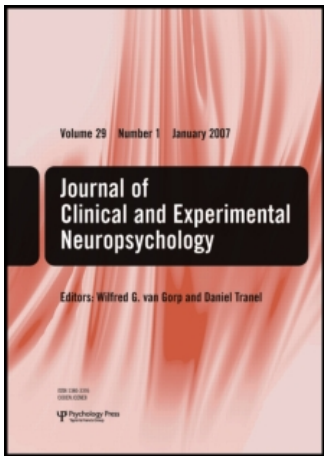

Journal of Clinical and Experimental Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713657736

Optimization of encoding specificity for the diagnosis of early AD: The RI-48 task

First Published on: 20 December 2006
To cite this Article: Adam, S., Van der Linden, M., Ivanoiu, A., Juillerat, A.-C., Bechet, S. and Salmon, E. , 'Optimization of encoding specificity for the diagnosis of early AD: The RI-48 task', Journal of Clinical and Experimental Neuropsychology, 29:5, 477 - 487
To link to this article: DOI: 10.1080/13803390600775339
URL: http://dx.doi.org/10.1080/13803390600775339

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
2007, 29 (5), 477–487

Psychology Press
Taylor & Francis Group

Optimization of encoding specificity for the diagnosis of early AD: The RI-48 task

S. Adam,[1] M. Van der Linden,[2,3,5] A. Ivanoiu,[4] A.-C. Juillerat,[5] S. Bechet,[6] and E. Salmon[6,7]

[1]Neuropsychology Unit, University of Liège, Belgium
[2]Cognitive Psychopathology Unit, University of Liège, Belgium
[3]Cognitive Psychopathology Unit, University of Geneva, Switzerland
[4]Neurology Department, Memory Clinic and Neuropsychological Rehabilitation Center, Saint Luc Hospital, University of Louvain-la-Neuve, Belgium
[5]Memory Clinic, Geriatrics Department, University of Geneva, Switzerland
[6]Memory Clinic, University of Liège, Belgium
[7]Department of Neurology and Cyclotron Research Center, University of Liège, Belgium

The aim of this study was to evaluate the discriminant validity of the *RI-48* test, a shorter French version of the Category Cued Recall portion of the Double Memory Test developed initially by Buschke and colleagues (1997), in the diagnosis of mild and very mild Alzheimer disease (AD). The distinctive feature of the *RI-48* task is that encoding specificity was increased by adding an immediate cued recall stage at the encoding phase. The results show that the *RI-48* task seems to be well adapted to the clinical context and to have good psychometric properties, in particular a lack of a ceiling effect. Moreover, this task appears to be especially well suited for the diagnosis of both mild and very mild AD (sensitivity of 93% and 83.8%). From a more theoretical point of view, this study confirms the importance of optimizing the encoding specificity for the diagnosis of very mild AD, since the more encoding specificity is accentuated, the more discriminating power is increased for the diagnosis of very mild AD.

INTRODUCTION

It is now widely acknowledged that deficits in episodic memory, i.e., memory for events that are separate in time and space, constitute a hallmark symptom of Alzheimer's disease (AD). However, although there is evidence of the usefulness of episodic memory tests in distinguishing AD patients from normal elderly adults, it has also been suggested that not all such tasks are equally useful. For example, several studies (Grober & Buschke, 1987; Grober, Buschke, Crystal, Bang, & Dresner, 1988; Petersen, Smith, Ivnik, Kokmen, & Tangalos, 1994; Pillon, Deweer, Agid, & Dubois, 1993; Tuokko & Crockett, 1989; Tuokko, Vernon-Wilkinson, Weir, & Beattie, 1991) showed that measures of cued recall, especially when the cues at retrieval matched those available at encoding (that is, when semantic category cues are used at both encoding and retrieval) are better at detecting memory impairments associated with dementia than episodic memory tasks involving less cognitive support.[1]

[1] These tasks are based on the theoretical principle of "encoding specificity" (Thomson & Tulving, 1970), which postulates that acquisition and retrieval conditions must be coordinated to maximally enhance spontaneous free recall in an episodic memory test (see also Schacter & Tulving, 1982).

S. Adam is supported by Grant 98/03–215 of the Government of the French-Language Community of Belgium (Actions de Recherche Concertées).

Address correspondence to S. Adam, Neuropsychology Unit, University of Liège, B33 Sart Tilman, B-4000, Liège, Belgium (E-mail: stephane.adam@ulg.ac.be).

http://www.psypress.com/jcen
DOI: 10.1080/13803390600775339

More specifically, Grober and Buschke (1987) designed the Free and Cued Selective Reminding Test (FCSRT) to control and coordinate the encoding and retrieval process. It consists of memorizing 16 words belonging to 16 different semantic categories. The FCSRT involves a controlled encoding stage consisting of the presentation of four items on a card (one item corresponds to the word printed in bold above a picture). Then, appropriate category cues are spoken aloud by the examiner. Participants are asked to find the corresponding picture and name it aloud (e.g., *cow* for *animal*; *Identification phase*). When all four items are correctly identified, the card is masked and immediate cued recall is tested by providing each category cue (e.g., "what was the animal on the card just before?"; *Immediate cued recall phase*). If an item is not recalled in response to its cue, then the procedure of pointing and naming is performed again for this item, and cued recall is tested again. If the participant is still unable to recall the item, the experimenter verbally represents the cue-item pair again, and a final immediate cued recall trial is attempted. The same encoding procedure is applied for the three other cards, each containing four items. After this encoding procedure, three recall phases are carried out, each consisting of two parts: (a) a period of free recall during which the participant has to recall as many words as possible from the encoding phase; and (b) a period of cued recall during which semantic categories are given as cues to elicit the recall of items not retrieved by free recall.

Grober and Buschke (1987) showed that the total cued recall (i.e., the sum of free and cued recall across the three recall phases) had the highest discriminating power for dementia. Indeed, although the demented patients had a weaker free recall performance than the normal elderly group, they also responded less to the semantic cueing. Therefore, the performance gap between the normal elderly and demented groups was higher for the sum of free and cued recall than for the free recall only, which explains the greater sensitivity of this measure for the diagnosis of dementia. However, Grober and Buschke's study was conducted with a relatively small and heterogeneous group of patients (n = 25) including nearly 50% patients with dementias other than AD (n = 12). In addition, these patients were not in the earliest stage of the disease. More recently, Tounsi and colleagues (1999) used the FCSRT to examine a much larger group (n = 131) of well-studied AD patients divided into four subgroups according to their MMSE scores. They confirmed the memory deficit in AD, even in a subgroup of predemented AD patients. More specifically, they showed that normal elderly controls responded correctly to about 94% of the categorical cues, while predemented patients responded to only about 53% of the cues. Consequently, they showed that sensitivity to semantic cuing is the most sensitive index of episodic memory dysfunction in early AD.

Thus the FCSRT seems useful for detecting memory impairments associated with AD. However, the discriminating power of this task is limited by ceiling effects in the performance of the normal elderly group. For example, in the Grober and Buschke (1987) and Tounsi et al. (1999) studies, the total cued recall scores for the nondemented group were respectively 47.8 ± 1.32 and 46.90 ± 1.12 (for a maximum score of 48). In view of these limits, Buschke, Sliwinski, Kuslansky, and Linpton (1997) developed a new memory task (the Double Memory Test: *DMT*) specifically designed to avoid the ceiling effect observed with the FCRST. To do this, the authors significantly increased the number of words to be memorized (64 words instead of 16 in the *FCSRT*). In particular, they increased the memory load on the cue by using four items per category cue (instead of one in the FCSRT, which appeared too easy for the healthy elderly participants). The *DMT* comprises two conditions (each with 64 different words): one condition where encoding specificity is optimized by using the same category cues for encoding and recall phases as in the *FCSRT* (the Category Cued Recall condition: *CCR*); and one condition which does not coordinate encoding and retrieval (that is, semantic category cues are used only for retrieval and not for encoding; the Item Cued Recall condition or *ICR*). Buschke et al. (1997) showed that the *CCR* condition score has substantially greater sensitivity and specificity for the diagnosis of people with dementia (93% sensitivity and 98% specificity) than a memory procedure that does not coordinate encoding and retrieval (the *ICR* condition, with 53% sensitivity and 94% specificity), and also than the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987), the Logical Memory story recall, or the WMS-R Verbal Paired Associates immediate recall task. However, as in the Grober and Buschke (1987) study, one limit of this study was that patient sample was heterogeneous, including 20% of patients with dementia other than AD (i.e., vascular dementia and mixed dementia) and patients who were not in the earliest stages of the disease.

Therefore, Brown and Storandt (2000) attempted more recently to extend Buschke et al.'s (1997) findings by administering the category cued recall portion of the *DMT* (i.e., the 64-items Category

Cued Recall task: *64-CCR*)[2] to a group of AD patients including patients with very mild AD (CDR of 0.5: Berg, 1988). They replicated the sensitivity and specificity of the Buschke et al. *CCR* procedure with respect to the differentiation of mild AD (CDR of 1) from healthy aging. However, the diagnostic accuracy of this test did not extend to the very mild stage of the disease. One possible reason for this weaker discriminating power for the diagnosis of early AD could be the methodology used to coordinate encoding and retrieval. Indeed, as we saw earlier, in the initial paradigm developed by Grober and Buschke (1987), the encoding procedure included both an identification of the items and an immediate cued recall stage. The aim of this encoding procedure was to ensure that items were correctly encoded and that cues were pertinent for the participants. By contrast, the encoding procedure in the Buschke et al. (1997) and Brown and Storandt (2000) studies comprised only the identification of items, without an immediate cued recall stage. Therefore, the authors had no real certainty that information was correctly encoded, and it might be considered that encoding specificity was less optimized in these two studies (i.e., encoding specificity –, or *ES*–) compared to the procedure used in Grober and Buschke's (1987) initial study (i.e., encoding specificity +, or *ES*+). This difference in encoding specificity optimization has more consequences for the performance of the healthy elderly group than for that of the AD group. Indeed, as we saw earlier, AD patients respond less well to the semantic cue and thus their performance varies relatively little with any increase (or decrease) in encoding specificity optimization. On the other hand, for normal elderly participants, the beneficial effect of encoding specificity is significant and thus their performance is reduced if encoding specificity decreases. Nevertheless, this differential sensitivity to cueing has few consequences when one considers the difference between normal aging and mild AD patients as the performance gap between these two groups is very large. In contrast, the difference in the rate of encoding specificity could have greater consequences for the diagnosis of very mild AD patients, as the contrast between the performance of very mild AD and healthy elderly participants is weaker. In other words, reducing encoding specificity optimization decreases the normal elderly group's performance, and thus also decreases the contrast between this group and the very mild AD group, which affect the task's discriminating power.

In this context, we have developed a new, shorter French version of Brown and Storandt's (2000) *64-CCR* task (i.e., the category cued recall portion of Buschke et al.'s DMT): the *RI-48* task (Adam et al., 2004). The special feature of the *RI-48* task is that we have kept the encoding procedure initially developed in the Grober and Buschke (1987) study, which forced encoding by including an immediate cued recall stage. This explains the reduction in the number of items compared to the *64-CCR* task. Since adding the immediate cued recall stage increases the duration of the task, we decreased the number of items (48 items instead of 64) in order to make our test readily applicable in a clinical context. Taking into account these methodological modifications, the first aim of this study is to verify whether the discriminating power of our task replicates that of the tasks used by Buschke et al. (1997) and Brown and Storandt (2000) for the diagnosis of mild AD. The second and more important purpose of this work is to assess whether the greater optimization of encoding specificity in the *RI-48* task (by adding the immediate cued recall stage; i.e., *ES*+) increases the discriminating power of this task for the diagnosis of very mild AD, compared to the Brown and Storandt's (2000) study where encoding included only the identification of items without the immediate cued recall stage (i.e., *ES*–).

METHOD

Participants

A total of 132 individuals participated in this study. All participants were selected in three Memory Clinics (Liège, Brussels and Geneva) between June 1998 and September 2002. They had to have French as their first language and had to be selected using the same procedure. Patients who had moderate to severe dementia, non-AD dementia, neurological or psychiatric conditions without dementia, severe general illnesses or incomplete data were excluded from the study. The screening procedure for dementia followed published guidelines (i.e., those of the Quality Standards Subcommittee of the American Academy of Neurology, 1994). This screening included a clinical examination by a trained clinician, a detailed interview with an informant who knew the subject and cognitive

[2] Brown and Storandt used only the category cued recall portion of the DMT (that is, the condition where the same semantic cue is given both at encoding and retrieval) as this was the condition that provided the greatest discriminating power for the diagnosis of dementia.

screening using the Dementia Rating Scale (DRS; Mattis, 1973), adapted in French (GRECO, 1994 and 1997). A complete neuropsychological examination was carried out in each case and included working and episodic memory tasks, attentional, and language tasks. A brain morphological imaging scan (CT or MRI) was performed for most of the patients. A careful evaluation of mood and anxiety levels was undertaken clinically with the help of specific scales and the patient was referred to a psychiatrist (and excluded from the study) if the results were doubtful.

On the basis of this neurological and neuropsychological examination, the 132 participants were classified into three separate groups (see Table 1): 38 patients with probable early AD (*Pr-AD*: mean DRS = 124.5 ± 9.12), 37 patients with questionable AD (*Q-AD*: mean DRS = 132.7 ± 5.65), and 57 clinically normal participants (*CN*: mean DRS = 140.0 ± 3.44). The three groups did not differ in terms of age [$F(2, 129) = 1.787$; $p = .172$] and education [$F(2, 129) = .703$; $p = .497$]. In contrast, a significant main effect for group was observed on the global DRS score [$F(2, 129) = 72.80$; $p < .0001$]. Post hoc comparisons (using the *Unequal N HSD* test) showed that the DRS score of the *CN* group was superior to the *Q-AD* group ($p < .0001$) and the *Pr-AD* group ($p < .0001$). In addition, the *Q-AD* group performed better than the *Pr-AD* group ($p < .0001$).

The diagnosis of probable AD was made by a senior neurologist according to the criteria developed by NINCDS-ADRDA (McKhann et al., 1984) and the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (*DSM-IV*; American Psychiatric Association, 1994). The vast majority of the *Pr-AD* patients (34 patients out of 38) had a score of 1 on the Clinical Dementia Rating scale (CDR; Berg, 1988) and were therefore in the mild stage of the disease, while the remaining 4 patients had a CDR score of 2. The *Q-AD* group comprised patients who presented to the Memory Clinic with complaints of episodic memory difficulties and/or cognitive decline but did not yet fulfil the criteria for probable AD according to either the DSM-IV or the NINCDS-ADRDA. Complaints were corroborated by an informant and neuropsychological examination showed abnormal performance in only one cognitive domain (most often, episodic memory). However, this abnormal performance did not interfere with daily living activities. These participants had a CDR score equal to 0.5 (see Perry & Hodges, 2000; or Thomson, Graham, Patterson, Sahakian, & Hodges, 2002; for justifications for using the term "questionable AD"). Finally, the *CN* group included 57 nondemented elderly patients who wanted to test their memory in the Memory Clinic as they presented with memory complaints. These participants did not fulfil the criteria for *Pr-AD* or *Q-AD*. Their complaints were not reliably corroborated by an informant and their cognitive screening, including the memory tasks and the memory subtest of DRS, was normal (CDR score of 0). All individuals in this latter group were free of major psychiatric or neurological illnesses but all of them expressed some anxiety about senescence, age-related cognitive decline or Alzheimer's disease.

Material

The task comprised 48 items belonging to 12 different categories (four words for each of the 12 categories; e.g., the "weapon" category had the words for "crossbow", "dagger", "bludgeon", and "pistol").[3] Words were selected according to two criteria. The first was that none of the items was a prototype of its category (the typicality ranged from 6 to 46: mean = 17.6 ± 8.8; from the database of Dubois, 1982). The second criterion was that items selected were not the most frequent in each category (as determined by the "Brulex" frequency database; Content, Mousty, & Radeau, 1990). Words occupying the first two positions in terms of frequency for each semantic category were excluded

[3] Compared to the *64-CCR* task, we preferred to decrease the number of semantic categories rather than the number of items per category. Indeed, as task complexity increases with the cue load, the risk to observing a ceiling effect decreases.

TABLE 1
General characteristics of the three participant groups

				Age			*Education*			*DRS score*		
Groups	*N*	*Sex Male/Female*	*CDR score*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*
CN	57	21/36	0	69.58	8.27	48–85	12.21	2.74	6–17	140.0	3.44	131–144
Q-AD	37	19/18	0.5	72.35	6.41	52–85	11.43	3.55	3–18	132.7	5.65	119–142
Pr-AD	38	13/25	≥1	71.39	6.21	51–85	11.71	3.55	6–17	124.5	9.12	98–140

from the task. The frequency ranged from 0 to 2063 (mean = 399.5 ± 405.6). These two criteria were used in order to avoid selecting the most common words, which participants might recall by guessing.

Procedure

Acquisition phase

The 48 words were presented to participants on 12 different cards (each card containing four items). The order of presentation of cards and words was fixed and unique for all participants. The first three cards contained one item from each category. The second item from each category was presented on the next three cards, recycling through the categories in the same order until all four items from each of the 12 categories had been presented.

Before beginning the encoding phase, participants were told that they had to learn four items from 12 categories and that they would be given category cues to assist their recall. The first card was then placed in front of the participants. They were asked to point to and read aloud each item (e.g., "ladybug") when its category cue was aurally presented (e.g., "insect"). When all four words on a card were correctly named, the card was removed and immediate verbal cued recall was done for just those four items (in the order of the identification), by providing each category cue (e.g., "what was the insect on the card just before?"). If participants were unable to give an item in response to its cue, the card was presented again and the procedure of pointing and naming aloud, followed by the cued recall, was repeated for this item alone. This procedure was repeated until a correct response was obtained. Once the immediate cued recall for a group of four items was completed, the next card with four new items was presented and encoded in the same way.

Cued recall phase

Before the cued recall phase and just after the encoding phase, participants performed an interference task (counting backward) for 20 seconds in order to ensure that the recall involved episodic memory. Afterwards, participants were asked to recall aloud, and in any order, the four items from each category when cues were given verbally by the examiner. They had 30 seconds to recall as many words as possible in each category. The presentation order of the categories was the same as for the encoding phase, thus permitting us to maintain the same interval between acquisition and retrieval.

Variables computed with the RI-48 scale

Different scores were computed with the *RI-48* scale. The *Identification score* corresponds to the number of items correctly identified on the cards at the encoding phase. The *Immediate Cued Recall score* (ICR score) corresponds to the number of items correctly recalled during the encoding phase when category cues were provided, just after the card was removed. The *Delayed Cued Recall score* (DCR score) corresponds to the number of items recalled in response to the category cues during the cued recall phase. The *Intrusion score* (Int score) is the number of extra-list intrusions produced by the participants during the cued recall phase.

In addition, it appeared to us that dissociating the DCR score into two subscores according to performance during the initial immediate cued recall phase could shed some interesting light on the question of the importance of encoding specificity optimization for the early diagnosis of AD. These two subscores corresponded to: (a) the proportion of items encoded at the first trial in the acquisition phase, which were recalled in the cued recall phase (*DCRa*); (b) the proportion of items for which a minimum of two trials were necessary in the acquisition phase, which were recalled in the cued recall phase (*DCRb*).[4] It could be considered that the DCRa subscore corresponds to a condition with greater encoding specificity optimization (*ES+*) than the DCRb subscore (*ES–*). Indeed, in the DCRa condition (with *ES+*), we are sure that the semantic category was a pertinent cue at encoding, and thus, that this condition induces an optimal coordination between encoding and retrieval. In contrast, in the DCRb condition (with *ES–*), the semantic category cue appeared less pertinent for the subjects at encoding. Various reasons may explain this (such as sensitivity to interference, language or semantic difficulties, etc.). However, whatever the reasons may be, it can be considered that we are not sure what operations were realized during encoding. There is therefore less coordination between encoding and retrieval for these items. More specifically, by dissociating these two delayed cued recall subscores, we reproduce a variable

[4] To illustrate these two subscores, we can, for example, consider the performance of one participant on the RI-48 scale. He had an ICR score of 34/48, and a DCR score of 26/48. Of the 34 words encoded with only one trial in the encoding phase, 22 were recalled in the delayed cued recall phase. The DCRa subscore was therefore equal to .65 (i.e., 22/34). Of the 14 words not recalled at the first trial in the encoding phase, 4 were recalled subsequently in the delayed cued recall phase. Therefore, the DCRb subscore was equal to .29 (i.e., 4/14).

modulating the encoding specificity (high versus low encoding specificity: ES+ vs. ES–). Finally, this variable is close to the distinction we made earlier between encoding with (ES+) or without (ES–) and immediate cued recall stage. Thus, the same hypothesis could be put forward, that is to say that higher encoding specificity (i.e., the DCRa subscore) provides greater discriminating power for the diagnosis of very mild AD.

RESULTS

Group differences on the RI-48 scores

Table 2 presents the mean performance of the three groups (*CN, Q-AD, Pr-AD*) for the different scores obtained with the RI-48 scale. A first glance at this table shows that all patients and controls obtained the maximum identification score, suggesting that they were able to correctly point to and read aloud the appropriate item in response to the semantic cue during the acquisition phase. Moreover, we did not observe a ceiling effect for either immediate cued recall (the maximum ICR score observed was 46/48) or delayed cued recall (the maximum DCR-score was 37/48).

A one way ANOVA showed a significant group effect on the ICR score [$F(2, 129) = 35.41$; $p < .0001$]. Post hoc comparisons (using the *Unequal N HSD* test) revealed significant differences between the three groups: the control group performed better than the Q-AD group ($p < .003$), and the Q-AD group performed better than the Pr-AD group ($p < .0001$). The same pattern of performance was observed with the DCR score, with a significant group effect [$F(2, 129) = 112.09$; $p < .0001$]. All post hoc comparisons were significant (CN > Q-AD > Pr-AD: all ps < .0001). The results were somewhat different for extra-list intrusions produced during the delayed cued recall phase. Indeed, a one-way ANOVA showed a significant group effect [$F(2, 129) = 5.64$; $p = .0045$] and post hoc comparisons revealed no difference between the CN and Q-AD groups ($p = .82$), while the Pr-AD group produced significantly more intrusions than the CN ($p = .0076$) and Q-AD groups ($p = .048$).

Concerning the two DCR-subscores (DCRa vs. DCRb), a 3 (Group: CN vs. Q-AD vs. Pr-AD) × 2 (DCRa vs. DCRb) ANOVA revealed a significant main effect of Group [$F(2, 129) = 75.47$; $p < .0001$] confirming that the global recall was better for the CN group than the other two groups, and that the Q-AD group had a better global performance than the Pr-AD group. In contrast, analysis showed no main effect for the DCR subscore [$F(1, 129) = 0.019$; $p = .89$], and no significant interaction [$F(2, 129) = 0.97$; $p = .38$]. The latter results suggest that delayed recall is equivalent whether or not items are correctly encoded.

Discriminant validity

Logistic regression analysis (see Table 3) was used to study the discriminant validity of the RI-48 task. The CN and Pr-AD groups were compared in one analysis; the CN and Q-AD groups in the other. For these analyses, specificity and sensitivity were computed by using the base rate represented in the sample.

The results presented in Table 3 showed a greater classification power for the DCR score than the other RI-48 index, for both the distinction between CN and Pr-AD (the sensitivity was 92.1% and the specificity 94.7% for a total correct classification of 93.7%) and between CN and Q-AD (specificity and sensitivity were respectively 83.8% and 91.2% for a total correct classification rate of 88.3%). Moreover, the DCR score had a greater classification rate than the DRS score for the distinction

TABLE 2
Mean performance of the CN, Q-AD, and Pr-AD groups for the different scores obtained with the RI-48 scale

	CN			*Q-AD*			*Pr-AD*		
	M	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*
Identification score	48.00	0	48–48	48	0	48–48	48	0	48–48
ICR	38.32	4.38	24–46	33.70	5.78	19–44	27.89	7.79	11–43
DCR	25.49	5.00	14–37	15.92	4.06	09–25	11.11	5.04	01–20
• DCRa	.54	.11	.30–.80	.34	.10	.13–.58	.26	.12	.00–.47
• DCRb	.56	.21	.00–1.0	.34	.15	.00–.71	.23	.13	.00–.50
Int	3.42	3.11	0–14	4.24	6.73	0–41	7.45	7.78	0–31

Note. ICR = Immediate Cued Recall; DCR = Delayed Cued Recall; DCRa = delayed cued recall for items correctly encoded; DCRb = delayed cued recall for items not correctly encoded; Int = number of extra-list intrusions during the delayed cued recall phase.

TABLE 3
Sensitivity, specificity, and overall hit rate for the distinction between CN vs. Pr-AD, and CN vs. Q-AD

		Q-AD			*Pr-AD*		
	Cut-off score	*Sensitivity*	*Specificity*	*Overall hit rate*	*Sensitivity*	*Specificity*	*Overall hit rate*
DRS	<133	67.6	82.5	76.6	89.5	96.5	93.7
ICR	<34/48	43.2	89.5	71.3	76.3	93.0	86.3
DCR	<19/48	83.8	91.2	88.3	92.1	94.7	93.7
• DCRa	<.40	81.1	89.5	86.2	86.8	93.0	90.5
• DCRb	<.40	67.6	78.9	74.5	84.2	86.0	85.3
Int	>5	2.7	100.0	61.7	31.6	93.0	68.4

Note. DRS = Dementia Rating Scale; ICR = Immediate Cued Recall; DCR = Delayed Cued Recall; DCRa = delayed cued recall for items correctly encoded; DCRb = delayed cued recall for items not correctly encoded; Int = number of extra-list intrusions during the delayed cued recall phase.

between the CN and Q-AD groups. On the other hand, the global classification rate was equivalent for the distinction between the CN and Pr-AD groups, but the DCR score had greater sensitivity (92.1% versus 89.5% for the DRS score), while the DRS score had a greater specificity (96.5% versus 94.7% for the DCR score).

Finally, the DRS and DCR scores give very close classification rates. In this context, we used the Receiver Operating Characteristic (ROC) curve method to more precisely assess the discriminant validity of the RI-48 as a diagnostic test for AD. This method allows a more thorough comparison of different tests than a single cut-off score, since the whole range of possible cut-off scores is taken into account (Sox, Blatt, Higgins, & Marton, 1988).[5] We plotted the sensitivity against 1-specificity for each of the RI-48 indices and for the DRS score (see Figure 1a for the Q-AD group and Figure 1b for the Pr-AD group). These figures showed that the DCR score had a better discriminant validity both for the distinction between the CN and Pr-AD groups (the area under the curve was 0.989 ± 0.007), and between the CN and Q-AD groups (the area under the curve was 0.933 ± 0.027) than the DRS score (the area under the curve was 0.973 ± 0.016 for the discrimination of Pr-AD, and 0.872 ± 0.036 for the discrimination of Q-AD) and the other RI-48 scores.

As suggested by Gifford and Cummings (1999), although calculating the sensitivity and specificity provides useful information on the diagnostic validity of a test, the clinician would rather know the probability that a patient with a positive (or negative) score has (or does not have) the disease (i.e., the positive predictive value [PPV] and the negative predictive value [NPV]). The PPV and NPV depend on the sensitivity and specificity of the test, but also and more importantly, on the prevalence of disease in the population being tested. In this context, to determine the discriminant validity of a test, it is also important to assess the positive (PPV) and negative (NPV) predictive values of each test across a range of disease prevalence. Thus, Table 4 shows changes in PPV and NPV for each RI-48 score and the DRS score when the prevalence of the disease decreases from 50% to 5%. The results showed that the decrease in PPV as a function of the decrease in prevalence of cases in the population appear weaker for the DCR score than for the DRS score and the other RI-48 scores. Indeed, when the expected prevalence of cases in the population decreases from 50% to 5%, the PPV of the DCR score decreases from 92% to 66% for the *Pr-AD* group, and from 86% to 43% for the *Q-AD* group. In contrast, for the DRS score, PPV decreased from 84% to 37% for the *Pr-AD* group, and from 73% to 30% for the *Q-AD* group (and this reduction was still more pronounced for the other RI-48 scores).

Distinction between DCRa and DCRb subscores

Whatever the level of analysis (sensitivity, specificity, area under the ROC curve, PPV, or NPV), the results showed greater discriminating power for the DCRa subscore than for the DCRb subscore. The sensitivity and specificity of the DCRa index (see Table 3) were respectively 85.8% and 93.0% for the *Pr-AD* group (versus 84.2% and 86.0% for the DRCb index), while for the *Q-AD* group, sensitivity and specificity were 81.1% and 89.5% for the DCRa index (versus 67.6% and 78.9% for the

[5] The discriminant validity of a test is determined by the area under the curve: the larger the area under the curve, the greater the diagnostic power.

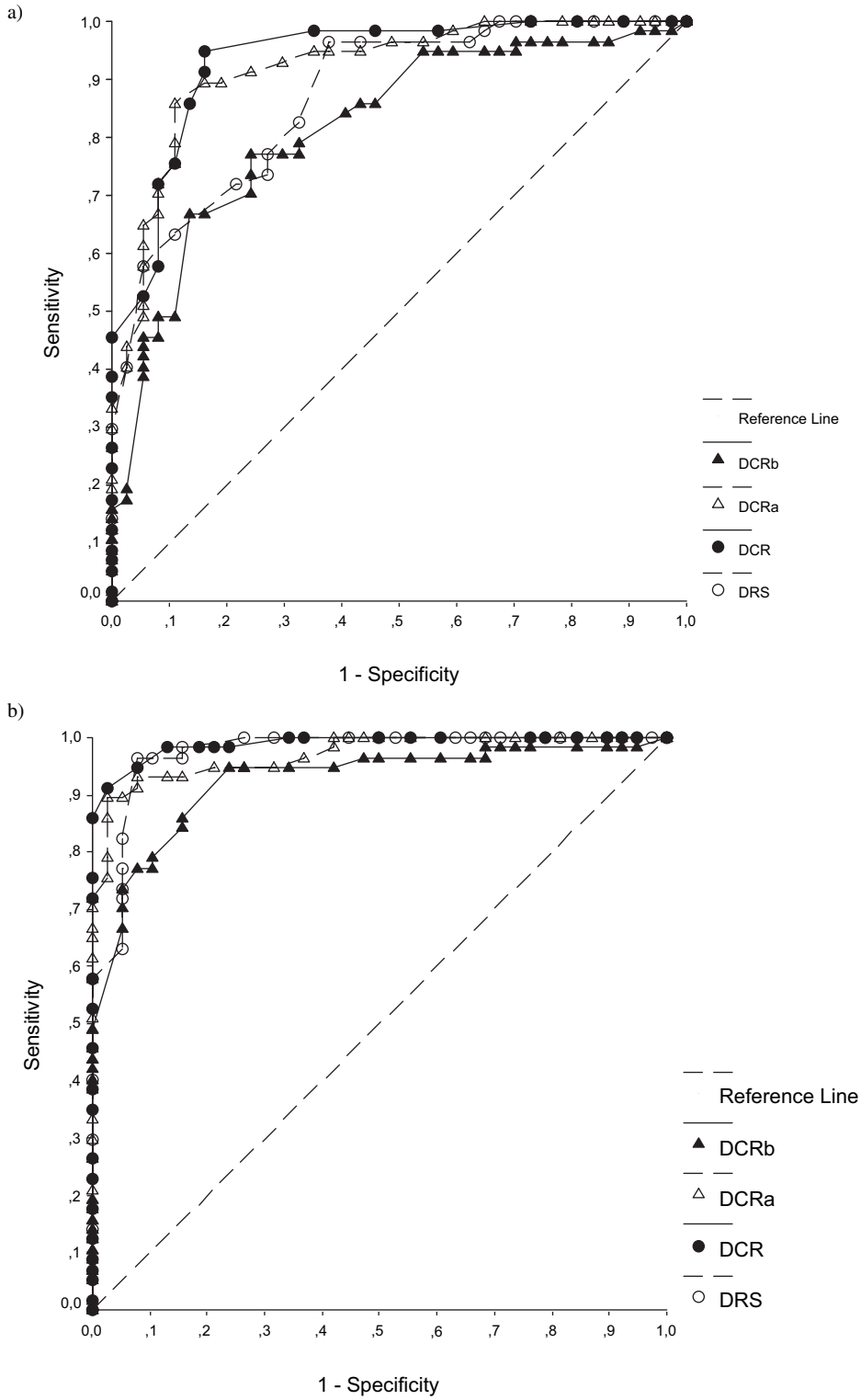

Figure 1. ROC curves for the RI-48 DCR, DCRa, and DRCb scores, and the DRS for discrimination between CN and Q-AD groups (a) and between CN and Pr-AD (b).

DCRb index). The area under the curve (see Figure 1a and 1b) was greater for the DCRa index (0.970 ± 0.015 for the *Pr-AD* group, and 0.919 ± 0.029 for the *Q-AD* group) than for the DCRb index (the area under the curve was 0.924 ± 0.028 for the *Pr-AD* group; and 0.818 ± 0.044 for the *Q-AD* group).

This superiority of the DCRa score was also observed in the stability of its PPV with a decrease

TABLE 4
Predictive value of the different RI-48 scores and the DRS score for different base rates of alzheimer disease (from 50% to 5%)

	Positive predictive value				*Negative predictive value*			
	50	*20*	*10*	*5*	*50*	*20*	*10*	*5*
CN vs. Pr-AD								
DRS	84	55	50	37	96	100	100	100
ICR	84	42	29	26	89	96	96	100
DCR	92	79	76	66	95	98	98	100
• DCRa	92	68	58	53	93	95	98	100
• DCRb	89	53	32	18	77	96	96	98
Int	39	13	5	2	81	100	100	100
CN vs. Q-AD								
DRS	73	38	32	30	77	96	100	100
ICR	70	21	5	3	74	96	98	100
DCR	86	65	54	43	86	98	98	98
• DCRa	89	65	49	24	84	95	96	100
• DCRb	78	24	5	0	70	96	98	100
Int	46	0	0	0	56	100	100	100

Note. DRS = Dementia Rating Scale; ICR = Immediate Cued Recall; DCR = Delayed Cued Recall; DCRa = delayed cued recall for items correctly encoded; DCRb = delayed cued recall for items not correctly encoded; Int = number of extra-list intrusions during the delayed cued recall phase.

in the disease prevalence rate (see Table 4). For the *Pr-AD* group, the PPV of the DCRa index decreases from 92% to 53% (versus from 89% to 18% for the DCRb index) when the disease prevalence rate decreases from 50% to 5%. For the *Q-AD* group, the PPV of the DCRa index decreases from 89% to 24% while the PPV of the DCRb index decreases from 78% to 0%. Conversely, when the disease prevalence increases from 5% to 50%, the NPV of the DCRb decreases more than the NPV of the DCRa. Indeed, for the *Pr-AD*, the NPV of the DCRb index decreases from 98% to 77% (versus from 100% to 93% for the DCRa index); and for the *Q-AD*, the NPV of the DCRb index decreases from 100% to 70% while the decrease was only from 100% to 84% for the DCRa index.

DISCUSSION

The aim of this study was to evaluate the psychometric properties and discriminant validity of the *RI-48* test, a shorter French version of the *64-CCR* task (i.e., the Category Cued Recall portion of the Double Memory Test developed initially by Buschke et al., 1997), in the diagnosis of mild and very mild AD. The particular feature of the *RI-48* task is that the encoding phase comprises an immediate cued recall stage, contrary to Buschke et al. (1997) and Brown and Storandt's (2000) studies in which participants are only asked to identify items. Indeed, we consider that encoding specificity is more optimized in the encoding procedure with an immediate cued recall stage (*ES+*) compared to the encoding procedure with only identifications of items (*ES–*). This reduction of encoding specificity in the 64-CCR task could explain the weaker discriminating power of this task for the diagnosis of very mild AD as observed in Brown and Storandt's (2000) study. In this context, our hypothesis was that the greater optimization of encoding specificity in the RI-48 task, because of the immediate cued recall stage at encoding, would make it more sensitive in the diagnosis of very mild AD.

Concerning the psychometric properties of the RI-48 task, our data showed that, despite the reduced number of items, no ceiling effect was observed even for the immediate cued recall score. With regard to discriminating power, our results replicate the sensitivity of the delayed cued recall score with respect to the distinction between healthy aging and mild AD, as observed in Buschke et al. (1997) and Brown and Storandt's (2000) studies with the 64-CCR task (i.e., sensitivity of 92.1% in our study versus 93% for Buschke et al., and 88% for Brown and Storandt). The specificity was, however, slightly weaker than in Buschke et al.'s study (94.1% versus 98.3%). In addition, the positive predictive value of our data decreased more with the decrease in disease prevalence than in the Buschke et al. study: with a decrease of disease prevalence from 50% to 5%, the PPV decreased from 92% to 66% in our study, compared to a drop from 98.8% to 81.6% in the Buschke et al. study. The main reasons for this reduction in specificity could be that, on the one hand, the patient group was more heterogeneous in Buschke et al.'s study (including not only AD patients but also subjects with vascular dementia and mixed dementia), and on the other hand, their control group was more homogeneous. Indeed, Buschke et al. (1997) recruited participants for their control group from the general population (i.e., people who had not consulted a memory clinic regarding a diagnosis of AD), while we included in this group only nondemented elderly patients who wanted to test their memory in a memory clinic but who were diagnosed as clinically normal. The reason for this choice was that we would like to situate our study in a more clinical context, and the interest for the clinician is to be able to classify elderly people with memory complaints as normal or beginning AD. However since we did this, our control group is probably more heterogeneous than Buschke et al.'s control group, which may

explain our weaker stability of PPV with a decrease in disease prevalence compared to Buschke et al.'s study.

Moreover, and more importantly, our data showed that the discriminant validity of the delayed cued recall score was extended to the questionable AD group. Indeed, the diagnostic advantage of the delayed cued recall score appeared to be slightly smaller than the distinction between the normal aging and probable AD groups. Nevertheless, the result suggests that the *RI-48* task also remains suitable for the diagnosis of the very mild stage of the disease (sensitivity of 83.8% and specificity of 91.2%). In addition, the discriminating power of our task for the diagnosis of early AD (Q-AD) is considerably higher than that observed in the Brown and Storandt's (2000) study (with a sensitivity of 62% and a specificity of 95%).

The major reason for this difference probably resides in the encoding procedure used. Indeed, the *64-CCR* task used in the Brown and Storandt's (2000) study comprised only the identification of items without an immediate cued recall stage. Therefore, in this encoding procedure, the identification of items is very easy and requires very few cognitive resources. Participants can be relatively passive and still perform this identification task correctly. By contrast, in the encoding procedure with an immediate cued recall, as initially developed by Grober and Buschke (1987), participants have to use semantic cues to engage in an active process of retrieval, which requires more cognitive resources. For Grober and Buschke (1997), in addition to the fact that the immediate cued recall stage serves to confirm the correct initial encoding of items, it also permits subjects to add semantic information to the memory trace (Rabinowitz & Craik, 1986), and consequently to accentuate the importance of the cue at encoding. Therefore, it can be argued that encoding specificity is greater in this condition than in the condition comprising only an identification of items.

Despite the absence of an immediate cued recall stage at encoding, the 64-CCR task has a high discriminating power for the diagnosis of mild AD (93% sensitivity in the Buschke et al., 1997, study; and 88% sensitivity in the Brown & Storandt, 2000, study). Nevertheless, the absence of the immediate cued recall stage (and thus, the reduction in encoding specificity) has deleterious effects for the diagnosis of very mild AD. As argued in the introduction, the smaller difference in performance between the very mild AD patients and the elderly participants, coupled with the differential benefit from encoding specificity for these two groups (with AD patients gaining no/or less benefit while elderly participants benefit greatly; see for example, Tounsi et al., 1999) could explain why the reduction in encoding specificity leads to a reduction in the diagnostic power of the 64-CCR task.

Moreover, we reproduced the same pattern of results (i.e., similar discriminating power for the diagnosis of mild AD between the ES+ and ES- conditions; and greater discriminating power for the diagnosis of very mild AD for the ES+ condition compared to the ES- condition), with the distinction between the DCRa subscore (i.e., the delayed cued recall of items correctly encoded in the immediate cued recall stage, which corresponds to the condition with higher encoding specificity optimization; ES+) and the DCRb subscore (i.e., the delayed cued recall of items for which a minimum of two trials was necessary in the acquisition phase and for which we hypothesized that encoding specificity is weaker; ES–).

In conclusion, the RI-48 task seems to be well adapted to the clinical context. It may be argued that the administration time is long (i.e., 20 to 25 minutes). However, it must be emphasized that this administration time is equivalent to that of the FCRST task (Grober & Buschke, 1987) and the 64-CCR task (Buschke et al., 1997). Moreover, the RI-48 task has good psychometric properties (in particular a lack of a ceiling effect) and is especially well suited for the diagnosis of both mild and very mild AD. In addition, we have recently conducted a study (Ivanoiu et al., 2004) where we directly compared the diagnostic power of our RI-48 task with other memory tasks including two with a delayed recall phase (one on verbal material and the other on visual material). The results confirmed the great discrimination power of the DCR score, and especially, they showed that the diagnostic power of the RI-48 task is greater than that of the other memory tasks. Therefore, although the RI-48 task is long, it is also more sensitive for the diagnosis of AD. Finally, from a more theoretical point of view, our study confirms the importance of optimizing the encoding specificity for the diagnosis of very mild AD, since the more encoding specificity is accentuated, the more discriminating power is increased for the diagnosis of very mild AD.

Original manuscript received 5 January 2004
Revised manuscript accepted 16 March 2004
First published online 20 December 2006

REFERENCES

Adam, S., Van der Linden, M., Poitrenaud, J., Kalafat, M., et les membres du groupe GRECO Mémoire (2004). L'épreuve de rappel indicé à 48 items (RI-48).

In M. Van der Linden, S. Adam, A. Agniel, C. Baisset Mouly, et al. (Eds.), *L'évaluation des troubles de la mémoire: Présentation de quatre tests de mémoire épisodique (avec leur étalonnage)*. Marseille, France: Solal.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders. Fourth edition*. Washington DC: American Psychiatric Association.

Berg, L. (1988). Clinical Dementia Rating (CDR). *Psychopharmacology Bulletin, 24*, 637–639.

Brown, L.B., & Storandt, M. (2000). Sensitivity of category cued recall to very mild dementia of the Alzheimer type. *Archives of Clinical Neuropsychology, 15*, 529–534.

Buschke, H., Sliwinski, M., Kuslansky, G., & Lipton, R.B. (1997). Diagnosis of early dementia by the Double Memory Test: Encoding specificity improves diagnostic sensitivity and specificity. *Neurology, 48*, 989–997.

Content, A., Mousty, P., & Radeau, M. (1990). BRULEX: Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique, 90*, 551–566.

Dubois, D. (1982). *Normes de production d'exemplaires appartenant à vingt-deux catégories sémantiques, à partir d'une consigne "classique" et d'une consigne d'imagerie*. Université de Paris VIII, unpublished manuscript.

Gifford, D.R., & Cummings, J.L. (1999). Evaluating dementia screening tests: Methodologic standards to rate their performance. *Neurology, 52*, 224–227.

Grober, E., & Buschke, H. (1987). Genuine memory deficits in dementia. *Developmental Neuropsychology, 3*, 13–36.

Grober, E., Buschke, H., Crystal, H., Bang, S., & Dresner, R. (1988). Screening for dementia by memory testing. *Neurology, 38*, 900–903.

Ivanoiu, A., Adam, S., Béchet, S., Georges, M., Godfrind, G., Jacquemin, A., Prairial, C., Juillerat, A.C., Mulligan, R., Salmon, E., & Van der Linden, M. (2004). The episodic memory evaluation in incipient Alzheimer's disease. *Journal of Neurology, 252*, 47–55.

Mattis, S. (1973). *Dementia rating scale*. Windsor, UK: NFER-Nelson.

McKahnn, 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.

Perry, R.J., & Hodges, J.R. (2000). Fate of patients with questionable (very mild) Alzheimer's disease: Longitudinal profiles of individual subjects' decline. *Dementia and Geriatric Cognitive Disorders, 11*, 342–349.

Petersen, R.C., Smith, G.E., Ivnik, R.J., Kokmen, E., & Tangalos, E.G. (1994). Memory function in very early Alzheimer's disease. *Neurology, 44*, 867–872.

Pillon, B., Deweer, B., Agid, Y., & Dubois, B. (1993). Explicit memory in Alzheimer's, Huntington's and Parkinson's diseases. *Archives of Neurology, 50*, 374–379.

Quality Standards Subcommittee of the American Academy of Neurology (1994). Practice parameter for diagnosis and evaluation of dementia (summary statement). *Neurology, 44*, 2203–2206.

Rabinowitz, J.C., & Craik, F.I.M. (1986). Prior retrieval effects in young and old adults. *Journal of Gerontology, 41*, 368–375.

Schacter, D.L., & Tulving, E. (1982).Amnesia and memory research. In L.S. Cermak (Ed.), *Human memory and amnesia* (pp. 1–32). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sox, H.C., Blatt, M.A., Higgins, M.C., & Marton, K.I. (1988). *Medical decision making*. Stoneham, MA: Butterworth-Heinemann.

Thomson, S.A., Graham, K.S., Patterson, K., Sahakian, B.J., & Hodges, J.R. (2002). Is knowledge of famous people disproportionately impaired in patients with early and questionable Alzheimer's disease? *Neuropsychology, 16*, 344–358.

Thomson, D.M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86*, 255–262.

Tounsi, H., Deweer, B., Ergis, A.-M., Van der Linden, M., Pillon, B., Michon, A., & Dubois, B. (1999). Sensitivity to semantic cuing: An index of episodic memory dysfunction in early Alzheimer disease. *Alzheimer Disease and Associated Disorders, 13*, 38–46.

Tuokko, H., & Crockett, D. (1989). Cued recall and memory disorders in dementia. *Journal of Clinical and Experimental Neuropsychology, 11*, 278–294.

Tuokko, H., Vernon-Wilkinson, R., Weir, J., & Beattie, B.L. (1991). Cued recall and early identification of dementia. *Journal of Clinical and Experimental Neuropsychology, 13*, 871–879.

Wechsler, D.A. (1987). *Manual: Wechsler Memory Scale revised*. New York: Psychological Corporation.