

Supplementary Data

Does Adaptive Cognitive Testing Combine Efficiency with Precision? Prospective Findings

Hans Wouters^{a,g,*}, Jos van Campen^f, Bregje Appels^e, Robert Lindeboom^a, Maarten Buijter^a, Rob J. de Haan^c, Aeilko H. Zwinderman^a, Willem A. van Gool^b and Ben Schmand^{b,d}

^a*Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, Amsterdam, the Netherlands*

^b*Department of Neurology Academic Medical Centre, Amsterdam, the Netherlands*

^c*Clinical Research Unit, Academic Medical Centre, Amsterdam, the Netherlands*

^d*Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands*

^e*Department of Medical Psychology Slotervaart Hospital, Amsterdam, the Netherlands*

^f*Department of Geriatric Medicine, Slotervaart Hospital, Amsterdam, the Netherlands*

^g*Department of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, the Netherlands*

Accepted 1 March 2011

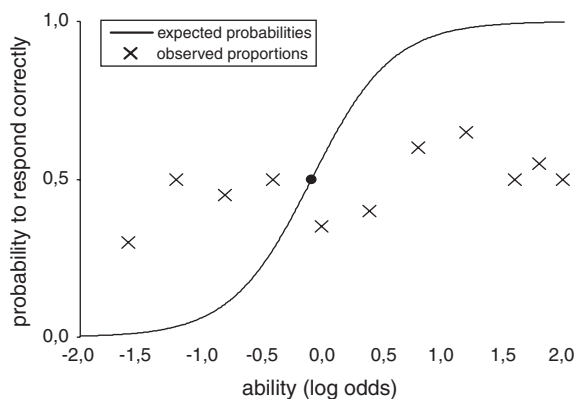
ESTIMATING AND EXAMINING VALIDITY OF ITEM DIFFICULTIES

Item difficulties of the CAMCOG and CAMCOG-Plus items were previously estimated with the one parameter logistic model (OPLM), a Rasch type of model [1, 2] using existing data. Item difficulties and ability levels are estimated on a common scale of global cognitive ability expressed in log odds that typically range from -2 to $+2$. The log odds have a one to one correspondence to the total score of the test(s) under study. In addition, the validity of item difficulties as measures of the global cognitive ability scale can be examined with the OPLM.

Examining the validity of the item difficulties is based on the following rationale. In the OPLM, the probability of responding correctly to an item is a function of an individual's estimated ability level relative to the estimated difficulty level of the item. The estimated item difficulty corresponds with an ability level which yields a 50% chance of responding correctly. An individual who has an ability that is exactly equal to the item difficulty level has a 50% probability of responding correctly. Individuals with lower ability have a lower probability to respond correctly, individuals with higher ability have a higher probability. Someone with low ability will have a low probability to respond correctly to most items, whereas a difficult item may have a low probability of responding correctly even for individuals with high ability levels. For different cumulative ability estimates the probability to respond correctly can be calculated. These are the *expected* probabilities to respond correctly. If the item

*Correspondence to: Hans Wouters, Department of Pharmacoepidemiology and Pharmacotherapy, Faculty of Science, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, 3584 CG Utrecht, The Netherlands. Tel.: +31 (0)6 22 73 60 73; E-mail: J.Wouters@uu.nl.

is a measure of an underlying dimension of global cognitive ability that is also assumed to be measured by the other items of the test, an increase of responding correctly would be expected with an increase of ability. Using χ^2 statistics the *observed* proportions of people who respond correctly can be compared with the expected probabilities. P -values < 0.01 indicate significant deviance of the *observed* proportions from the *expected* probabilities indicating that the item is not measuring the underlying dimension of global cognitive ability. The figure below illustrates this idea. It shows the probability of drawing a spiral correctly. The item difficulty is about zero ability, which indicates that copying a spiral is intermediately difficult. Memory items would probably be more difficult (located at ability levels > 0), whereas complying to a verbal command is probably easier (located at ability levels < 0). The solid line represents the probability of responding correctly that is assumed to increase with an increase of ability. The crosses represent the observed proportions of people from cumulative ability classes who do this item correctly. These observed proportions do not suggest that the probability to respond correctly increases with an increase in ability level. Apparently the item is not measuring the underlying dimension of global cognitive ability. After all, the dimension of global cognitive ability would predict an increase of responding correctly with an increase in ability. Calculating a χ^2 statistic of these differences is likely to result in significant deviances of the observed proportions from the expected probabilities. In such case, the easiest solution is to exclude the item from the item set or test. If possible, it would be worthwhile to study why the item does not measure the underlying the dimension. The item could be more difficult for older individuals than younger individuals who have the same ability.



Supplementary Figure 1. Item copy spiral (constructional praxis).

THE CAMCOG ITEM SET

The CAMCOG item set contains 47 items from the original 60 item CAMCOG [3] that includes the MMSE [4]. Items came from the CAMCOG subtests abstract thinking (1 item), attention, concentration and calculation (3 items), comprehension (8 items), expression (4 items), incidental learning (3 items), orientation (9 items), perception (3 items) praxis (7 items), recent memory (4 items) and remote memory (5 items). For 13 items, item difficulty levels could not be validly estimated. OPLM results for the CAMCOG are described elsewhere [5].

THE CAMCOG-PLUS ITEM SET

The CAMCOG-Plus has 51 items or tests out of a total of 64 items or tests that were initially considered for inclusion in the CAMCOG-Plus. Items came from the CAMCOG, MMSE ADAS-cog [6] or were neuropsychological tests (see Supplementary Table 1).

The neuropsychological tests of the CAMCOG-Plus included the digit span forward [7] and the trail making test part A [8] to assess attention, the digit span backward [7] to assess working memory, the visual association test [9] to assess episodic memory, and the trail making test part B [8], the Stroop color word test [10], insect fluency 60 s [11], the digit symbol test [7] and the mazes test [12] to assess executive function.

The remaining items or tests had to be excluded for various reasons. First of all, there were a number of duplicate tests from the itembank that had to be excluded. Neuropsychological Tests: animal fluency 30 and 120 s; ADAS-cog: temporal and spatial orientation, delayed recall; CAMCOG: ideational praxis, verbal commands, and recall of the address. Second, the incidental learning and delayed recall conditions of the digit symbol test were also not included, because it is difficult to administer these tests adaptively with a computerized adaptive testing algorithm. The reason is that the adaptive testing algorithm can select these items before presenting the digit symbol test. Third, two items, namely tactile perception and gnosis (objects from an unusual angle from the CAMCOG) were also excluded because they were found to be unreliable. Fourth, the naming item of the MMSE had more than 95% of its responses in one category, precluding a reliable estimate of its difficulty. Fifth, for 1 item (Language comprehension, ADAS-cog) the difficulty level could not be estimated validly and it was therefore misfitting the OPLM (see Supplementary Table 1).

Supplementary Table 1
Items of the CAMCOG-Plus

Item	Short Description	Domain	OPLM cutpoint (Max)	Item Difficulty	Instrument
Q1	Serial 7 s	WM	3/4(5) c	0.47	MMSE
Q2	Spell backwards 'world'	WM	3/4(5) c	-0.18	MMSE
Q3	No ifs and or buts	LA/AT	-/(1) c	0.03	MMSE
Q4	Close your eyes	LA	-/(1) c	-0.91	MMSE
Q5	Three stage command	LA/AT	1/2(3) c	-0.54	MMSE
Q6	Orientation in Time (total score)	OM	3/4(5) c	0.32	MMSE
Q7	Orientation in Place (total score)	OM	4/5(5) c	0.74	MMSE
Q8	Copy 2 Pentagons	CP	-/(1) c	0.16	MMSE
Q9	Write a sentence	LA	-/(1) c	-0.30	MMSE
Q10	Immediate recall of 3 unrelated words	AT	2/3(3) c	-2.68	MMSE
Q11	Delayed recall of 3 unrelated words	EM	1/2(3) c	0.54	MMSE
Q12	Digit Cancellation (either of 2 numbers)	EF	30/31(40) c	-0.38	ADAS-cog
Q13	Naming: fingers and objects	LA	5/6(17) e	-0.86	ADAS-cog
Q14	Following Commands	LA	2/3(5) e	-1.14	ADAS-cog
Q15	Remembering Test Instructions	AT	3/4(5) rs	-1.20	ADAS-cog
Q16	Spoken language ability	LA	1/2(5) rs	-1.16	ADAS-cog
Q17	Word finding difficulty	LA	3/4(5) rs	-1.60	ADAS-cog
Q20	Constructional Praxis: drawing 4 figures	CP	2/3(4) e	-1.36	ADAS-cog
Q21	Ideational Praxis	IP	3/4(5) e	-1.75	ADAS-cog
Q22	Immediate Word Recall	EM	5/6(10) e	0.22	ADAS-cog
Q24	Word Recognition Test	EM	6/7(24) e	-0.15	ADAS-cog
Q25	176-177 Calculation Money	AT/WM	1/2(2) c	-0.37	CAMCOG
Q26	Animal fluency 60 seconds	EF/SM	14/15(-) c	0.78	CAMCOG
Q27	Count Backwards 20-1	WM	1/2(2) c	-0.25	CAMCOG
Q28	Similarities: apple banana	EF/SM	1/2(2) c	-0.23	CAMCOG
Q29	Similarity: shirt dress	EF/SM	1/2(2) c	0.23	CAMCOG
Q30	Similarity: table chair	EF/SM	1/2(2) c	0.56	CAMCOG
Q31	Similarities: plant animal	EF/SM	0/1(2) c	0.58	CAMCOG
Q33	134-136 Semantic Knowledge: Factual Comparisons	SM	2/3(3) c	-0.50	CAMCOG
Q34	138 Semantic Knowledge: Naming objects	LA/SM	5/6(6) c	0.16	CAMCOG
Q35	140-143 Semantic Knowledge: Factual Knowledge	LA/SM	4/5(6) c	0.14	CAMCOG
Q36	Executing verbal Command	LA	-/(1) c	-0.80	CAMCOG
Q37	165-166 Copy figures: 3 d house & spiral	CP	1/2(2) c	0.55	CAMCOG
Q38	Clock Drawing	EF/CP	2/3(3) c	0.56	CAMCOG
Q41	146 Delayed recall of six objects Q34	EM	1/2(6) c	0.43	CAMCOG
Q42	147 Delayed recognition Q34	EM	3/4(6) c	-0.11	CAMCOG
Q43	148-153 Remote Memory	SM	4/5(6) c	0.53	CAMCOG
Q44	154-157 Recent Episodic Memory	EM	3/4(4) c	0.70	CAMCOG
Q46	Recognition famous persons	SM	1/2(2) c	-0.46	CAMCOG
Q50	Insect Fluency 60 seconds	EF/SM	5/6(-) ts	0.91	NPE
Q51	WISC Mazes number of errors	EF	6/7(-) e	0.64	NPE
Q52	WAIS Digit Span Forward	AT/WM	9/10(21) c	0.44	NPE
Q53	WAIS Digit Span Backward	WM	7/8(21) c	1.06	NPE
Q54	Stroop word reading	AT	62/63(-) ts	0.42	NPE
Q55	Stroop color naming	AT	85/86(-) ts	0.48	NPE
Q56	Stroop color-word interference	EF	150/151(-) ts	0.92	NPE
Q57	Trailmaking A numbers.	AT	65/66(-) ts	0.85	NPE
Q58	Trailmaking B numbers + letters	EF	167/168(-) ts	1.01	NPE
Q59	Visual Association Test A	EM	8/9(12) c	0.74	NPE
Q60	Visual Association Test B	EM	8/9(12) c	1.02	NPE
Q61	WAIS Symbol Substitution	EF	33/34(133) c	0.76	NPE
Excluded items/tests					
Q48	Animal fluency 30 seconds	EF/SM	duplicate		NPE
Q49	Animal fluency 120 seconds	EF/SM	duplicate		NPE
Q23	CERAD Delayed Recall	EM	duplicate		ADAS-cog
Q45	178 Recall address	EM	duplicate		CAMCOG
Q19	Orientation (Time/Place)	OM	duplicate		ADAS-cog
Q39	170-174 Ideational and Ideomotor Praxis	IP	duplicate		CAMCOG
Q32	130-133 Following Commands	LA/SM	duplicate		CAMCOG
Q62	Wais Substitution Incidental Learning	EM	see text		NPE

Supplementary Table 1
Continued

Item	Short Description	Domain	OPLM cutpoint (Max)	Item Difficulty	Instrument
Q63	Wais Substitution free recall	EM	see text		NPE
Q40	Tactile Perception (coins)	SM	unreliable		CAMCOG
Q47	Visual perception unusual view	SM	unreliable		CAMCOG
Q18	Language comprehension	LA	misfitting		ADAS-cog
Q64	Naming two objects	LA	>95% of responses in 1 category		MMSE

Domains: AT, attention; CP, constructional praxis; EF, executive functioning; EM, episodic memory; IP, ideational and ideomotor praxis; LA, language; OM, orientation memory; SM, semantic memory; WM, working memory. Instruments: MMSE, mini mental state examination; ADAS-cog, cognitive part of the Alzheimer disease assessment scale; CAMCOG, Cambridge cognitive examination; NPE, neuropsychological examination. WISC, Wechsler intelligence scale for children; WAIS, Wechsler intelligence scale for adults. Notes: item numbers in descriptions refer to original CAMCOG numbering, *** c = correct answers e = errors, rs = rating scale, ts = time in seconds. Correct answers: >cut-off point = 1, <cut-off point = 0, errors, rating scales & time in seconds: <cut-off point = 1, >cut-off point = 0.

For matter of simplicity, the CAMCOG-Plus items were dichotomized and for every item a difficulty level was estimated for the cut-off point at which the item was dichotomized. 18 are CAMCOG items in their original form or assemblies of related items (for Q25, Q33, Q35, Q37, Q43, and Q44), 11 are MMSE items, 10 are ADAS-cog items, and 12 are neuropsychological tests.

ADAPTIVE ADMINISTRATION OF THE CAMCOG AND CAMCOG-PLUS

A computerized adaptive testing (CAT) algorithm using these item difficulties was programmed to tailor the administration of either the CAMCOG or CAMCOG-Plus to each individual participant. CAT estimates a participant's cognitive ability level by only selecting items of appropriate difficulty.

The algorithm was validated with already collected CAMCOG data of patients with Alzheimer's disease (AD) and vascular dementia (VaD) [13;14]. A standard set of items was always administered at the beginning of the test. The set consisted of naming items and delayed recall items and items to assess executive function. There were two practical reasons to administer this set at the beginning of the test. Firstly, CAT often selects delayed recall of previously named and encoded stimuli prior to the naming of the stimuli which makes no sense. Secondly, the items of the set assured the measurement of the important domains of episodic visual and verbal memory, perception and executive function. For the CAMCOG a set of six items (item numbers correspond to original CAMCOG numbers) were presented in the following order:

1. item 138 "naming/encoding pictures"
2. item 171 "writing an address"

3. item 146 "delayed recall of encoded pictures of item 138"
4. item 147 "recognition of objects item 138"
5. item 178 "recall written address of item 171"
6. item 139 "naming as many animals in 1 minute"

For the CAMCOG-Plus a set of eight items (item numbers correspond to item numbers in Supplementary Table 1) were administered:

1. Q34 item 138 "naming/encoding pictures" (CAMCOG).
2. Q1 item "serial 7 s" (MMSE).
3. Q41 item 146 "delayed recall of encoded pictures of item 138" (CAMCOG).
4. Q42 item 147 "recognition of objects item 138" (CAMCOG).
5. Q10 item "encoding of 3 words" (MMSE).
6. Q38 item 167 "clock drawing" (CAMCOG).
7. Q11 item "delayed recall of 3 words" (MMSE).
8. Q26 item 139 "Animal Fluency 60 seconds" (CAMCOG).

After administering these initial item sets, the CAT made a provisional estimate of a patient's ability along with a standard error as a measure of reliability using either the CAMCOG or the CAMCOG-Plus items. Subsequently, the CAT updated the ability estimate and the standard error after each response made by the patient by selecting an item with a difficulty nearest to the patient's provisionally estimated ability. With the administration of each subsequent item, the estimate of the patient's ability became more reliable (a smaller standard error) thereby increasing the adequacy of the item selection. The CAT process terminated after reaching a maximum of 25 items or a standard error of 0.15 which reflects 90% reliability.

REFERENCES

- [1] Fischer GH, Molenaar IW (1995) *Rasch Models: Foundations, Recent Developments and Applications*. Springer-Verlag, New York.
- [2] Verhelst ND, Glas CAW (1995). *The one parameter logistic model*. in: *Rasch Models, Foundations, Recent Developments and Applications*. Fischer GH, Molenaar IW, eds. Springer Verlag, New York.
- [3] Roth M, Tym E, Mountjoy CQ, Huppert FA, 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. *Br J Psychiatry* **149**, 698-709.
- [4] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.
- [5] Lindeboom R, Schmand B, Holman R, de Haan RJ, Vermeulen M (2004) Improved brief assessment of cognition in aging and dementia. *Neurology* **63**, 543-546.
- [6] Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. *Am J Psychiatry* **141**, 1356-1364.
- [7] Wechsler Adult Intelligence Scale (1997) - III. The Psychological Corporation, San Antonio.
- [8] Reitan RM (1955) The relation of the trail making test to organic brain damage. *J Consult Psychol*, **19**, 393-394.
- [9] Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C (2002) Visual association test to detect early dementia of the Alzheimer type. *J Neurol Neurosurg Psychiatry* **73**, 126-133.
- [10] Stroop JR (1935) Studies of interference in serial verbal reactions. *J Exp Psychol* **18**, 643-662.
- [11] Lezak MD (1995) *Neuropsychological assessment*. Oxford University Press, New York.
- [12] Wechsler Intelligence Scale for Children – Third Edition (WISC - III) (1991). The Psychological Corporation, San Antonio.
- [13] Wouters H, Zwinderman AH, Van Gool WA, Schmand B, Lindeboom R (2009) Adaptive cognitive testing in dementia. *Int J Meth Psychiatr Res* **18**, 118-127.
- [14] Wouters H, De Koning I, Zwinderman AH, Van Gool WA, Schmand B, Buitter M, Lindeboom R (2009) Adaptive cognitive testing in cerebrovascular disease and vascular dementia. *Dement Geriatr Cogn Disord* **28**, 486-492.