Journal of the International Neuropsychological Society (2012), **18**, 314–322.

doi:10.1017/S1355617711001755

Improving the Accuracy and Precision of Cognitive Testing in Mild Dementia

Hans Wouters,[1,2] Bregje Appels,[3] Wiesje M. van der Flier,[4] Jos van Campen,[5] Martin Klein,[6] Aeilko H. Zwinderman,[1] Ben Schmand,[7,8] Willem A. van Gool,[7] Philip Scheltens,[4] AND Robert Lindeboom[1]

[1]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands
[2]Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, the Netherlands
[3]Department of Medical Psychology, Slotervaart Hospital, Amsterdam, the Netherlands
[4]Department of Neurology and Alzheimer Center, VU University Medical Center, Amsterdam, the Netherlands
[5]Department of Geriatric Medicine, Slotervaart Hospital, Amsterdam, the Netherlands
[6]Department of Medical Psychology, VU University Medical Center, Amsterdam, the Netherlands
[7]Department of Neurology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands
[8]Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

(RECEIVED April 22, 2011; FINAL REVISION November 21, 2011; ACCEPTED November 21, 2011)

Abstract

The CAMCOG, ADAS-cog, and MMSE, designed to grade global cognitive ability in dementia have inadequate precision and accuracy in distinguishing mild dementia from normal ageing. Adding neuropsychological tests to their scale might improve precision and accuracy in mild dementia. We, therefore, pooled neuropsychological test-batteries from two memory clinics (ns = 135 and 186) with CAMCOG data from a population study and 2 memory clinics (n = 829) and ADAS-cog data from 3 randomized controlled trials (n = 713) to estimate a common dimension of global cognitive ability using Rasch analysis. Item difficulties and individuals' global cognitive ability levels were estimated. Difficulties of 57 items (of 64) could be validly estimated. Neuropsychological tests were more difficult than the CAMCOG, ADAS-cog, and MMSE items. Most neuropsychological tests had difficulties in the ability range of normal ageing to mild dementia. Higher than average ability levels were more precisely measured when neuropsychological tests were added to the MMSE than when these were measured with the MMSE alone. Diagnostic accuracy in mild dementia was consistently better after adding neuropsychological tests to the MMSE. We conclude that extending dementia specific instruments with neuropsychological tests improves measurement precision and accuracy of cognitive impairment in mild dementia. (*JINS*, 2012, *18*, 314–322)

Keywords: Neuropsychology methods, Neuropsychology classification, Dementia diagnosis, Dementia epidemiology, Dementia psychology, Cognition disorders, Psychometrics methods, Models logistic

INTRODUCTION

Precise grading of the severity of global cognitive decline in mild dementia is important in view of developing new therapies for the earliest stage. However, evidence shows several instruments designed for this purpose to have inadequate precision or diminished capacity to distinguish between mild dementia and normal ageing (Vellas, Andrieu, Sampaio, Coley, & Wilcock, 2008), and consequently, also suboptimal accuracy for correct classification of mild dementia versus normal ageing.

The cognitive part of the Alzheimer's Disease Assessment Scale (ADAS-cog) (Rosen, Mohs, & Davis, 1984) designed to assess dementia severity and predominantly used in clinical trials, the Cambridge Cognitive Examination (CAMCOG) (Roth et al., 1986) designed to screen for dementia and to assess dementia severity, and the Mini Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975) were all shown to have imprecise grading of dementia severity in mild dementia (De Jager, Milwain, & Budge, 2002; Harrison et al., 2007; Tombaugh & McIntyre, 1992). This reduces the capacity to detect improvement of cognitive function in mildly demented patients following therapy, it limits the screening for and monitoring of preclinical and incipient dementia, and it weakens the relationship between early global cognitive decline and other clinical outcome

Correspondence and reprint requests to: Hans Wouters, Department of Pharmacoepidemiology and Clinical Pharmacology, Faculty of Science, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, 3584 CG, Utrecht, The Netherlands. E-mail: j.wouters@uu.nl

variables. At the same time, neuropsychological tests designed to assess memory and executive functioning were shown to better measure mild degrees of cognitive decline than the ADAS-cog (Harrison et al., 2007) and to better discriminate between incipient dementia and normal ageing than the CAMCOG (De Jager et al., 2002). Regarding the MMSE, Tombaugh and McIntyre (1992) suggested adding neuropsychological tests to enhance its precision and accuracy in mild dementia.

Using the Rasch measurement theory (Fischer & Molenaar, 1995; Verhelst & Glas, 1995), we examine whether adding neuropsychological tests to the ADAS-cog, CAMCOG, and MMSE improves the precision of grading mild degrees of global cognitive decline and the detection of mild dementia. We examine this in particular for neuropsychological tests that measure episodic or semantic memory and executive function. The Rasch measurement theory is well known in the field of educational measurement and is increasingly in use to improve the validity of patient reported outcomes (Hobart, Cano, Zajicek, & Thompson, 2007; Jenkinson, Fitzpatrick, Garratt, Peto, & Stewart-Brown, 2001).

We extend previous findings about the additional value of neuropsychological tests in three important ways. First, we compare the neuropsychological tests with the ADAS-cog, the CAMCOG, and the MMSE simultaneously rather than separately. Second, the Rasch measurement theory assumes a single underlying dimension of global cognitive ability that is common to the CAMCOG, ADAS-cog, MMSE, and neuropsychological tests. If corroborated, it permits quantitative comparisons of individuals' total scores on either of these instruments. Third, we examine how all items and tests relate to global cognitive ability in terms of difficulty. We examine if neuropsychological tests are more difficult than the other instruments and, therefore, if they are more suitable to grade mild degrees of global cognitive decline.

METHOD

Participants

We studied data from 2061 participants from 4 data sets: (I) the AMSTEL study, a population based study and two memory clinics ($n = 852$) (Frankfort et al., 2006; Jonker, Schmand, Lindeboom, Havekes, & Launer 1998; Walstra, Teunisse, Van Gool, & Van Crevel, 1997), (II) three multicenter randomized controlled trials ($n = 714$) (Raskind, Peskind, Wessel, & Yuan, 2000; Tariot et al., 2000; Wilcock, Lilienfeld, & Gaens, 2000), (III) data from the diagnostic geriatric day clinic of the Slotervaart Hospital in Amsterdam ($n = 182$) (Meulen et al., 2004) and (IV) data from the Alzheimer Center of the VU University medical center in Amsterdam ($n = 313$) (see Van der Vlies et al., 2007 for a description of part of the data and diagnostic workup). A total of 198 individuals (data set I, $n = 23$; data set II, $n = 1$; data set III, $n = 47$; data set 4, $n = 127$) responded to less than 20% of the items or tests and were excluded from further analysis, leaving data from 1863 individuals available. Dementia was diagnosed according to consensus guidelines (American Psychiatric Association, 1987; McKhann et al., 1984; McKeith et al., 1996; Neary et al., 1998; Román et al., 1993). Specific details on the diagnostic workup can be found in the original publications. All data were collected according to the principles of the Helsinki Declaration. Whenever data were collected outside the context of standard patient care, ethical approval was obtained from local medical ethical committees or institutional review boards.

Instruments

Data set I contained the CAMCOG, data set II the 13-item ADAS-cog, that is, the original 11-item ADAS-cog extended with a concentration task and a verbal delayed recall task (Mohs et al., 1997) and data sets III–IV neuropsychological tests. Data set III included the Trailmaking Test part A and B, (Reitan, 1955), the Visual Association Test (Lindeboom, Schmand, Tulner, Walstra, & Jonker, 2002), and the Stroop Color Word Test (Stroop, 1935). In addition to these tests, data set IV contained Animal Fluency (30s, 60s, 120s), Insect Fluency 60s (Lezak, 1995), the Mazes Test (Wechsler, 1991), Digit Span Forward & Backward (Wechsler, 1997), and the Digit Symbol Test (Wechsler, 1997). The MMSE was available in every data set (see Table 1). Neuropsychological tests were administered by neuropsychologists or trained assessors under supervision of a neuropsychologist.

Estimation of Item Difficulties and Individual Abilities

Two approaches from the Rasch measurement theory were used in conjunction with each other to pool the data sets and to estimate an underlying dimension of global cognitive ability common to all items or tests. A problem for pooling of data sets which have only part of their items in common are the resulting blocks of missing data where the data sets have unique rather than common items. Fortunately, as long as data sets have items in common, they can still be pooled and the whole item set can be equated and analyzed with the common items. The method is hence called "common item equating" (Kelderman, 1988). See Holman, Lindeboom, Glas, Vermeulen, & De Haan (2003) for application and discussion. In our study, the MMSE items (all four data sets), animal fluency 60s (Data sets I and IV), the Stroop Color Word test, Trailmaking Test and the Visual Association Test version A (Data set III and IV) were the common items or tests.

Subsequently, the One Parameter Logistic Model (OPLM) (Verhelst & Glas, 1995), a Rasch measurement model, was used to estimate a single difficulty level for each item on this dimension of global cognitive ability. In addition, the OPLM estimates an ability level for each individual on this same dimension. The item difficulty and individual ability estimates are maximum likelihood estimates of the log transformed odds (log p/(1-p)) of responding correctly. In the OPLM, the probability to respond correctly is modeled as a

https://doi.org/10.1017/S1355617711001755 Published online by Cambridge University Press

Table 1. Overview of MMSE, ADAS-cog, CAMCOG items and Neuropsychological tests

Item	Description	Domain	Cutoff point* (max)	χ^2 (df) goodness of fit	Instrument	Data set
		Included Items				
Q1	Serial 7 s	WM	3/4 (5) c	16.7 (16)	MMSE	1–4
Q3	No ifs and or buts	LA/AT	—/— (1) c	5.2 (5)	MMSE	1–4
Q4	Close your eyes	LA	—/— (1) c	1.5 (3)	MMSE	1–4
Q5	Three stage command	LA/AT	1/2 (3) c	9.6 (5)	MMSE	1–4
Q6	Orientation in Time (total score)	OM	3/4 (5) c	5.7 (4)	MMSE	1–4
Q7	Orientation in Place (total score)	OM	4/5 (5) c	20.6 (12)	MMSE	1–4
Q8	Copy 2 Pentagons	CP	—/— (1) c	10.8 (11)	MMSE	1–4
Q9	Write a sentence	LA	—/— (1) c	1.6 (5)	MMSE	1–4
Q10	Immediate recall of 3 unrelated words	AT	2/3 (3) c	2.4 (3)	MMSE	1–4
Q11	Delayed recall of 3 unrelated words	EM	1/2 (3) c	6.5 (6)	MMSE	1–4
Q12	Digit Cancellation (either of 2 numbers)	EF	30/31 (40) c	10.5 (7)	ADAS-cog	2
Q13	Naming: fingers and objects	LA	5/6 (17) e	2.2 (3)	ADAS-cog	2
Q14	Following Commands	LA	2/3 (5) e	3.2 (2)	ADAS-cog	2
Q15	Remembering Test Instructions	AT	3/4 (5) rs	0.4 (2)	ADAS-cog	2
Q16	Spoken language ability	LA	1/2 (5) rs	2.3 (3)	ADAS-cog	2
Q17	Word finding difficulty	LA	3/4 (5) rs	0.2 (1)	ADAS-cog	2
Q18	Language comprehension	LA	1/2 (5) rs	2.4 (4)	ADAS-cog	2
Q20	Constructional Praxis: drawing 4 figures	CP	2/3 (4) e	8.2 (5)	ADAS-cog	2
Q21	Ideational Praxis	IP	3/4 (5) e	2.8 (2)	ADAS-cog	2
Q22	Immediate Word Recall	EM	5/6 (10) e	7.1 (3)	ADAS-cog	2
Q23	CERAD Delayed Recall	EM	6/7 (10) e	0 (1)	ADAS-cog	2
Q24	Word Recognition Test	EM	6/7 (24) e	6.7 (7)	ADAS-cog	2
Q25	176–177 Calculation Money	AT	1/2 (2) c	1.7 (4)	CAMCOG	1
Q26	Animal fluency 60 seconds	EF/SM	14/15 (—) c	3.9 (6)	CAMCOG	1 & 4
Q27	Count Backwards 20–1	WM	1/2 (2) c	4.7 (5)	CAMCOG	1
Q28	Similarities: apple banana	EF/SM	1/2 (2) c	3.5 (7)	CAMCOG	1
Q31	Similarities: plant animal	EF/SM	0/1 (2) c	9.1 (7)	CAMCOG	1
Q32	130–133 Following Commands	LA/SM	3/4 (4) c	7.2 (7)	CAMCOG	1
Q33	134–136 Semantic Knowledge: Factual Comparisons	SM	2/3 (3) c	5.4 (4)	CAMCOG	1
Q34	Semantic Knowledge: Naming objects	LA	5/6 (6) c	6.4 (6)	CAMCOG	1
Q35	140–143 Semantic Knowledge: Factual Knowledge	LA/SM	4/5 (6) c	2.6 (7)	CAMCOG	1
Q36	Executing verbal Command	LA	–/– (1) c	2.4 (2)	CAMCOG	1
Q37	165–166 Copy figures: 3d house & spiral	CP	1/2 (2) c	9.7 (6)	CAMCOG	1
Q38	Clock Drawing	EF/CP	2/3 (3) c	1.3 (5)	CAMCOG	1
Q39	170–174 Ideational and Ideomotor Praxis	IP	4/5 (8) c	1.4 (2)	CAMCOG	1
Q40	Tactile Perception (coins)	SM	1/2 (2) c	13.6 (6)	CAMCOG	1
Q41	Delayed recall of six objects Q34	EM	1/2 (6) c	8 (7)	CAMCOG	1
Q42	Delayed recognition Q34	EM	3/4 (6) c	7.2 (6)	CAMCOG	1
Q43	148–153 Remote Memory	SM	4/5 (6) c	6.1 (6)	CAMCOG	1
Q44	154–157 Recent Episodic Memory	EM	3/4 (4) c	1.9 (5)	CAMCOG	1
Q46	Recognition famous persons	SM	1/2 (2) c	4 (4)	CAMCOG	1
Q48	Animal fluency 30 seconds	EF/SM	11/12 (—) ts	5.4 (3)	NPE	4
Q49	Animal fluency 120 seconds	EF/SM	24/25 (—) ts	1.2 (3)	NPE	4
Q50	Insect Fluency 60 seconds	EF/SM	5/6 (—) ts	2.5 (3)	NPE	4
Q51	WISC Mazes number of errors	EF	6/7 (—) e	2.2 (3)	NPE	4
Q52	WAIS Digit Span Forward	AT	9/10 (21) c	9.3 (3)	NPE	4
Q53	WAIS Digit Span Backward	WM	7/8 (21) c	12.8 (4)	NPE	4
Q54	Stroop word reading	AT	62/63 (—) ts	0.3 (1)	NPE	3–4
Q55	Stroop color naming	AT	85/86 (—) ts	1 (2)	NPE	3–4
Q56	Stroop color-word interference	EF	150/151 (—) ts	1.1 (3)	NPE	3–4
Q57	Trailmaking A numbers.	AT	65/66 (—) ts	0.8 (2)	NPE	3–4
Q58	Trailmaking B numbers + letters	EF	167/168 (—) ts	5.7 (2)	NPE	3–4
Q59	Visual Association Test A	EM	8/9 (12) c	4.6 (2)	NPE	3–4
Q60	Visual Association Test B	EM	8/9 (12) c	1.5 (3)	NPE	4

(Continued)

https://doi.org/10.1017/S1355617711001755 Published online by Cambridge University Press

Table 1. Continued

Item	Description	Domain	Cutoff point* (max)	χ^2 (df) goodness of fit	Instrument	Data set
Q61	WAIS Symbol Substitution	EF	33/34 (133) c	1.1 (2)	NPE	4
Q62	Wais Substitution Incidental Learning	EM	3/4 (18) c	1 (3)	NPE	4
Q63	Wais Substitution free recall	EM	4/5 (10) c	4.7 (3)	NPE	4
		Excluded Items				
Q2	Spell backwards 'world'	WM	Misfitting		MMSE	2
Q19	Orientation (Time/Place)	OM	Misfitting		ADAS-cog	2
Q29	Similarity: shirt dress	EF/SM	Misfitting		CAMCOG	1
Q30	Similarity: table chair	EF/SM	Misfitting		CAMCOG	1
Q45	178 Recall address	EM	Misfitting		CAMCOG	1
Q47	Visual perception unusual view	SM	Misfitting		CAMCOG	1
Q64	Name two objects	LA	>95% of responses in 1 category		MMSE	1–4

Notes. Domains: AT, Attention; CP, Constructional Praxis; EF, Executive Functioning; EM, Episodic Memory; IP, Ideational Praxis; LA, Language; OM, Orientation Memory; SM, Semantic Memory; WM, Working Memory. Instruments: MMSE, Mini Mental State Examination; ADAS-cog, Alzheimer Disease Assessment Scale; CAMCOG, Cambridge Cognitive Examination; NPE, Neuropsychological Examination. Item numbers in descriptions refer to original CAMCOG numbering, * As determined with the One Parameter Logistic Model (OPLM). c = correct answers: > cutoff point = 1, < cutoff point = 0, e = errors rating scales & ts = time in seconds: < cutoff point = 1, > cutoff point = 0. Data sets: 1, Amstel Study and Memory Clinics; 2, RCT data Janssen research foundation; 3, Neuropsychological test data diagnostic geriatric day clinic Slotervaarthospital; 4, Neuropsychological test data Alzheimer Center VU University Medical Center.

logistic function or s-shaped curve of the difference between the estimated item difficulty and an individual's estimated ability. By definition, the item difficulty corresponds to an ability level at which the probability of responding correctly is 50%. Someone with less ability than the item difficulty has a lower probability of a correct response; someone with more ability than the item difficulty has a higher probability of a correct response. With an increase in ability, the probability to respond correctly to any item increases, also for more difficult items. With an increase in item difficulty, the probability to respond correctly decreases for anyone even for more able individuals. With respect to these logistic s-shaped curves, a potential problem of the basic Rasch model is that the curves are assumed to have equal slopes for all of the items. However, this assumption is not realistic as some items may have better capacity to discriminate between the estimated ability levels, that is, have steeper slopes than items with less capacity, that is, those with lower slopes. The OPLM resolves this problem by extending the basic Rasch model with a slope, one for each item, to express its capacity to discriminate between estimated ability levels.

Analyses

Preparatory analyses as described elsewhere were done to facilitate the analysis (Wouters, Van Gool, Schmand, Zwinderman, & Lindeboom, 2009). One item (Q64, see Table 1) was excluded from further analyses, because 95% of the responses were correct, precluding reliable difficulty estimates. Items scoring the number of seconds or errors were recoded in the scoring direction of the CAMCOG and MMSE which score the number of correct responses. Deciles were calculated for the continuous items (e.g., those measuring time in seconds) to obtain ten response categories. Categories of items were reduced by collapsing adjacent ones with the rationale as described elsewhere (Wouters, Van Gool, Schmand, & Lindeboom, 2008). For matter of simplicity, all items that were still polytomous after these revisions were then dichotomized to get a single cutoff point (incorrect *vs*. correct) for each item (see Table 1) for which we estimated a single difficulty. The validity of the item difficulties as measures of the underlying dimension of global cognitive ability was examined.

Under the assumption that a single dimension of global cognitive ability explains the data, it is expected (a) that with an increase in ability or total score, a patient's probability to respond correctly to a cognitive task should also increase, and (b) for the common items, the chance to respond correctly for people with equal ability levels should be similar in all data sets. The second expectation reflects the absence of a phenomenon called differential item functioning which suggests that the common items are not equally difficult across data sets. Both expectations, if met, would corroborate the common dimension of global cognitive ability. Using χ^2-tests implemented in the OPLM software package (Verhelst & Glas, 1995), the fit of the individual items and overall fit to these expectations was examined. It should be noted that good fit supports the null hypothesis that the observed values are according to the expected values by the OPLM. Thus, p values $> .01$ are desirable, because they indicate acceptable fit. This is unlike the usual "aim" to detect significant p values. Good fit was defined as $p > .01$ rather than $p > .05$ because of the large number of items.

https://doi.org/10.1017/S1355617711001755 Published online by Cambridge University Press

To facilitate interpretation, the item difficulties and patient ability levels expressed in log transformed odds (the unit of measurement of the OPLM) were linearly transformed to resemble a T-score metric (*M* 50; *SD* 10). To infer which items had the best precision for which ability level, the item difficulties were plotted against the mean ability levels (+ and – 2 standard deviations) of normal ageing individuals, of patients with MCI or mild dementia (total score MMSE > 22) and of patients with more advanced dementia (total score MMSE < 22).

We examined the level of precision of the ability estimates along the dimension using the so called "information" of the OPLM ability estimates (Mungas, Reed, & Kramer, 2003). Amount of information was calculated as one divided by the error variance of each ability estimate and plotted against each ability level. As a rule, the higher the level of information, the more reliable the ability estimate. The amount of information of a certain ability level is related to classical reliability. Information levels of 10, 20, 30, and 60 would correspond to classical reliability levels of respectively 0.50, 0.75, 0.83, and 0.92 in samples with an observed variance in ability of approximately 0.20. However, the advantage of information compared to classical reliability is that it can be calculated for each ability level, which results in informative trends showing not only how reliable a test is overall, but also how reliable it is for a range of ability levels. For classical reliability this is not possible, since classical reliability reflects the proportion of true variance of the total observed variance of ability. Also, as classical reliability is based on variance, it is variance dependent. Consequently, a test that was found to have adequate reliability in a particular sample can have quite disappointing reliability in samples with lower variance in ability. We compared the information of the ability estimates based on the neuropsychological tests and MMSE items (data sets III-IV) with that of the ability estimates based on the CAMCOG and MMSE (data set I), the ADAS-cog and MMSE (data set II) (all items were dichotomized, see above) and also with that of the original polytomous MMSE (all data sets), CAMCOG (data set I), and ADAS-cog (data set II).

Diagnostic accuracy of the ability estimates was examined in two series of ROC curve analyses. First, the diagnosis of dementia was taken as the reference standard. Second, the diagnosis of either MCI or mild dementia, defined according to MMSE cutoff points ranging from 20 to 25, was taken as the reference standard. Thus, the reference standard was actually a series of composite diagnoses. The *p* values were obtained by bootstrapping, that is, calculating Areas Under the Curve (AUCs) for 1000 samples drawn with replacement from the actual sample. Accuracy could not be examined for ability estimates based on the MMSE and the ADAS-cog items from data set II because that data set had only AD patients.

RESULTS

Demographic and Clinical Characteristics

Of the 1863 individuals included in the analysis, 585 (31.4%) were normal ageing individuals or had subjective complaints only, 66 (3.5%) were diagnosed with MCI, 1012 (54.3%) with AD, 133 (7.1%) with another form of dementia, and 67 (3.6%) had an unknown or psychiatric diagnosis. Individuals from data set IV were higher educated, younger and less cognitively impaired (see Table 2).

Validity Analyses

The validity of all item difficulties was examined (except Q64, see Methods). Several of the common items had biased difficulties, that is, systematic differences in difficulty across the data sets and were removed from the data set where they had bias (data set I: Q3, Q6, Q7, data set II: Q3, Q4, Q5, Q6, Q8, Q9, and Q10, data set III: Q9, Q11, Q55, and Q59 and data set IV: Q1, Q11, Q26, and Q54). Hence they were abandoned as common items and not used to equate the four data sets. However, they were retained in the remaining data

Table 2. Demographic and clinical characteristics of the participants from the four data sets

	Data sets				
	I AMSTEL study & Memory Clinics	II Janssen Trials	III Geriatric Day Clinic Slotervaart Hospital	IVAlzheimer Center VU Medical Center	Total
n (%)	829 (44.5)	713 (38.3)	135 (7.2)	186 (10.0)	1863 (100)
n Women (%)	529 (63.8)	440 (61.7)	80 (59.3)	87 (46.8)	1136 (61)
M Education (*SD*)*	3.2 (1.5)	*Unavailable*	3.7 (1.6)	5.1 (1.4)	3.6 (1.6)
Diagnosis					
n Normal ageing (%)	492 (59.3)	—	7 (5.2)	86 (46.2)	585 (31.4)
n MCI (%)	—	—	14 (10.4)	52 (28.0)	66 (3.5)
n AD (%)	226 (27.3)	713 (100)	25 (18.5)	48 (25.8)	1012 (54.3)
n Other Dementia (%)	76 (9.2)	—	57 (42.2)	—	133 (7.1)
n Unknown/other (%)	35 (4.2)	—	32 (23.7)	—	67 (3.6)
M Age (*SD*)	76.44 (6.2)	75.21 (8.1)	79.03 (6.9)	65.16 (10.3)	74.96 (8.3)
M MMSE (*SD*)	22.68 (5.6)	17.03 (3.8)	22.59 (4.2)	27.16 (2.4)	21.0 (5.7)

Notes. * Education measured on a 7-point scale, MCI = mild cognitive impairment; AD = Alzheimer's disease; MMSE = Mini Mental State Examination.

https://doi.org/10.1017/S1355617711001755 Published online by Cambridge University Press

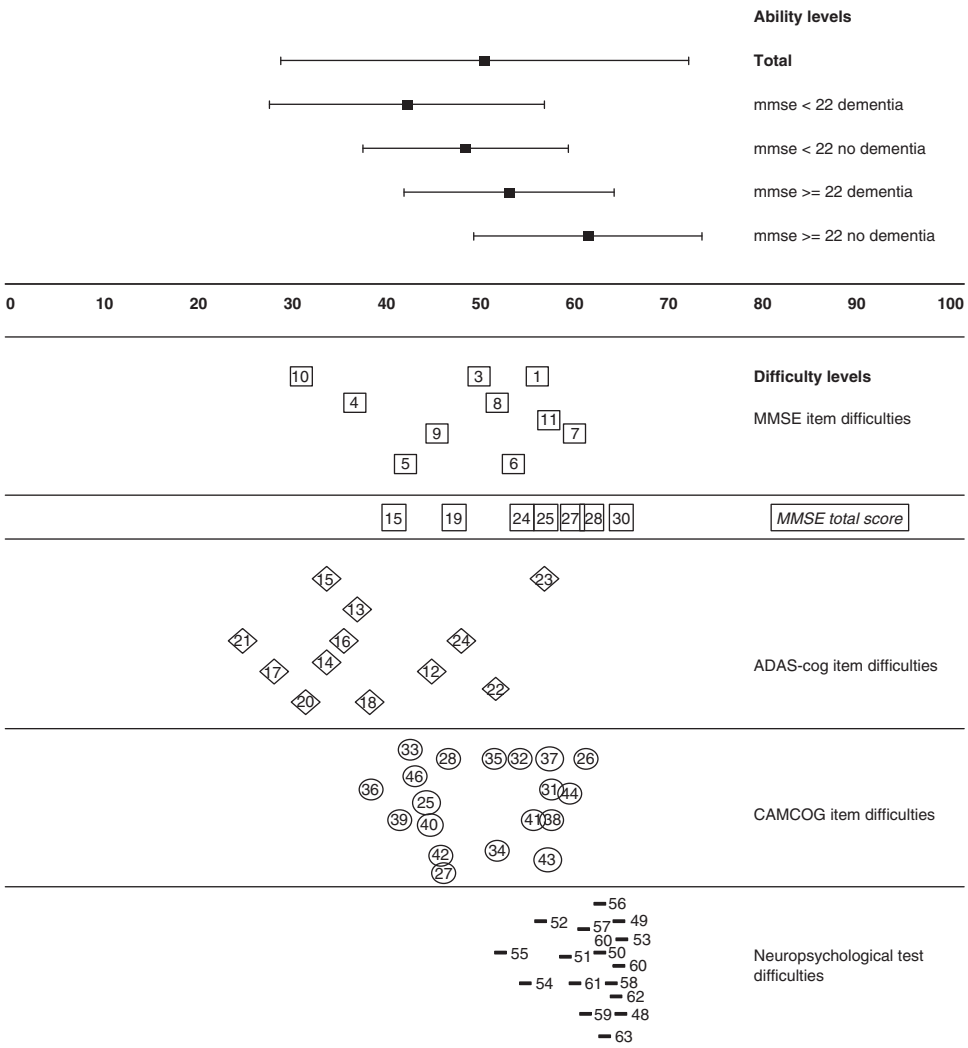

Fig. 1. Items in terms of their difficulty and persons in terms of their ability level arranged on the dimension of global cognitive ability (X-axis). Upper panel: Mean person ability levels (+ and − 2 standard deviations) for demented individuals (MMSE <22; $n = 861$; MMSE >22; $n = 284$) and non-demented individuals (MMSE <22; $n = 64$; >22, $n = 594$) in relation to the estimate of overall cognitive ability for the entire population ($n = 1803$) (on X-axis, mean: 50, *SD* 10, by definition). Lower panel: MMSE items (squares) and MMSE cutoff points, ADAS-cog items (diamonds), CAMCOG items (circles) and neuropsychological tests (stripes) arranged on the dimension of global cognitive ability. Item numbering corresponds with that of Table 1. Correlation ability estimates (0–100) and MMSE total score = 0.86.

set(s) to keep as many items as possible in the overall data set and because their difficulty could be validly estimated. For six items, difficulty levels could not be validly estimated (Q2, Q19, Q29, Q30, Q45, and Q47); these items were excluded from the whole data set (see Table 1).

Item Difficulties *versus* Patient Abilities

The neuropsychological tests had higher difficulty levels ($M = 59.62$; $SD = 3.68$) than the MMSE ($M = 48.06$; $SD = 8.83$), the CAMCOG ($M = 49.92$; $SD = 6.64$), and the ADAS-cog items ($M = 38.93$; $SD = 9.10$). Along the dimension of global cognitive ability, the most difficult neuropsychological tests measured memory (e.g., Q59-60, Q62-63) and executive functioning (e.g., Q56-Q58). A graphical inspection of Figure 1, showed the difficulties of the neuropsychological tests to be in the ability range of normal ageing individuals and patients with MCI or mild dementia.

Precision of the Ability Estimates

Plotting the ability estimates against their amount of information revealed the ability estimates based on the dichotomized MMSE items and either the dichotomized CAMCOG, ADAS-cog and neuropsychological tests to have consistently more information than the original MMSE alone (see Figure 2a–d). The abilities based on the dichotomized neuropsychological tests and MMSE items (Figure 2cd) also had more information in the ability range above average than the combinations of CAMCOG or ADAS-cog plus MMSE items (both dichotomous), and the polytomous ADAS-cog and CAMCOG (Figure 2ab).

https://doi.org/10.1017/S1355617711001755 Published online by Cambridge University Press

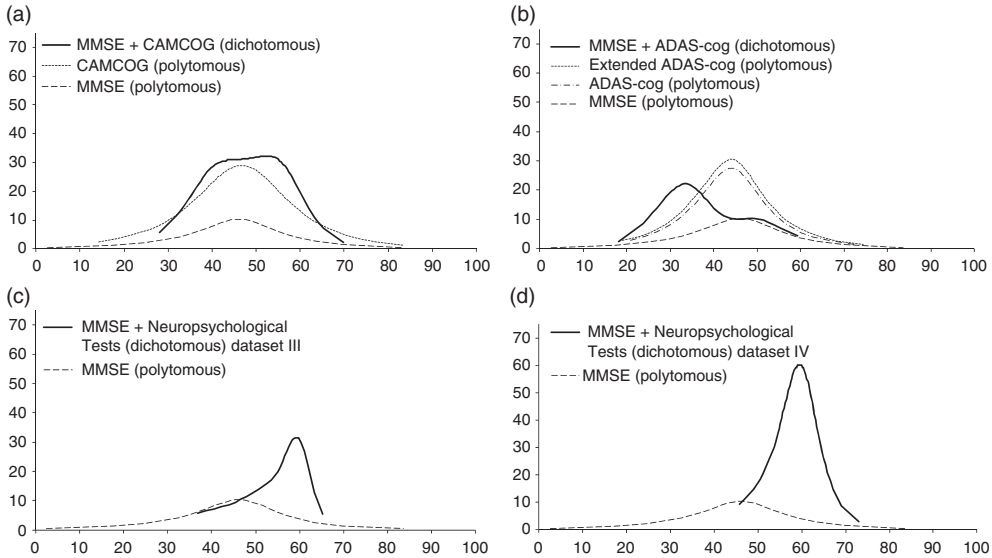

Fig. 2. Information, that is, $1/(SE^2 ability)$, (Y-axis, see Methods) along the range of global cognitive ability (X-axis, analogous to Figure 1) of the dichotomized MMSE plus either the dichotomized CAMCOG (a), the ADAS-cog (b) or the two data sets with neuropsychological tests (c–d) and of the MMSE, CAMCOG, the extended and original ADAS-cog in their polytomous form.

Accuracy

For the diagnosis of dementia, including all patients at all levels of global cognitive ability, the OPLM ability estimates did not have better accuracy than the MMSE alone. Neither when they were based on the CAMCOG and MMSE items (both AUCs 0.93, bootstrapped p value $>.05$) nor when they were based on the neuropsychological tests plus the MMSE items (AUCs 0.91 *vs.* 0.88, bootstrapped p value $>.05$).

However, if the analysis was restricted to the accuracy of diagnosing mild dementia according to the MMSE cutoff points 20–25, the OPLM ability estimates detected dementia better when based on the combined dichotomized CAMCOG and MMSE items than when based on the polytomous MMSE for the cutoff point of 25 (bootstrapped p values $<.05$). AUCs of ability estimates based on CAMCOG and MMSE items ranged from 0.88 to 0.79, AUCs of the original MMSE ranged from 0.86 to 0.64. When the ability estimates were based on the dichotomized neuropsychological tests and the MMSE, they detected dementia consistently better than the polytomous MMSE for the cutoff points 21–25 (all bootstrapped p values $<.05$). AUCs of the neuropsychological tests and MMSE ranged from 0.86 to 0.80, while AUCs of the MMSE alone ranged from 0.81 to 0.66 (see Figure 3).

DISCUSSION

We found different rates of measurement precision for the MMSE, the CAMCOG, and the ADAS-cog and specific neuropsychological tests (see Figure 2). Adding the CAMCOG to the MMSE improved the precision in the below average to average range of estimated global cognitive ability (Figure 2a; T-score range of approximately 35 to 60), whereas the precision in the below average range of estimated global cognitive ability was improved when the ADAS-cog was added to the MMSE (Figure 2b; T-score range of approximately 30 to 40). When we added the neuropsychological tests to the MMSE, the precision at levels of estimated global cognitive ability above average was improved (Figure 2c and 2d; T-score ranges of approximately 50 to 60). Moreover, the accuracy of detecting mild dementia improved consistently after adding the neuropsychological tests to the MMSE (Figure 3 lower panel) and also after adding the CAMCOG (Figure 3 upper panel). More neuropsychological tests than CAMCOG, MMSE, and ADAS-cog items had estimated difficulty levels in the range of global cognitive ability seen in normal ageing to mild cognitive impairment and in mild dementia (Figure 1, T-score range of 50–65).

These findings are important given that the usefulness of cognitive tests designed to grade the severity of dementia was questioned for early dementia (Vellas et al., 2008). The accompanying appeal was to improve their measurement precision to increase their usefulness for diagnosis and evaluation of therapy effects in patients with early dementia.

Our results are consistent with other findings (Harrison et al., 2007; De Jager et al., 2002; Tombaugh & McIntyre, 1992), yet they also extend current knowledge in two ways. First, we examined these three widely used instruments together rather than separately, as was done previously, using a common dimension of global cognitive ability. The advantage of an overall dimension of global cognitive ability is obvious: it can be measured simply with a total score. This is practical for detecting dementia, monitoring patients, and evaluating effects of medication. Our results suggest to use the most difficult CAMCOG items and neuropsychological tests for the screening and grading of early and mild dementia, and to administer the easier CAMCOG and MMSE

https://doi.org/10.1017/S1355617711001755 Published online by Cambridge University Press

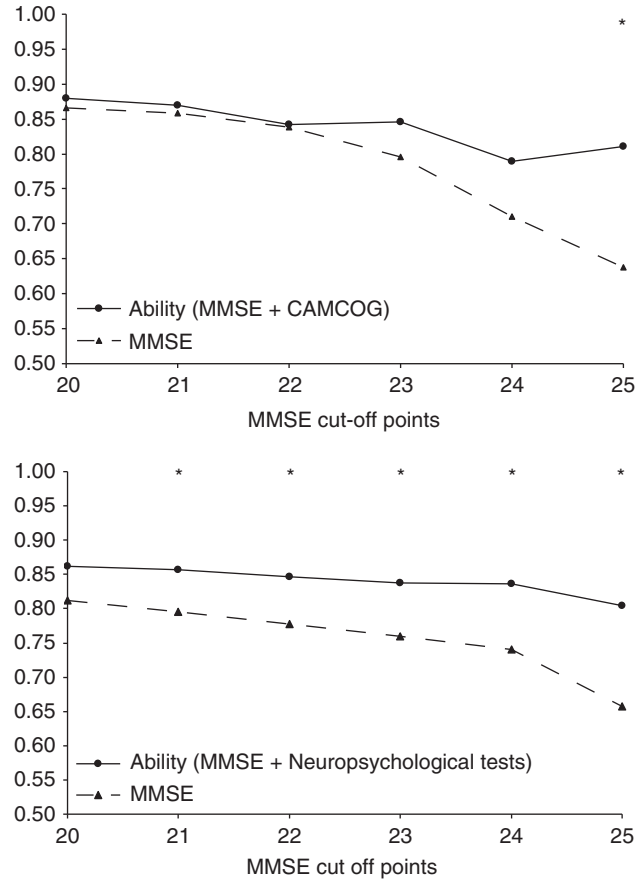

Fig. 3. Diagnostic accuracy of the ability estimates expressed as Areas under the Curve obtained with ROC analysis (Y-axis) and the original MMSE-score (X-axis) for the diagnosis of MCI and mild dementia according to the MMSE cutoff points of 20–25. Upper panel: circles represent ability estimates based on the MMSE plus the CAMCOG, triangles represent MMSE. Lower panel: circles represent ability estimates based on the MMSE plus the neuropsychological tests, triangles represent MMSE. Asterisk (*) above MMSE cutoff points indicate a significant difference between AUCs.

items and most ADAS-cog tests in patients with moderate to advanced dementia. Second, our results do not suggest replacement of the MMSE, the ADAS-cog, or the CAMCOG with more complex neuropsychological tests. After all, adding the neuropsychological tests did not improve accuracy and precision in already progressed dementia. Based on our results, a practical approach to obtain a more precise estimate of early decline of overall cognitive ability than obtained with, for example, the MMSE alone, could be to first administer the MMSE as the set of common items to every patient. Subsequently, only the highest scoring patients with, say, an MMSE score of 25–30, could be administered neuropsychological tests of executive functioning, language, and episodic and working memory, that would likely give additional information. On the other hand, lower scoring patients with, say, an MMSE score below 20, could be administered the easier items from the ADAS-cog or CAMCOG. Using the common item equating approach as described here, it would still be possible to compare patients' estimates of global cognitive ability despite the fact that for some patients the MMSE was complemented with neuropsychological tests, whereas for others it was complemented with the ADAS-cog or CAMCOG.

Our findings are limited in some respects. First, in the memory clinics diagnosis was made using the cognitive test results. Thus, the clinical diagnosis of dementia as the reference standard and the cognitive tests as the index tests were not totally independent. However, dementia assessment encompasses much more than cognitive testing alone. It also includes patient history taking, informant interview, assessment of activities of daily living, neurological and psychiatric examination, laboratory testing, and neuroimaging. Second, other tests than the ones studied here may have better psychometric properties and accuracy in detecting early dementia. Third, item difficulties may have been influenced in part by the dichotomous recoding based on the OPLM analyses. This affected the precision of the ADAS-cog compared to the polytomous ADAS-cog (see Figure 2). However, even in its polytomous form the ADAS-cog had less precision in mild dementia than the neuropsychological tests. Finally, the relatively small number of patients with other forms of dementia than AD is a potential limitation for the external validity of the results.

Taken together, our findings demonstrate that grading precision of global cognitive ability and diagnostic accuracy for early and mild dementia can be improved by adding neuropsychological tests. At the same time, administration time can be kept short using subtests with difficulty levels that are tailored to the patient's global cognitive ability.

ACKNOWLEDGMENTS

Part of the findings were presented as a poster at the 38th annual meeting of the International Neuropsychological Society, February 3–6, 2010, Acapulco de Juárez, México. This study was funded by a grant from the Internationale Stichting Alzheimer Onderzoek (ISAO). We thank Prof. Jonker and associates and the Janssen research foundation for access to the data. We also thank the Dutch Council for Educational Research and Prof. Verhelst for making the OPLM software available to us. The authors report no conflicts of interest.

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders, revised third edition*. Washington, DC: Author.

De Jager, C.A., Milwain, E., & Budge, M. (2002). Early detection of isolated memory deficits in the elderly: The need for more sensitive neuropsychological tests. *Psychological Medicine*, *32*, 483–491.

Fischer, G.H., & Molenaar, I.W. (1995). *Rasch models: Foundations, recent developments and applications*. New York: Springer-Verlag.

Folstein, M.F., Folstein, S.E., & 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.

Frankfort, S.V., Appels, B.A., De Boer, A., Tulner, L.R., Van Campen, J.P., Koks, C.H., & Beijnen, J.H. (2006). Treatment effects of rivastigmine on cognition, performance of daily living activities and behaviour in Alzheimer's disease in an outpatient

https://doi.org/10.1017/S1355617711001755 Published online by Cambridge University Press

geriatric setting. *International Journal of Clinical Practice*, *60*, 646–654.

Harrison, J., Minassian, S.L., Jenkins, L., Black, R.S., Koller, M., & Grundman, M. (2007). A neuropsychological test battery for use in Alzheimer disease clinical trials. *Archives of Neurology*, *64*, 1323–1329.

Hobart, J.C., Cano, S.J., Zajicek, J.P., & Thompson, A.J. (2007). Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *Lancet Neurology*, *6*, 1094–1105.

Holman, R., Lindeboom, R., Glas, C.A.W., Vermeulen, M., & De Haan, R.J. (2003). Constructing an item bank using Item Response Theory: The AMC linear disability score project. *Health Services and Outcomes Research Methodology*, *4*, 19–33.

Jenkinson, C., Fitzpatrick, R., Garratt, A., Peto, V., & Stewart-Brown, S. (2001). Can item response theory reduce patient burden when measuring health status in neurological disorders? Results from Rasch analysis of the SF-36 physical functioning scale (PF-10). *Journal of Neurology, Neurosurgery, and Psychiatry*, *71*, 220–224.

Jonker, C., Schmand, B., Lindeboom, J., Havekes, L.M., & Launer, L.J. (1998). Association between apolipoprotein E epsilon4 and the rate of cognitive decline in community-dwelling elderly individuals with and without dementia. *Archives of Neurology*, *55*, 1065–1069.

Kelderman, H. (1988). Common item equating: Using the loglinear Rasch model. *Journal of Educational Statistics*, *13*, 319–336.

Lezak, M.D. (1995). *Neuropsychological assessment*. New York: Oxford University Press.

Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery, and Psychiatry*, *73*, 126–133.

McKeith, I.G., Galasko, D., Kosaka, K., Perry, E.K., Dickson, D.W., Hansen, L.A., … Perry, R.H. (1996). Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): Report of the consortium on DLB international workshop. *Neurology*, *47*, 1113–1124 .

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

Meulen, E.F., Schmand, B., Van Campen, J.P., De Koning, S.J., Ponds, R.W., Scheltens, P., & Verhey, F.R. (2004). The seven minute screen: A neurocognitive screening test highly sensitive to various types of dementia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *75*, 700–705.

Mohs, R.C., Knopman, D., Petersen, R.C., Ferris, S.H., Ernesto, C., Grundman, M., … Thal, L.J. (1997). Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer's Disease and Associated Disorders*, *11*(Suppl. 2), S13–S21 .

Mungas, D., Reed, B.R., & Kramer, J.H. (2003). Psychometrically matched measures of global cognition, memory, and executive function for assessment of cognitive decline in older persons. *Neuropsychology*, *17*, 380–392.

Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., … Benson, D.F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51*, 1546–1554 .

Raskind, M.A., Peskind, E.R., Wessel, T., & Yuan, W. (2000). Galantamine in AD: A 6-month randomized, placebo-controlled trial with a 6-month extension. The Galantamine USA-1 Study Group. *Neurology*, *54*, 2261–2268.

Reitan, R.M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology*, *19*, 393–394.

Román, G.C., Tatemichi, T.K., Erkinjuntti, T., Cummings, J.L., Masdeu, J.C., Garcia, J.H., … Hofman, A. (1993). Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*, *43*, 250–260 .

Rosen, W.G., Mohs, R.C., & Davis, K.L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, *141*, 1356–1364.

Roth, M., Tym, E., Mountjoy, C.Q., Huppert, F.A., Hendrie, H., Verma, S., & Goddard, R. (1986). CAMDEX. A standardized instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *British Journal of Psychiatry*, *149*, 698–709.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Tariot, P.N., Solomon, P.R., Morris, J.C., Kershaw, P., Lilienfeld, S., & Ding, C. (2000). A 5-month, randomized, placebo-controlled trial of galantamine in AD. The Galantamine USA-10 Study Group. *Neurology*, *54*, 2269–2276.

Tombaugh, T.N., & McIntyre, N.J. (1992). The mini-mental state examination: A comprehensive review. *Journal of the American Geriatrics Society*, *40*, 922–935.

Van der Vlies, A.E., Pijnenburg, Y.A., Koene, T., Klein, M., Kok, A., Scheltens, P., & van der Flier, W.M. (2007). Cognitive impairment in Alzheimer's disease is modified by APOE genotype. *Dementia and Geriatric Cognitive Disorders*, *24*, 98–103.

Vellas, B., Andrieu, S., Sampaio, C., Coley, N., & Wilcock, G. (2008). Endpoints for trials in Alzheimer's disease: A European task force consensus. *Lancet Neurology*, *7*, 436–450.

Verhelst, N.D., & Glas, C.A.W. (1995). The one parameter logistic model. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models, foundations, recent developments and applications* (pp. 215–237) New York: Springer Verlag.

Walstra, G.J., Teunisse, S., Van Gool, W.A., & Van Crevel, H. (1997). Reversible dementia in elderly patients referred to a memory clinic. *Journal of Neurology*, *244*, 17–22.

Wechsler, D. (1991). *Wechsler intelligence scale for children — third edition (WISC - III)*. San Antonio: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler adult intelligence scale - III*. San Antonio: The Psychological Corporation.

Wilcock, G.K., Lilienfeld, S., & Gaens, E. (2000). Efficacy and safety of galantamine in patients with mild to moderate Alzheimer's disease: Multicentre randomized controlled trial. Galantamine International-1 Study Group. *British Medical Journal*, *321*, 1–7.

Wouters, H., Van Gool, W.A., Schmand, B., & Lindeboom, R. (2008). Revising the ADAS-cog for a More Accurate Assessment of Cognitive Impairment. *Alzheimer's Disease and Associated Disorders*, *22*, 236–244.

Wouters, H., Van Gool, W.A., Schmand, B., Zwinderman, A.H., & Lindeboom, R. (2009). Three sides of the same coin: Measuring global cognitive impairment with the MMSE, ADAS-cog and CAMCOG. *International Journal of Geriatric Psychiatry*, *25*, 770–779.