko *et al.*, 1992), and the Montreal Cognitive Assessment (MoCA; Nasreddine *et al.*, 2005). Unfortunately, the MMSE is insensitive to the mild cognitive impairment (MCI) that may be an early marker of a dementing illness, whereas the MoCA shows a strong floor effect in moderate to severe dementia (Tombaugh and McIntyre, 1992; Nasreddine *et al.*, 2005). The problem of "one size does not fit all" has resulted in practice guidelines recommending a two-stage screening process in which a normal score on the MMSE (24–30 points) warrants subsequent administration of the MoCA (Nasreddine *et al.*, 2005; Chertkow *et al.*, 2007).

Correspondence should be addressed to: Dr. Lisa Koski, Royal Victoria Hospital, R4.74, 687 Pine Avenue West, Montreal, Quebec, H3A 1A1, Canada. Phone: +1 (514) 934-1934 ext. 35652; Fax: +1 (514) 934-1934 ext. 31734. Email: lisa.koski@mcgill.ca. Received 24 Sep 2010; revision requested 6 Dec 2010; revised version received 4 Mar 2011; accepted 6 Mar 2011. First published online 4 April 2011.

https://doi.org/10.1017/S1041610211000615 Published online by Cambridge University Press

The primary goal of administering cognitive screening tests is to contribute evidence for a diagnosis of MCI or dementia based on the summed total score reaching a certain cut-off criterion. In the clinic, it has become customary to use the change in summed score over repeated assessments as a measure of improving, stable, or declining cognitive ability; however, we know from measurement theory that counts and frequencies may not produce a unit of continuous quantity that can be used to measure a trait or an attribute (Bond and Fox, 2007). Indeed, it has been shown using Rasch analysis (Rasch, 1960) that the total score on the MMSE does not measure cognitive ability as a global quantity (Jones and Gallo, 2000; Schultz-Larsen *et al.*, 2007a; 2007b). In other words, an individual who scores 10/30 on the MMSE is not necessarily 'twice' as impaired as an individual who scores 20/30 (Smith and Smith, 2004). We have demonstrated that the total score on the MoCA test yields an acceptable measure of global cognitive ability (Koski *et al.*, 2009), whose measurement precision can be improved by including additional information obtained from MMSE screening (Koski *et al.*, 2011). Unfortunately, the difficulty of the cognitive tasks administered in the MoCA makes it inappropriate for use in more cognitively impaired patients. Thus, the field of geriatric medicine is lacking a cognitive assessment tool that can assess patients across a broad range of cognitive ability levels and accurately monitor changes in cognitive ability over time. In addition to being brief, easy to administer and acceptable to patients, such a tool must quantify the global construct of "cognitive ability" so that progression of cognitive deficits can be tracked longitudinally.

We hypothesized that the efficiency of measurement of cognitive ability in the elderly could be improved through the use of adaptive testing. In adaptive testing, performance on the current question determines which question, either easier or more difficult, should be administered next (Smith and Smith, 2004). Thus, the patient is only asked questions that lie within close range of his/her ability and not all items need to be administered. Well-designed adaptive tests provide a standardized method for "skipping" questions that are too easy or too difficult for a given respondent, thereby arriving quickly at a reliable and valid final score.

The primary objective of the current study was to contribute evidence towards the interpretability of cognitive ability estimates obtained from adaptive administration of questions from the MMSE and MoCA. First, the questions were combined onto an interval scale using Rasch analysis and an adaptive method of administering the items was developed. The novel adaptive method, which we call the Geriatric Rapid Adaptive Cognitive Estimate (GRACE), was then piloted in a sample of patients undergoing geriatric cognitive assessment, and performance of the method was evaluated. We aimed to demonstrate that administering cognitive screening tests in a brief, adaptive format provides a reliable and valid estimate of cognitive ability in a geriatric clinic population.

Methods

The present study is an analysis of existing data arising from the routine assessment of 457 patients attending a geriatric cognitive disorders clinic in Montreal. Following standard practice guidelines (Nasreddine *et al.*, 2005), patients without a prior diagnosis of dementia who presented to the clinic with memory or other cognitive complaints were screened using the MMSE and those scoring in the normal range were subsequently administered the MoCA. Items that appeared in both tests (MMSE date, day of the week, month, year, place, and city) were administered only once. The test form for piloting the adaptive GRACE was developed using data from all patients assessed with the MMSE and MoCA screening tests between February, 2006 and November, 2007 (Historic Cohort, n = 185).

In the pilot GRACE method, the Clock Drawing Test was administered and scored for performance on its three elements: drawing the circle, inserting the numbers, and setting the hands to ten minutes after 11. Based on the score pattern on the Clock test, the patient was then directed to one of four sections or sub-boxes of the test: sub-box A, B, C, or D, which contained the easiest, relatively easy, relatively hard, and hardest items from the MMSE and MoCA, respectively. Patients were administered the items in the sub-box to which they were directed. The clinician then went back and administered any remaining MMSE and MoCA items, in order to derive a total score for the MMSE and MoCA tests so that no relevant clinical information was lost. To view a sample form of the GRACE adaptive method, please refer to the supplementary material (Appendix S1) that is attached to the online version of this paper at http://www.journals.cambridge.org/ipg. The piloted GRACE form was derived from the Historic cohort while the amended form is based on the values obtained from the full sample, to ensure maximum precision of cognitive ability estimates.

Among patients attending the clinic from November 2007 to August 2009, 137 individuals were administered the MMSE and the MoCA using the GRACE pilot test method (Pilot Cohort). During the same period, 135 patients (Concurrent Cohort) were administered the MMSE and MoCA

https://doi.org/10.1017/S1041610211000615 Published online by Cambridge University Press

serially in the traditional format, either because the tests were administered by clinic personnel who had not been trained in the GRACE pilot administration format, or because they appeared on initial presentation as being unlikely to pass the MMSE. The Concurrent Cohort was included to validate the method of adaptive test scoring in a sample of patients who were independent of those used to develop the pilot test form. A potential bias existed in the Concurrent Cohort in that it included patients who appeared to be lower functioning than the Pilot Cohort patients; therefore, reliability and validity comparisons were made separately for the three cohorts.

All data were coded in a clinical database. This study was approved by the McGill University Health Center (MUHC) Research Ethics Board. The following clinical and demographic information was extracted from the database: date of birth, date of test, sex, diagnosis at time of test, years of education, language of test (English or French), and itemized test scores for the MMSE and MoCA. Age was coded as <80 and 80+. Years of education were coded as low (0–8 yrs), medium (9–12 yrs), and high (13+ yrs). Patients who took the MoCA in languages other than English or French were excluded (two from the Historic Cohort and four from the Concurrent Cohort).

Each patient had a diagnosis recorded by the geriatrician at the time of the visit. As described previously (Koski *et al.*, 2011), the diagnosis was made according to DSM-IV criteria for dementia (American Psychiatric Association, 2000) and revised criteria for MCI (Petersen, 2004) based on a clinical evaluation that included a full chart review, history, physical examination, assessment of basic and instrumental activities of daily living, blood tests, brain computerized tomography (CT scan), as well as the results of the cognitive screening tests. For the purposes of this study, diagnosis was coded as normal (cognition within normal limits), MCI without dementia, and dementia, which included Alzheimer's, vascular dementia, mixed dementia, dementia with Lewy bodies, and frontotemporal dementia. People with cognitive impairment of unknown severity and etiology were classified as unspecified; those for whom cognitive impairment was secondary to a primary psychiatric disorder were classified as psychiatric; and those with no diagnostic information regarding cognition recorded in the chart at the time of assessment were labeled as "missing data".

Data analysis

Itemized responses to the questions from the MMSE and the MoCA screening tests were entered into RUMM2020 software (RUMM Laboratory Pty Ltd, version 5.1.0) and Rasch analyses were used to position the questions along a continuous scale in hierarchical order by level of difficulty, as described previously (Koski *et al.*, 2011). Kendall's tau statistic was used to evaluate the reliability of the item difficulties across the three cohorts (PASW, version 18.0).

Rasch analysis of the full set of combined MMSE and MoCA responses was used to derive cognitive ability scores for individuals in each cohort (referred to as GRACE-full scores). These scores represent the level of difficulty at which an individual has a 50% probability of responding correctly. For simplicity, the ability scores output from Rasch analysis, which ranged from −6 to +6 logits, were linearly transformed onto a scale of 0 to 30.

The reliability of the GRACE-full scores across cohorts was evaluated using one-way analysis of variance in RUMM. Next, the GRACE-full scores were compared against the current standard test method, i.e. scores obtained for the MMSE questions alone and for the MoCA questions alone. Reliability of scores across methods was estimated using paired t-tests and Pearson's correlations, as well as Kendall's tau to estimate the correlation of rank-orders of the patients.

To validate the use of an adaptive approach to assessment, in which only a subset of questions are administered, a sparse data set was created that contained only responses to items from the sub-box of the GRACE pilot form to which Pilot Cohort patients were initially directed. The scores estimated by RUMM from the sparse data set (GRACE-sparse scores) were compared to GRACE-full scores and MoCA scores in the Pilot Cohort using paired t-tests of mean scores, Pearson's correlation, and Kendall's tau correlation of rank-order.

Finally, we evaluated two methods for adaptive administration and scoring in a pencil-and-paper format. In the Clock rule, the test is administered as described for the Pilot form and the score corresponds to the most difficult item passed in the sub-box. The Middle Item method begins with administration of the question of middle difficulty (MoCA Tapping item), followed by subsequent administration of one of five specific sub-boxes, and scoring based on the most difficult item in the sub-box. The validity of each method was evaluated by comparison with the GRACE-full ability estimates derived from RUMM and also with the MoCA total scores using paired t-tests of mean scores, Pearson correlation, and Kendall's tau correlation of rank-order. This final analysis was conducted on patients from all three cohorts (N = 455 and N = 439 for the Clock and the Middle Item rules, respectively; note

https://doi.org/10.1017/S1041610211000615 Published online by Cambridge University Press

Table 1. Demographic and clinical characteristics of patients in each cohort

	COHORT			
	ALL, N = 457 NUMBER (%)	HISTORIC, N = 185 NUMBER (%)	PILOT, N = 137 NUMBER (%)	CONCURRENT, N = 135 NUMBER (%)
Age in years				
Mean [SD]	80 [6.7]	80 [6.7]	79 [7.1]	81 [6.3]
<80	191 (41.8)	71 (38.4)	62 (45.3)	58 (43.0)
80+	266 (58.2)	114 (61.6)	75 (54.7)	77 (57.0)
Sex				
Female	258 (56.5)	101 (54.6)	77 (56.2)	80 (59.3)
Male	199 (43.5)	84 (45.4)	60 (43.58)	55 (40.7)
Diagnostic class				
Normal	52 (11.4)	19 (10.3)	15 (10.9)	18 (13.3)
MCI*	156 (34.1)	65 (35.1)	45 (32.8)	46 (34.1)
Dementia	147 (32.2)	58 (31.4)	54 (39.4)	35 (25.9)
Unspecified†	57 (12.5)	25 (13.5)	10 (7.3)	22 (16.3)
Psychiatric‡	27 (5.9)	12 (6.5)	9 (6.6)	6 (4.4)
Missing data§	18 (3.9)	6 (3.2)	4 (2.9)	8 (5.9)
Education level				
Low (0–8 yrs)	98 (24.1)	32 (17.3)	34 (24.8)	32 (23.7)
Med (9–12 yrs)	139 (30.4)	56 (30.3)	40 (29.2)	43 (31.9)
High (13+ yrs)	173 (37.9)	72 (38.9)	56 (40.9)	45 (33.3)
Missing	47 (10.3)	25 (13.5)	7 (5.1)	15 (11.1)
Test language				
English	322 (70.5)	132 (71.4)	92 (67.2)	98 (72.6)
French	135 (29.5)	53 (28.6)	45 (32.8)	37 (27.4)

*MCI = mild cognitive impairment without dementia.
†Unspecified = objective evidence of cognitive impairment without yet a definitive determination between MCI and dementia.
‡Psychiatric = cognitive impairment attributed to primary psychiatric impairment.
§Missing data = diagnostic information relevant to cognition at time of assessment not available in chart.

that patients with missing data for the first item were excluded from the analysis).

Results

Description of sample

Characteristics of individuals in the entire sample as well as by cohort are shown in Table 1. Mean age was similar across the three cohorts (~80 yrs). All three cohorts had a slightly larger proportion of females than males. As reported previously in this population (Koski *et al.*, 2011), women were older (Pearson $\chi^2 = 14.4$, df = 1, $p < 0.001$) and less educated ($\chi^2 = 23.0$, df = 3, $p < 0.001$) than men. MCI was the most common diagnostic classification across the three cohorts (32–35%), followed by dementia (26–39%), with the remaining patients having normal cognition (10–13%), cognitive impairment secondary to a psychiatric disorder (4–7%), cognitive impairment of unspecified severity and etiology (7–16%), or no mention of cognition in the diagnostic conclusions at the time of the assessment (3–6%). Across the three cohorts, 17–25% of patients had 0–8 years of formal education, 30% had 9–12 years, 33–40% of patients had more than 13 years of education, and 5–13% of cases were missing information on education. Approximately 70% of each cohort took the screening tests in English and 30% in French. While test language showed a similar distribution across sex and age, French-language test-takers tended to be more educated ($\chi^2 = 13.8$, df = 3, $p = 0.003$) than English-language test-takers.

Reliability of cognitive ability scores across time and across order of administration of test questions

Rank-order correlations of 0.80 to 0.87 (all $p < 0.01$) indicate that the relative difficulty of the test questions was stable and reliable across different groups of patients and across time. Table 2 shows the mean scores obtained from the GRACE-full method, the MoCA, and the MMSE for each of the cohorts. The mean GRACE-full score did not differ across cohorts ($F = 0.90$, $p = 0.41$), indicating that the test questions can be administered in a different order than usually presented in the MMSE and MoCA tests, without invalidating the assessment.

https://doi.org/10.1017/S1041610211000615 Published online by Cambridge University Press

Table 2. Comparison of scores across 3 Cohorts (rows) and across three test methods (columns)

COHORT	GRACE*	MoCA	MMSE
Historic			
Observed Score Range†	14.7–29.2	9–30	17–30
% at floor/ceiling	0/0	0/1	0/8
Mean (SD)	**19.4 (2.2)**	**19.7 (4.2)**	**26.2 (2.6)**
Test bias contrast, t-value (95% CI)			
GRACE-MoCA		−1.72 (−0.62–0.04)	
GRACE-MMSE			−46.47 (−7.08– −6.50)
Pilot			
Observed Score Range†	10.5–27.1	3–30	11–30
% at floor/ceiling	0/0	0/1	0/6
Mean (SD)	**19.3 (2.8)**	**19.8 (5.2)**	**25.4 (3.4)**
Test bias contrast, t-value (95% CI)			
GRACE-MoCA		−2.06 (−0.95–0.02)	
GRACE-MMSE			−37.65 (−6.49–5.84)
Concurrent			
Observed Score Range†	9.6–25.6	0–29	10—30
% at floor/ceiling	0/0	1/0	0/13
Mean (SD)	**19.0 (2.7)**	**18.4 (5.7)**	**25.9 (3.7)**
Test bias contrast, t-value (95% CI)			
GRACE-MoCA		2.07 (0.03–1.18)	
GRACE-MMSE			−38.02 (−7.18–6.47)

*GRACE-full ability estimates (in logits) obtained from RUMM were linearly transformed onto a scale of 0 to 30; †Theoretical score range for all methods was 0–30.
GRACE = Geriatric Rapid Adaptive Cognitive Estimate; MoCA = Montreal Cognitive Assessment; MMSE = Mini-Mental State Examination.

Table 3. Correlation of total scores obtained with the GRACE method vs. conventional test methods*

COHORT	GRACE-MoCA	GRACE-MMSE	MoCA-MMSE
Historic	0.94 (0.92–0.96)	0.67 (0.58–0.74)	0.52 (0.41–0.62)
Pilot	0.94 (0.91–0.95)	0.83 (0.77–0.88)	0.74 (0.65–0.81)
Concurrent	0.93 (0.90–0.95)	0.84 (0.79–0.89)	0.75 (0.66–0.81)

*Value in brackets indicates 95% confidence interval of rho.
GRACE = Geriatric Rapid Adaptive Cognitive Estimate; MoCA = Montreal Cognitive Assessment; MMSE = Mini-Mental State Examination.

GRACE-full scores fell between the scores obtained with the MoCA and the MMSE, which was expected as it was based on the combination of questions from both screening tests. The pattern of mean scores across the three test methods in the Pilot cohort was comparable to that of the other two cohorts, providing further evidence that estimates of cognitive ability based on these questions are reliable across order of administration. In all three cohorts, the MMSE yielded a narrower range of scores and a larger percentage of individuals at ceiling (6–13%) compared with the GRACE-full and the MoCA.

Table 3 presents the correlations between scores obtained from the GRACE-full method and scores obtained from the MoCA and the MMSE, for each cohort. GRACE-full and MoCA scores were highly correlated in all cohorts (0.93–0.94). Correlations between GRACE-full and MMSE scores were slightly lower, ranging from 0.67 to 0.84. Correlations of total scores on the MoCA and MMSE were lowest, ranging from 0.52 to 0.75. Correlations of the rank-order of persons within each cohort by GRACE-full and MoCA scores were all above 0.87 ($p < 0.01$), indicating that the two methods rank individuals according to ability in a similar manner. Correlations between GRACE-full and MMSE rankings were lower and ranged between 0.51 and 0.68 ($p < 0.01$). These results indicate convergent validity between GRACE-full scores and scores obtained from validated measures used in the clinic.

https://doi.org/10.1017/S1041610211000615 Published online by Cambridge University Press

Reliability of cognitive ability scores obtained from administering a reduced set of test questions

In the Pilot Cohort, the mean ability estimates derived in RUMM from responding to only a subset of items (GRACE-sparse) did not differ from the mean estimates derived from responding to all items (GRACE-full) ($t = -0.09$, 95% CI $= -0.26$–0.23), nor did they differ from mean scores on the MoCA ($t = -1.86$, 95% CI $= -0.99$–0.03). The GRACE-sparse scores correlated highly with the GRACE-full scores (0.90) and with the MoCA scores (0.84). The GRACE-sparse method also ranked people in a similar order to their ranking based on GRACE-full (0.74) and MoCA scores (0.71). Therefore, in addition to being able to administer the items out of order, not all of the items need to be administered to obtain reliable and valid scores.

Assessment of adaptive administration and scoring rules

The average number of questions administered to arrive at a final score was 8 for the Clock rule and 11 for the Middle Item rule. The mean of prospectively calculated scores using the Clock rule (mean $\pm$ SD $= 19.0 \pm 2.8$) was marginally lower than the GRACE-full ability score estimated from RUMM ($t = -2.95$, 95% CI $= -0.48$–-0.10) and the MoCA score ($t = -2.22$, 95% CI $= -0.71$–-0.04). The mean score obtained from the Middle Item rule (mean $\pm$ SD $= 19.0 \pm 3.0$) was marginally lower than the GRACE-full score ($t = -3.10$, 95% CI $= -0.61$–-0.14), but was not significantly different from the MoCA score ($t = -1.60$, 95% CI $= -0.74$–0.08). Hand-scoring based on the Clock rule yielded correlations of 0.70 and 0.71 with the scores obtained from the computer-calculated GRACE-full data and the MoCA, respectively. The Middle Item rule fared less well, with correlations of 0.61 and 0.52 with the GRACE-full and MoCA, respectively. The rank-order of persons by Clock rule score correlated at +0.52 with that of GRACE-full and at +0.53 with the MoCA. Rank-ordering of persons using the Middle Item rule yielded correlations of only 0.46 (GRACE-full) and 0.33 (MoCA).

Discussion

In this study, we developed and pilot tested an adaptive testing method for measuring global cognitive ability in a geriatric outpatient clinic. Based on questions from widely used cognitive screening tests, the GRACE method was found to produce reliable and valid scores when compared with traditional methods for brief assessment of cognition. Stronger correlations were observed between the GRACE and the traditional MoCA than between the GRACE and the MMSE, due to ceiling effects associated with the MMSE in patients with milder cognitive impairment. Unlike traditional methods, the GRACE requires administration of only eight questions on average. The specific questions asked depend on the ability level of the individual patient, allowing for evaluation of patients from a wide range of ability, while avoiding the frustration felt by both clinicians and patients at having to ask questions that are patently too easy or too difficult for a given person.

When performing cognitive screening of a newly referred patient, it may be desirable to maximize the data collected, particularly when building the case for a diagnosis based on cognitive impairment. Knowledge of which questions are failed may provide clues to deficits in specific cognitive domains, which could be used to clarify diagnosis and provide prognostic information to the clinician (Jacova *et al.*, 2007). For example, Montero-Odasso and Muir (2010) recently showed that scores from individual cognitive domains assessed by the MoCA can be used to classify subtypes of MCI. For this reason, the clinician may prefer to administer all questions at a patient's initial visit, but the use of the GRACE method at subsequent visits may be an attractive alternative to administering full screening tests when time is limited. The set of questions from which the GRACE score is derived was previously shown to provide a precise and valid estimate of global cognitive ability (Koski *et al.*, 2011). This means that the score obtained from the GRACE method could be used to track changes in cognitive ability over time.

The benefits of adaptive testing for the precise and rapid assessment of large populations have long been recognized in the fields of education and psychology, but are only now being realized in the health sciences for patient-reported outcomes as well as performance-based assessments (Cinamon *et al.*, 2010, Higgins *et al.*, 2010). In the current study, correlations with traditional screening tests were stronger when the individual GRACE scores were calculated by computer than when they were selected "by hand". This is because the cognitive ability score calculated by computer takes into account the pattern of success and failure on all questions asked, as compared with simply selecting the value for the most difficult question answered correctly. While the GRACE method is currently designed as a paper-and-pencil assessment for reasons of practicality and in order to facilitate worldwide dissemination of the adaptive method, there is no reason why it could not be modified into a computerized format for greater precision

https://doi.org/10.1017/S1041610211000615 Published online by Cambridge University Press

in estimating cognitive ability in geriatric patients. With the expanding use of laptop and hand-held computers, the growth of computer-based testing is increasing rapidly. Recently, Wouters and colleagues (2009) simulated a computer adaptive test (CAT) for cognitive assessment in dementia using 47 items from the Cambridge Cognitive Examination. With the ability to estimate and minimize standard error, CATs are very precise and can be conveniently administered.

In the present study, two different rules for administering and scoring the GRACE method adaptively were evaluated with the aim of selecting the more reliable and valid rule. The Clock rule outperformed the Middle Item rule on multiple measures. This may be explained by the fact that the Clock rule selects the sub-box to which an individual is directed based on their performance on three initial items, rather than the single initial item used in the Middle Item method. Use of the Clock test as the first item is further warranted by the fact that it is already in widespread use as a stand-alone test that is well accepted by both patients and clinicians (Shulman, 2000). We did not test other possible rules for adaptive administration in the pencil-and-paper format. A different rule might outperform the Clock rule in terms of its approximation of the computer-generated score, but would likely come at the expense of more complex methods for administration and scoring.

Some limitations of this work must be acknowledged. First, the order of difficulty of the test questions was determined based on data collected in a particular group of patients, namely those referred to a specialty clinic for evaluation of memory or other cognitive changes. Rasch analysis of the full set of questions largely supported the use of the adaptive test score as a measure of global cognitive ability; however, questions in the memory domain showed a tendency to cluster together at the difficult end of the scale. Therefore, validation of the GRACE method in a less memory-biased sample is required. A further improvement would be obtained by adding more difficult questions that would allow improved precision of measurement at the normal/high ability end of the cognitive spectrum. Second, the integration of the MoCA questions into this method was only possible by restricting the sample to those patients who performed within the high end of the MMSE score range. Thus, the method must be validated for use in patients with moderate to severe dementia. Finally, it is an unavoidable feature of the GRACE method that the word list must be administered to all patients prior to administering the Clock question, regardless of whether they will later be asked to recall the words. While onerous for those who are not required to do recall, it is a minor set-back and the GRACE method still provides a major reduction in test burden for all patients.

In conclusion, the GRACE method is a novel, adaptive method of administering common screening test questions that responds to the current needs of global cognitive assessment. It is brief, easy to administer, and reduces test-taker burden. The resulting numeric score quantifies cognitive ability, allowing clinicians to compare patients and providing the potential to monitor change in global cognition over time. This single method may provide a means for evaluating persons of a wider range of cognitive ability levels than previously possible.

Conflict of interest

None.

Description of authors' roles

SK contributed to study design and planning of statistical analyses, and was responsible for maintenance of the clinical database, data extraction, conducting all statistical analyses, writing the first and subsequent drafts of the paper, and creating the GRACE sample form. HX assisted with maintenance of the clinical database and created the test form for the pilot version of the GRACE method that was administered in the clinic. JH assisted with interpretation of Rasch analysis results. NM contributed to the statistical design of the study and assisted with statistical analyses. LK contributed the original concept for this study, assisted with study design, supervised data collection, and contributed to revisions of the paper during preparation.

Acknowledgements

The authors thank the clinicians of the McGill University Health Center Geriatric Cognitive Disorders Clinics for enthusiastically supporting this research project and for their willingness to adapt their methods of test administration in support of this work; Aisha Bassett for assistance with data entry and maintenance of the clinical database; and Susan C. Scott for providing SAS macros for running factor and parallel analyses. The MoCA test was used for the purposes of this study with the permission of its copyright holders (see Nasreddine *et al.*, 2005).

This study was funded by a Canadian Institutes of Health Research student award and by CIHR operating grant MOP-97810.

https://doi.org/10.1017/S1041610211000615 Published online by Cambridge University Press

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, text revision. Washington, DC: American Psychiatric Association.

Bond, T. G. and Fox, C. M. (2007). *Applying the Rasch model: Fundamental Measurement in the Human Sciences*, New Jersey: Lawrence Erlbaum Associates, Inc.

Chertkow, H. *et al.* (2007). Mild cognitive impairment and cognitive impairment, no dementia. Part A: Concept and diagnosis. *Alzheimers and Dementia*, 3, 266–282. doi: 10.1016/j.jalz.2007.07.013

Cinamon, J. S., Finch, L., Miller, S., Higgins, J. and Mayo, N. (2010). Preliminary evidence for the development of a stroke specific geriatric depression scale. *International Journal of Geriatric Psychiatry*, 26, 188–198. doi: 10.1002/gps.2513

Folstein, M. F., Folstein, S. E. and Mchugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

Higgins, J., Finch, L. E., Kopec, J. and Mayo, N. E. (2010). Development and initial psychometric evaluation of an item bank created to measure upper extremity function in persons with stroke. *Journal of Rehabilitation Medicine*, 42, 170–178. doi: 10.2340/16501977-0501

Jacova, C., Kertesz, A., Blair, M., Fisk, J. D. and Feldman, H. H. (2007). Neuropsychological testing and assessment for dementia. *Alzheimers and Dementia*, 3, 299–317. doi: 10.1016/j.jalz.2007.07.011

Jones, R. N. and Gallo, J. J. (2000). Dimensions of the Mini-Mental State Examination among community dwelling older adults. *Psychological Medicine*, 30, 605–618.

Koski, L., Xie, H. Q. and Finch, L. (2009). Measuring cognition in a geriatric outpatient clinic: Rasch analysis of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology*, 22, 151–160. doi: 10.1177/0891988709332944

Koski, L., Xie, H. and Konsztowicz, S. (2011). Improving precision in the quantification of cognition using the Montreal Cognitive Assessment and the Mini-Mental State Examination. *International Psychogeriatrics*. E-published ahead of print. doi: 10.1017/S1041610210002450

Montero-Odasso, M. and Muir, S. W. (2010). Simplifying detection of mild cognitive impairment subtypes. *Journal of the American Geriatrics Society*, 58, 992–994. doi: 10.1111/j.1532-5415.2010.02823.x

Nasreddine, Z. S. *et al.* (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53, 695–699.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256, 183–94. doi: 10.1111/j.1365-2796.2004.01388.x

Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*, expanded edn. Chicago: The University of Chicago Press.

Schultz-Larsen, K., Kreiner, S. and Lomholt, R. K. (2007a). Mini-Mental Status Examination: Mixed Rasch model item analysis derived two different cognitive dimensions of the MMSE. *Journal of Clinical Epidemiology*, 60, 268–279. doi: 10.1016/j.jclinepi.2006.06.007

Schultz-Larsen, K., Lomholt, R. K. and Kreiner, S. (2007b). Mini-Mental Status Examination: a short form of MMSE was as accurate as the original MMSE in predicting dementia. *Journal of Clinical Epidemiology*, 60, 260–267. doi: 10.1016/j.jclinepi.2006.06.008

Shulman, K. I. (2000). Clock-drawing: is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, 15, 548–561.

Smith, Jr., E. V. and Smith, R. M. (2004). *Introduction to Rasch Measurement*, Maple Grove: JAM Press.

Tombaugh, T. N. and McIntyre, N. J. (1992). The Mini-Mental State Examination: a comprehensive review. *Journal of the American Geriatrics Society*, 40, 922–935.

Tuokko, H., Hadjistavropoulos, T., Miller, J. A. and Beattie, B. L. (1992). The Clock Test: a sensitive measure to differentiate normal elderly from those with Alzheimer's disease. *Journal of the American Geriatrics Society*, 40, 579–584.

Wouters, H., Zwinderman, A. H., van Gool, W. A., Schmand, B. and Lindeboom, R. (2009). Adaptive cognitive testing in dementia. *International Journal of Methods in Psychiatric Research*, 18, 118–127. doi: 10.1002/Mpr.283

https://doi.org/10.1017/S1041610211000615 Published online by Cambridge University Press