OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 31 (2016) 165–175

Archives of CLINICAL NEUROPSYCHOLOGY

Comparison of the Psychometric Properties of the "Word" and "Picture" Versions of the Free and Cued Selective Reminding Test in a Spanish-Speaking Cohort of Patients with Mild Alzheimer's Disease and Cognitively Healthy Controls

Carolina Delgado[1,2,*], Carlos Muñoz-Neira[3], Ambar Soto[4], Melissa Martínez[1,2], Fernando Henríquez[3], Patricia Flores[5], Andrea Slachevsky[3,5,6,7,8]

[1]*Neurology and Neurosurgery Department, Hospital Clínico Universidad de Chile, Santiago 8380456, Chile*
[2]*Neurology Department, Clínica Santa Maria, Santiago 8420485, Chile*
[3]*Cognitive Neurology and Dementia Unit. Neurology Department Hospital del Salvador Santiago 7500922 Chile*
[4]*Mutual de Seguridad, Santiago, Chile*
[5]*Program of Pathophysiology ICBM and east Neuroscience Department Faculty of Medicine University of Chile Santiago 7500922 Chile*
[6]*FONDAP Geroscience Center for Brain Health and Metabolism Santiago 7500922 Chile*
[7]*Centro de Investigación Avanzada en Educación, Universidad de Chile, Santiago 8330014 Chile*
[8]*Neurology Department, Clinica Alemana Santiago 7650568 Chile*

*Corresponding author at: Tel.: +56-229788716
E-mail address: carodede@gmail.com (C. Delgado).

Received on September 10, 2015; revised on December 3, 2015; Accepted 9 December 2015

Abstract

The aim of this study was to compare the psychometric properties of the "Word" and "Picture" versions of the Spanish FCSRT across the same sample of mild Alzheimer disease (AD) patients and controls. Mild AD patients ($N = 50$, 27 CDR = 0.5; 23 CDR = 1) and controls ($N = 42$, CDR = 0) were assessed with an extensive clinical and neuropsychological evaluation. Psychometric characteristics for both versions of the FCSRT were compared. Free recall (FR) and total recall (TR) across both versions of the FCSRT showed areas under the curve >0.9 and did not significantly differ between them. The scores of both versions were well correlated, although the scores for the Picture version were greater than those for the Word version, particularly for the TR scores of the mild AD group. Both versions of the FCSRT showed an appropriate accuracy to distinguish mild AD patients and controls. Visual cues were easier to recall than verbal cues, especially in the memory impaired patients.

Keywords: Alzheimer's disease; Neuropsychological tests; Memory disorders; Cues; Psychometrics; ROC curve

Introduction

Dementia rates have been progressively increasing worldwide, especially in Latin America and Asia (Prince et al., 2013). Alzheimer's disease (AD) is the most common type of dementia (Fargo & Bleiler, 2014), and one of its earliest manifestations is episodic memory impairment (Di Paola et al., 2007). Therefore, memory assessments are essential for the early diagnosis of AD (Dubois et al., 2007, 2014). Grober and Buschke (1987) stressed that to identify the genuine storage deficits that are characteristic of hippocampal impairment due to AD using neuropsychological evaluations of memory performance, it is critical to control for the effective encoding of all of the items that should be recalled. Similarly, these specific amnesic deficits must be distinguished from the memory deficits that result from normal ageing, secondary to impaired attention, due to inefficient information processing, or that can be explained by ineffective retrieval (Grober & Buschke, 1987).

The Free and Cued Selective Reminding Test (FCSRT) is a memory assessment that includes a specific procedure to ensure that attention is focused on the task. In addition, it encourages further semantic processing during the encoding phase (controlled encoding). Correspondingly, the FCSRT includes particular administration conditions in which encoding specificity is increased by

doi:10.1093/arclin/acv107 Advance Access publication on 11 January 2016

coordinating encoding and retrieval (Grober, Buschke, Crystal, Bang, & Dresner, 1988). Previous studies have demonstrated that the FCSRT is an appropriate tool to detect Alzheimer's disease at the early stages (Grober et al., 1988; Petersen, Smith, Ivnik, Kokmen, & Tangalos, 1994), predict future cases of dementia (Auriacombe et al., 2010; Grober, Lipton, Hall, & Crystal, 2000; Sarazin et al., 2007), identify patients with mild cognitive impairment who are at a higher risk for developing AD (Sarazin et al., 2007), and differentiate AD from other types of dementia (Grober, Hall, Sanders, & Lipton, 2008; Lemos, Duro, Simões & Santana, 2014; Pillon, Deweer, Agid, & Dubois, 1993).

Although the aforementioned characteristics of the FCSRT may support its extended use in clinical contexts, the use of considerably different memory measurements, such as the Free Recall (FR) or the Total Recall (TR), together with the different cut-off points (COPs) that have been proposed to identify dementia, has resulted in discrepancies in the interpretation of the FCSRT (Grober, Sanders, Hall, & Lipton, 2010). Such discrepancies may be explained, in brief, by three main arguments. Firstly, due to its application to diverse populations, i.e., memory clinics versus primary care cohorts; secondly, because of differences in the application procedure of the task (immediate recall inclusion or not); thirdly, as a result of the use of two versions of the task with dissimilar stimuli, namely the "Word" (verbal) and "Picture" (visual) versions (Grober et al., 2010).

Recently, Zimmerman and colleagues (2015) began to explore this issue by applying both the "Word" and "Picture" approaches of the FCSRT (including immediate recall) in the same population of healthy older adults. This study indicated that pictures were recalled more easily than words. As a consequence, the scores for the Picture version were higher than those for the Word version (Zimmerman et al., 2015). Although this was an important finding, the study did not include patients with AD, and the two versions of the task have not yet been directly compared across the same cohort of patients with memory impairments and cognitively healthy elderly controls. The present study attempted to compare the performances on the Word (verbal) and Picture (visual) versions of the Spanish FCSRT across the same population. Thus, two main objectives were proposed: (i) to compare the psychometric properties of the Word and Picture versions of the Spanish FCSRT in differentiating patients with mild AD from cognitively healthy controls and (ii) to compare the results of the different memory measures obtained for both versions of the FCSRT across the same cohort of mild AD and cognitively healthy controls.

Subjects and Methods

Participants

All patients included in this study were recruited from two memory clinics: the Cognitive Neurology and Dementia Unit of the Neurology Department, Hospital del Salvador, and the Neuropsychology Unit of the Neurology and Neurosurgery Department, Hospital Clínico Universidad de Chile (HCUCH), which are both located in Santiago, Chile. The controls were recruited from a variety of sources, including spouses or relatives of the patients with dementia and older adults who regularly attended community groups of elderly people. The study was carried out in a convenience sample of 92 Spanish-speaking participants who were older than 59 years of age. The inclusion criteria considered subjects with an appropriate capacity to provide consent for research and included either patients who had been diagnosed with mild AD or cognitively healthy individuals. The exclusion criteria included illiteracy, underlying neurological or psychiatric illness that could affect cognition (except for AD), physical disability, sensory disturbance, or severe cognitive impairment that could interfere with the neuropsychological assessment, and the absence of a reliable proxy. The sample was divided into two groups. The mild AD group includes 50 patients, and the control group includes 42 individuals.

Each participant was evaluated for dementia by a neurologist based on detailed neurological, neuropsychological, laboratory, and brain imaging data for each participant. Specifically, the neurologist determined whether the participants fulfilled the National Institute on Aging and the Alzheimer's Association work group criteria of dementia due to AD (McKhann et al., 2011). To avoid diagnostic circularity, the results of the FCSRT were not considered to diagnose dementia.

The severity of dementia was rated based on the clinical dementia rating (CDR), a measure of clinical progression and staging (Morris, 1993). Notably, only participants with dementia that was rated between 0.5 and 1 on the CDR scale were considered to suffer from mild AD and included in this study. The control group consisted of cognitively healthy subjects who had CDR scores of 0. Additionally, controls did not report memory complaints, and their cognitive performance was considered normal according to local normative Mini-Mental State Examination (MMSE) scores (>21) (Folstein, Folstein, & McHugh, 1975; Quiroga, Albala, & Klaasen, 2004) and Addenbrooke's Cognitive Examination-Revised (ACE-R) scores (>76) (Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006; Muñoz-Neira et al., 2012).

All participants had a proxy who had known them for at least 5 years. Specifically, a person that was able to provide information about the activities of daily living performance and the behavior of the participants as well as a general medical history was considered a proxy. All proxies were interviewed together with the participants to obtain the CDR score.

Standard Protocol Approval and Patient Consent

All participants provided informed consent in accordance with the Declaration of Helsinki. The protocol utilized in this study was approved by the Ethical and Scientific Committee of Servicio de Salud Metropolitano Oriente and HCUCH.

Neuropsychological Evaluation

a. The FCSRT: Experienced clinical psychologists who were extensively trained in the conduction of neuropsychological evaluations and who were blinded to the condition of each subject (dementia patient or normal control) administered the Spanish versions of the FCSRT. They first administered the Word version of the Spanish FCSRT (words as stimuli). Seven days later, the Picture version was administered (black and white line drawings). A committee of experts in neuropsychological assessment carefully selected each stimulus included in both versions of the task, considering neutral stimuli adapted to the Spanish language. The frequencies of use in the Spanish language of the stimuli used in both versions of the FCSRT were assessed, and no significant differences between the Word and the Picture versions of the task were found. Each version consisted of different semantic categories (items) to avoid a learning effect between each cognitive assessment. Thus, both Spanish versions of the FCSRT were applied according to the procedure described by Grober and Buschke (1987). The subjects were required to memorize a set of items as follows: (i) Study phase: 16 elements were presented in groups of four on successive cards. Items associated to a unique semantic category cue, such as "metal" or "fruit," were presented in each quadrant of the card. In the Word version, the items consisted of words, such as "aluminum," whereas in the Picture version, the items consisted of drawings, such as a drawing of grapes. The subject was asked to verbally name and note aloud each item after its cue was aurally and contingently presented. Once all of the first four items had been identified and appropriately encoded, the card was removed, and immediate cued recall of the first four items was tested by mentioning aloud the cues to the subject to control the encoding process. If the subject was not able to recall the presented items, the same procedure was repeated only for the items that had been forgotten. Once immediate recall for the first group of four items was completed, the next set of four items was presented until all 16 items had been encoded. This first phase of the test allows encoding to be controlled and provides an immediate recall (IR) score. (ii) Test phase: The subsequent memory phase of the test consisted of three successive recall trials. The first trial was conducted 60 s after the study phase, and Trials 2 and 3 were conducted 20 s after the first and second recall trials, respectively. Subjects counted backwards from 100 while they waited for each trial to begin. Each recall trial consisted of two parts: first, each subject was allowed up to 2 min to freely recall as many items as possible; second, a semantic category ("What was the name of the metal?" or "What was the name of the fruit?") was aurally provided for the items that had not been spontaneously retrieved by the subject. This phase thus provided an FR score and a TR score, where FR score consisted of the total number of freely recalled items across the three trials, and TR score consisted of the total number of both the free and cued recalls across the three trials. The maximum total score was 48 for both the FR and TR scores when the sum of the three trials is considered. A sensitivity to cueing index, which was calculated as a percentage estimated with the following formula: $(TR - FR)/(48 - FR) \times 100$, was used to evaluate the efficacy of the semantic cues in facilitating retrieval of stored information.

Describing the administration procedure of the other two remaining stages of the FCSRT, namely the delayed recall and recognition phases, goes beyond the scope of this study.

Overall, both versions of the FCSRT provide several memory measurements, such as IR within the study phase, FR, TR, and sensitivity to cueing index within the test phase.

Statistical Analysis

The Statistical Package for the Social Sciences (SPSS®) version 20 for Windows (IBM Corp., Armonk, NY, USA) was used to analyze the data.

Demographic and neuropsychological data were compared between the mild AD and control groups using χ^2 tests, unpaired two-tailed *t*-tests and ANCOVA (including age and years of education as covariates) where indicated. Differences with a $p < .05$ were considered significant. Notably, only the FR and TR of the test phase were considered for the analyses of discriminant validity for several reasons. First, the FR and TR of the short-term memory phase have been the most widely used indicators to reflect the capability of the FCSRT to differentiate between AD patients and cognitively healthy controls (Grober et al., 2010). Second, the cued recall, TR, and sensitivity to cueing index of the FCSRT are all considered acceptable measures of the efficacy of semantic cues in facilitating retrieval of already stored information, and these parameters have also been shown to be highly correlated among them. Third, in this study, the FR and TR were preferred because they are easier to analyze. To estimate the ability of the verbal and visual versions of the FCSRT to distinguish mild AD patients from controls, standardized mean differences and effect sizes (Cohen's *d* and Pearson's *r*, respectively) as well as 95% confidence intervals were calculated to assess the magnitude

of the differences in the scores for the visual and verbal versions of the FCSRT between the different subject groups. According to Cohen, effect sizes between 0.2 and 0.49 are considered small; those between 0.5 and 0.79, moderate; and those >0.8, large (Cohen, 1988). Similarly, the FR and TR of the Word (WO) and Picture (PI) versions (WO_FR, WO_TR, PI_FR, PI_TR) of the FCSRT were compared. Receiver-operating characteristics (ROC) curves were generated to calculate areas under the curve (AUCs) and the best cut-off point (COP) considering a specificity >0.9 and the best possible sensitivity. AUCs were used as a measure of diagnostic utility and were classified as having excellent (>0.9), good (>0.8), fair (>0.7), or poor (>0.6) utility (Gifford & Cummings, 1999). The method suggested by Hanley and McNeil (1983) was used to compare AUCs between the memory measures. Specifically, z scores were estimated, and differences between AUCs were considered significant when the z scores were ≥ 1.5. Additionally, the McNemar Test was used to compare sensitivities across the different memory measures (McNemar, 1947). The convergent validity of both versions of the FCSRT was evaluated using Pearson's r correlation coefficients between the Word and Picture FCSRT scores and the results of the other administered neuropsychological tests.

Global cognitive impairment was measured with the Chilean versions of the MMSE (Quiroga et al., 2004) and ACE-R (Mioshi et al., 2006; Muñoz-Neira et al., 2012). The Frontal Assessment Battery (FAB; Dubois, Slachevsky, Litvan, & Pillon, 2000) as well as portions of the Fluency domain of the ACE-R (1-min phonemic verbal fluency with the letter P and 1-min category fluency with animals) (Canning, Leach, Stuss, Ngo, & Black, 2004) were used to assess executive function and processing speed. The Memory domain of the ACE-R was used as an indicator of episodic memory. Internal consistency was measured by calculating Cronbach's α coefficients. The various memory measures of the Word version were compared with the corresponding measures of the Picture version using paired t-tests.

Results

Demographic and Clinical Data

Table 1 provides the clinical characteristics of the two groups. The mild AD group consisted of 50 patients; 27 with a CDR = 0.5, and 23 patients with a CDR = 1. The control group consisted of 42 participants without cognitive impairment and a CDR = 0. Gender (56% female in the mild AD group versus 71% in the control group, $\chi^2 = 2.3$ $p = .14$), years of education (12 ± 5 versus 13 ± 4; $t = -1.7$, $p = .09$), or years of age (74 ± 6 versus 71 ± 6; $t = 1.8$, $p = .07$) did not significantly differ between the groups.

Psychometric Properties of the Word and Picture Versions of the Spanish FCSRT

Comparison of the properties of the Word and Picture versions of the Spanish FCSRT to differentiate mild AD patients from cognitively healthy controls. The FR and TR scores of the Word and Picture versions of the Spanish FCSRT for different stages of AD

Table 1. Comparison of demographic characteristics and neuropsychological evaluations between patients and controls

CDR algorithm	Controls	Mild AD patients		Total
	0 (A)	0.5 (B)	1 (C)	
Number of subjects	42	27	23	50
Years of age	71 ± 6	73 ± 5	74 ± 7	74 ± 6
Years of education	13 ± 4	11 ± 5	12 ± 5	12 ± 5
% (*N*) of females	71 (30)	59 (16)	52 (12)	56 (28)
WO_Free Recall	28 ± 6$^{(BC)}$	10 ± 6$^{(C)}$	5 ± 4	7 ± 6
WO_Total Recall	45 ± 3$^{(BC)}$	26 ± 11$^{(C)}$	18 ± 11	22 ± 12
PI_Free Recall	33 ± 5$^{(BC)}$	15 ± 9$^{(C)}$	8 ± 6	12 ± 9
PI_Total Recall	47 ± 3$^{(BC)}$	37 ± 8$^{(C)}$	28 ± 13	33 ± 11
CDR sum of boxes	0.0 ± 0	3.20 ± 1.02	6.26 ± 1.23	4.61 ± 1.9
ACE-R Total score	93 ± 5$^{(BC)}$	72 ± 12$^{(C)}$	60 ± 13	67 ± 14
MMSE Total score	28 ± 2$^{(BC)}$	23 ± 4$^{(C)}$	20 ± 4	22 ± 4
P Phonetic Fluency	16 ± 5$^{(BC)}$	11 ± 5	10 ± 5	11 ± 5
Category Fluency	19 ± 6$^{(BC)}$	12 ± 6	9 ± 4	10 ± 5
FAB Total score	16 ± 2$^{(BC)}$	13 ± 3	12 ± 3	12 ± 3
ACE-R Memory score	23 ± 2$^{(BC)}$	14 ± 5$^{(C)}$	10 ± 4	12 ± 5

Notes: Mean ± standard deviations are shown for continuous variables, and % (number) for gender distribution. The comparisons between cognitive groups are made with ANOVA and χ^2.

A = Controls CDR score = 0; B = Mild AD patients, CDR score = 0.5; BC = score larger than those of patients with mild AD at CDR = 0.5 and 1; C = score larger than those of patients with mild AD at CDR = 1.

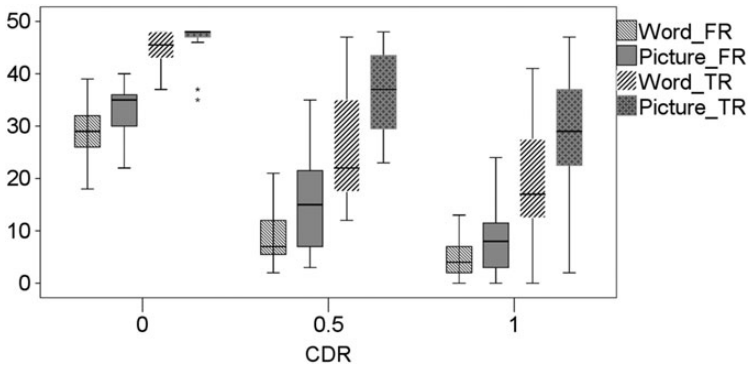

Fig. 1. Box plots comparing the FR and TR scores of the Word and Picture versions of the Spanish FCSRT between CDR score groups. Bars show medians ± 25–75 CI.

Table 2. Effect sizes and confidence intervals for comparisons between patients with mild AD and cognitively healthy elderly controls concerning Free Recall and Total Recall of the FCSRT Word and Picture versions

Variables		Comparison mild AD versus Controls			
		d	*r*	95% CI	
				Lower	Upper
FCSRT Word Version	Free Recall	3.54	.87	−0.34	2.08
	Total Recall	2.56	.79	−1.05	2.62
FCSRT Picture Version	Free Recall	2.85	.82	−0.69	2.32
	Total Recall	1.69	.64	−1.05	2.33

Note: d = Cohen's d coefficient. r = Pearson's r coefficient.

patients and controls are compared in Table 1 and Fig. 1. Age and years of education were considered covariates in this analysis. Significant differences in the FR and TR for both versions of the FCSRT were observed between the groups: WO_FR ($F = 76.4$, $p < .001$); WO_TR ($F = 44.8$, $p < .001$); PI_FR ($F = 58.7$, $p < .001$); PI_TR ($F = 27.2$, $p < .001$). The scores of the control subjects were significantly higher than those of the patients with mild AD across both versions of the FCSRT (CDR = 0.5 and CDR = 1). The scores of the patients with CDR = 0.5 were also higher than those of the patients with CDR = 1.

Effect sizes of group on the FR and TR of the Word and Picture versions of the Spanish FCSRT. Diagnosis had a large effect on the FR (Cohen's d of 3.84 and 2.85, and Pearson's r of .87 and .82 for the verbal and visual versions of the FCSRT, respectively). The corresponding effect of diagnosis on the TR was moderate (Cohen's d of 2.56 and 1.69 and Pearson's of .79 and .64 for the verbal and visual versions of the FCSRT, respectively) (Table 2).

Diagnostic utility of the FR and TR of the Word and Picture versions of the Spanish FCSRT. ROC curves are shown in Fig. 2, whereas the AUCs are compared in Table 3A. The FR and TR for both versions had excellent diagnostic utility with AUCs >0.9 and did not significantly differ between both versions of the test, although the AUC for the PI_TR tended to be slightly smaller than that for the PI_FR and WO_TR: (PI_FR versus PI_TR, z = 1.48); (WO_TR versus PI_TR, $z = 1.49$); (WO_FR versus WO_TR, $z = 0.48$); (WO_FR versus PI_FR, $z = 0.61$).

The COP and diagnostic accuracy in discriminating the mild AD patients from the controls across the FR and TR of both versions of the FCSRT are shown in Table 3B. When selecting the most appropriate COP, specificity was prioritized over sensitivity. Likewise, when specificity was fixed at ≥0.9, the sensitivity was good (>85%) for the WO_TR and PI_TR and excellent (>90%) for the WO_FR and PI_FR. Accordingly, the McNemar test did not identify any differences in the sensitivity of the different measure-version combinations.

Convergent validity of the Word and Picture versions of the Spanish FCSRT. The convergent validity of the Word and Picture versions of the Spanish FCSRT is shown in Table 4. Only the FR scores of the Word version minimally correlated with age (WO_FR $r = -.21$, $p = .04$). Overall, none of the memory measures, neither the FR nor TR of either version of the FCSRT, correlated with years of education. Scores on both versions of the test were highly correlated with measures of global cognitive status

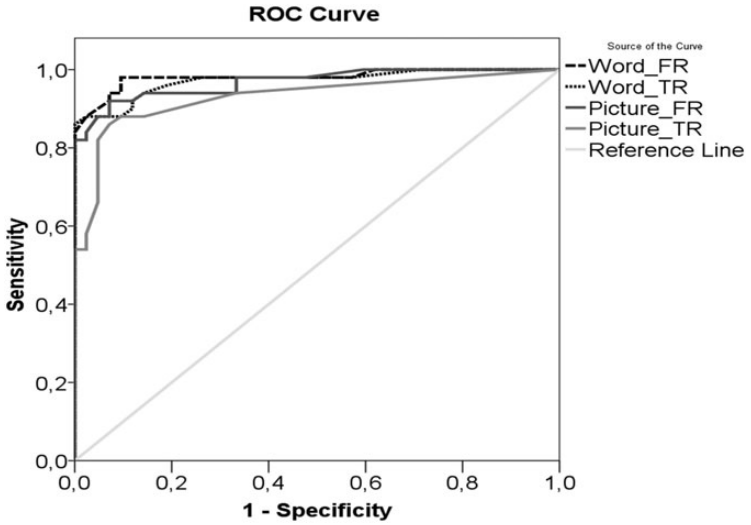

Fig. 2. Comparison of ROC curves for the FR and TR of both versions of the FCSRT for discriminating mild AD patients from cognitively healthy controls.

Table 3. Diagnostic utility of the FCSRT for discriminating mild AD patients from controls

FCSRT	A				B		
	AUC	*SE*	95% CI		COP	Sensitivity	Specificity
Word_ FR	0.980	0.013	0.956	1.000	≤22	0.98	0.91
Word _TR	0.972	0.015	0.938	1.000	≤39	0.88	0.95
Picture _FR	0.969	0.016	0.938	0.999	≤26	0.92	0.93
Picture_TR	0.930	0.028	0.876	0.984	≤46	0.88	0.91

Note: A, area under the curve. B, cut-off point, sensitivity and specificity.

Table 4. Pearson correlations between the FCSRT, demographic, and neuropsychological test data

FCSRT		Age	Years of education	CDR	Total ACE-R	MMSE	ACE-R memory	FAB	Phonetic fluency	Category fluency
Word_FR	*r*	−.207*	.203	−.842**	.820**	.718**	.865**	.612**	.492**	.712**
	p	.048	.053	.000	.000	.000	.000	.000	.000	.000
Word_TR	*r*	−.074	.167	−.796**	.843**	.766**	.883**	.559**	.435**	.660**
	p	.484	.115	.000	.000	.000	.000	.000	.000	.000
Picture_FR	*r*	−.118	.190	−.829**	.849**	.784**	.902**	.606**	.424**	.700**
	p	.263	.071	.000	.000	.000	.000	.000	.000	.000
Picture_TR	*r*	.086	.174	−.704**	.804**	.822**	.811**	.611**	.372**	.608**
	p	.415	.099	.000	.000	.000	.000	.000	.000	.000

Note: *$p < 0.01$; **$p < 0.001$.

(CDR, ACE-R, and MMSE) and episodic memory (scores for the episodic memory domain of the ACE-R) and were acceptably correlated with measures of executive function (FAB and verbal fluencies). As a consequence, the FR and TR of both versions demonstrated good convergent validity.

Internal consistency of the Word and Picture versions of the Spanish FCSRT. Cronbach's α for the Word version of the Spanish FCSRT was 0.85, whereas it was 0.82 for the Picture version. These estimations were calculated by considering three items, i.e., IR, FR, and TR. Hence, the internal consistency was good for both versions of the test.

Comparison of the Results of the Word and Picture Versions of the Spanish FCSRT Across the Same Cohort of Mild AD Patients and Cognitively Healthy Controls

Score comparisons and correlations between the memory measures of both versions of the test across the same population. The FR, TR, and index of sensitivity to cueing of both versions of the FCSRT were highly correlated across the entire sample of this

study. All of the scores for the Picture version of the FCSRT were significantly higher than those for the Word version. Furthermore, these measures were only correlated between the different versions of the test in the mild AD group, not in the control group. In contrast, the FR scores for both versions of the test were moderately correlated in the control group ($r = .59$, $p < .001$), whereas the TR scores and the index of sensitivity to cueing for the two versions were not correlated in this group. The Picture version continued to yield higher scores in the mild AD and control groups.

Comparison of individual FR and TR scores for the Word and Picture versions of the Spanish FCSRT). For FR, 37 of the 50 patients with mild AD (74%) and 37 of the 42 cognitively healthy controls (88%) included in this study performed better on the Picture version than on the Word version of the test. Seven of the patients with mild AD (14%) and none of the control participants exhibited similar performances on both versions of the test, while six mild AD patients and five control subjects (12%) performed better on the Word version than on the Picture version ($\chi^2 = 6.9$, $p = .04$). For TR, 46 of the 50 patients with mild AD (92%) and 28 of the 42 controls (67%) performed better on the Picture version than on the Word version of the test, and 1 patient with mild AD (2%) and 10 controls (23%) exhibited similar performances on both versions. Conversely, three patients with mild AD (6%) and four controls (10%) performed better on the Word version ($\chi^2 = 11.3$, $p = .004$). Notably, 28 controls (66%) achieved the maximum score on the Picture version, whereas 12 controls (29%) did it on the Word version. Three patients with mild AD (5%) achieved the maximum on the Picture version, but none of the mild AD patients achieved the maximum score on the Word version.

Overall, these results indicate that in both groups of subjects, the visual cues were more easily recalled than the verbal cues during both FR and cued recall. Additionally, compared with the control subjects, a greater number of mild AD patients exhibited improved cued recall performance with the use of visual instead of verbal cues, although the FR scores of the groups did not differ. This finding suggests that picture cues might improve encoding to a similar extent in both groups as well as facilitate retrieval (cued recall), especially in the AD group. This effect may not have been evident in the control group due to a ceiling effect in the TR scores on both versions of the test.

Discussion

This study addressed several important issues. First, the FR and TR of the Word and Picture versions of the Spanish FCSRT exhibited appropriate and similar discriminant validity for distinguishing between mild AD patients and cognitively healthy elderly controls. Second, both versions of the FCSRT exhibited appropriate convergent validity and reliability. Third, the scores obtained with the FCSRT across the same cohort of mild AD patients and controls varied in accordance with the type of stimuli: verbal (words) or visual (pictures) cues. Specifically, all of the scores for the Picture version were higher than those for the Word version. Fourth, the memory processes for the verbal (words) and visual (pictures) cues differed between the mild AD patients and the controls, as demonstrated by the greater benefit of pictorial cues during cued recall in the AD patients than in the controls.

Psychometric Properties of the Word and Picture Versions of the Spanish FCSRT

In this study, the Word and Picture versions of the Spanish FCSRT showed excellent capacity to discriminate mild AD patients from cognitively healthy elderly controls. This finding was supported by the medium to large effect of group (diagnosis) on the memory measures and the acceptable AUCs encountered, where the AUCs for the TR and FR for each version of the test were larger than 0.9 and did not differ significantly. Additionally, both versions of the test showed good to excellent diagnostic utility, with sensitivities >90% for the FR and >85% for the TR and specificities >90% for both the FR and TR (Table 3). To estimate the best COP for the FR and TR of the Word and Picture versions of the Spanish FCSRT, specificity was optimized instead of sensitivity because the possibility of making a false-positive diagnosis might result in larger economical costs and a greater emotional burden for the relatives of the patient than making a false-negative diagnosis (Grober et al., 2010). In other studies of the discriminant validity of the FCSRT for differentiating dementia patients from controls, the most appropriate memory measure varied according to the population studied, the administration procedure of the task, and/or the version of the FCSRT that was used (Grober et al., 2010), as will be discussed below.

Influence of the Population Studied on the Scores of the FCSRT

In studies using the Word version of the FCSRT in memory clinic settings, TR was better than FR in discriminating mild AD patients from controls and in predicting the progression of amnestic MCI patients to dementia syndromes (Sarazin et al., 2007; Tounsi et al., 1999), although in a community cohort the FR outperformed the TR (Auriacombe et al., 2010). In studies in

which the Picture version of the FCSRT was used, FR outperformed TR in detecting mild dementia and predicting future dementia in primary care populations (Grober et al., 2010); however, the TR of the Picture version may be superior in identifying patients with moderate AD in memory clinic settings (Grober et al., 1988; Petersen et al., 1994). These results all suggest that FR, of either version, is a more suitable parameter to use to detect dementia in settings with a higher prevalence of non-AD dementias, because the associated TR significantly improves upon the introduction of semantic cues in these dementias, making it less sensitive for these cases. Non-AD dementias frequently present with memory disorders associated with the dysfunction of non-hippocampal regions. Therefore, the performance of these patients on the FCSRT improves with cues, suggesting a detriment of retrieval rather than storage. Conversely, the TR is a more specific memory measure in settings with a higher prevalence of amnestic MCI or AD patients (Sarazin et al., 2007; Tounsi et al., 1999). Accordingly, because this study included a convenience sample of patients who met the criteria for mild AD (McKhann et al., 2011) (CDR = 05 and 1) and because all other causes of dementia were excluded, TR should have identified more cases of AD dementia than FR as it was suggested in the aforementioned studies. All the same, there were no notable differences in the discriminant validity of the FR and TR of either version of the test for detecting mild AD in the present study. In fact, FR in the Word version predicted the risk for mild AD dementia slightly better than the other memory measures (Figs 1 and 2, Tables 2 and 3). These results may be due to the high level of education and lack of memory complaints of the control group used in this study, which make them "super controls" that are less susceptible to problems of encoding or retrieval than "normal" elderly persons. Alternatively, the inclusion of very mild AD patients, for which FR may be more sensitive than TR (Grober et al., 2010), may be responsible for the results obtained here. Finally, as discussed below, the mild AD patients benefitted more from the picture cues than the controls (Fig. 3 and Table 5), especially during cued recall. Consequently, TR for the Picture version could be less sensitive to diagnosis of AD dementia than the other memory measures.

Influence of the Administration Procedure on the Scores of the FCSRT

The results in this study were similar to the findings of studies that, on the one hand, also used the same procedure of administration for the FCSRT introducing the IR within a study phase to control an adequate encoding and, on the other, included sample with similar characteristics to the cohort used here. For instance, Auriacombe and colleagues (2010) using the Word version of the task and Grober and colleagues (2010) using its Picture version reported results for controls that resemble the findings obtained in the present study. On the other hand, Sarazin and collaborators (2010) presented data that match with the results shown here when the Word version was administered to AD patients. Conversely, other analyses conducted with the Word version of the FCSRT in a Portuguese study (Lemos et al., 2014) and a Spanish cohort (Peña-Cassnova et al., 2009) suggested much lower scores for controls compared with those reported here. This difference might be explained by considering that, unlike this study and others (Auriacombe et al., 2010; Grober et al., 2010; Sarazin, *et al.*, 2010), they did not include the IR and test phase prior to the study phase, and the IR improves recall since it permits retrieval practice when the items are still in working memory, which might facilitate the encoding of items into medial temporal areas (Grober et al., 2010; Swerdlow & Jicha, 2012).

Comparison of the Memory Scores Obtained Using the Word and Picture Versions of the FCSRT Across the Same Cohort of Mild AD Patients and Controls

The Word and Picture versions of the FCSRT were compared across the same population of dementia patients. This comparison demonstrated that all of the memory scores for the Picture version of the FCSRT were higher than those for the Word version (Table 5, Figs 1 and 3). This result corroborates the findings of Zimmerman and colleagues (2015), who compared both versions of the task in a sample of older adults without cognitive impairment. Even though the scores of the control group in the present study were very similar to those obtained by Zimmerman and colleagues (2015), it should be noticed that the Word version scores in this study were slightly higher than those reported in their study. This difference might be explained due to the older age of the cohort analyzed by Zimmerman and colleagues (2015), which is in agreement with the negative correlation found here between the FR of the Word version and age (Table 4).

These findings confirm that the two versions of the task are not equivalent, and that the COP should be interpreted based on the version employed.

Studies using other memory assessments in children, normal adults, older adults, and dementia patients have also reported that picture cues are recalled more easily than verbal or word cues (Cherry et al., 2008). In fact, amnestic patients receive a greater benefit from the use of pictures than healthy older adults (Ally 2012).

The performance improvement observed in tests of memory that utilized pictures rather than words as cues might be related to the "dual coding theory," which proposes that pictures are more beneficial than words because pictures evoke both a verbal and an image code, whereas words only trigger an abstract verbal code (Paivio, 1995). Other theories suggest that the sensory codes for pictures are richer than those for words, leading to a more differentiated representation that is less susceptible to interference.

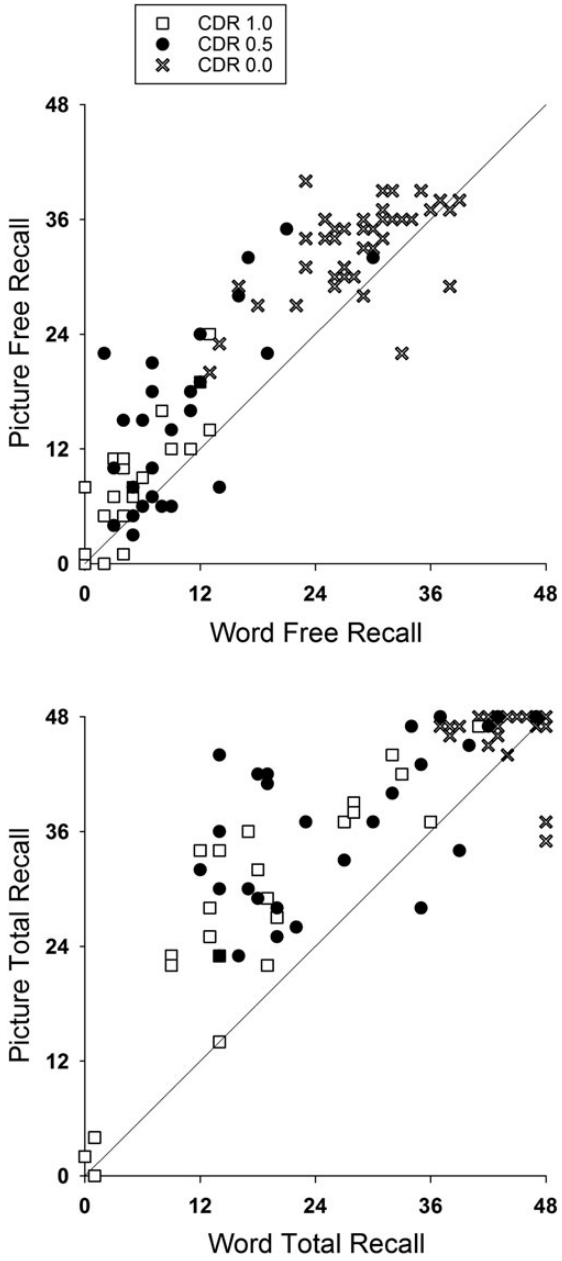

Fig. 3. Scatter plot comparing individual performance on the Picture and Word versions of the Spanish FCSRT between CDR groups.

Table 5. Comparison of the two versions of the FCSRT

Subjects	Memory measure	Picture version	Word version	Paired *t*-test				Pearson correlation	
				Mean	*SD*	*t*	*p*	*r*	*p*
Global group	Free Recall	28 ± 13	17 ± 12	4.7	5.2	8.5	<.001	.9	<.001
	Total Recall	39 ± 11	33 ± 14	6.5	7.5	8.2	<.001	.85	<.001
	Sensitivity to cueing index	77 ± 25	60 ± 30	16.2	19	8.1	<.001	.78	<.001
Ad patients	Free Recall	12 ± 9	8 ± 6	4.5	5.4	5.9	<.001	.79	<.001
	Total Recall	33 ± 11	22 ± 12	10.1	7.6	9.3	<.001	.78	<.001
	Sensitivity to cueing index	61 ± 24	39 ± 23	22.4	16.2	9.6	<.001	.76	<.001
Control group	Free Recall	33 ± 5	28 ± 6	4.8	5	6.2	<.001	.59	<.001
	Total Recall	47 ± 3	45 ± 3	2.1	4.3	3.1	.003	−.07	.67
	Sensitivity to cueing index	94 ± 13	85 ± 16	9.3	19	3.0	.004	.027	.8

Pictures may also elicit deeper and more elaborate conceptual processing than words, thus, this difference could produce the picture superiority effect (Ally, 2012).

With respect to the differences in the memory processing of word versus picture cues between the mild AD patients and controls, this study showed that a greater proportion of the AD patients exhibited better performance in PI_TR than in WO_TR, indicating that the greater sensitivity to picture compared with verbal cues was more important in the memory retrieval process of the AD group than in the control group. These differences could be related to a ceiling effect in TR performance in the control group, which in almost 30% achieved the maximal score in both versions of the test, reducing the possibility of identifying significant differences between them; and/or a specific increase in memory performance with the use of picture cues in patients with hippocampal-dependent memory impairment due to AD, because pictures may activate larger areas of extra-hippocampal cortex than word cues (Ally, 2012).

These issues are especially relevant for issues such as (i) the diagnostic utility, which could explain the poorer sensitivity of the PI_TR in discriminating mild AD patients from controls; (ii) the nature of the stimuli used in memory assessment, which must be considered in addition to the studied population when interpreting the COP of any study that uses the FCSRT, as discussed above; and (iii) cognitive stimulation, in that the use of images rather than words may be preferable in patients with memory impairment (Ally, 2012; Embree, Budson, & Ally, 2012)

Because a sample of convenience was used in this study, the conclusions drawn based on this population should be generalized to other populations with caution. The findings presented here should be considered independently of those of studies of subjects with diverse educational, social, and ethnic backgrounds as well as distinct clinical diagnoses.

Several final conclusions can be drawn from this study. First, the Word and Picture versions of the Spanish FCSRT had similar and excellent discriminant validity, diagnostic utility, convergent validity, and reliability. Second, the two versions could yield different results across the same population of mild AD patients and controls. Third, picture cues were recalled more easily than word cues, which resulted in higher scores in the Picture version than in the Word version of the test. Fourth, compared with that of the controls, the AD patients showed greater improvement in cued recall with the picture cues than the word cues, possibly due to the stimulation of preserved extra-hippocampal areas in these patients.

Funding

This work was supported by Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) (1140423 to C.D., P.F and A.S.) of the Comisión Nacional de Investigación Científica y Tecnológica, Gobierno de Chile, the FONDAP grant 15150012 (to A.S.) and the Basal Funds for Excellence Centers from the Associative Research Program (FB 0003 to A.S.) of the Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), Gobierno de Chile.

Conflict of Interest

None declared.

Acknowledgements

The authors thank Dr Juan Luis Moya, who designed all the drawings of the Picture version of the Free and Cued Selective Reminding Test used in this study, Dr Paul Delano that helped out with the figures, and Mr Kimon Butzbach who assisted in the proofreading of the manuscript as an English native speaker.

References

Ally, B. A. (2012). Using pictures and words to understand recognition memory deterioration in amnestic mild cognitive impairment and Alzheimer's disease: A review. *Current Neurology and Neuroscience Reports*, *12*, 687–694.

Auriacombe, S., Helmer, C., Amieva, H., Berr, C., Dubois, B., & Dartigues, J. F. (2010). Validity of the free and cued selective reminding test in predicting dementia: The 3C study. *Neurology*, *22*, 1760–1767.

Canning, S. J., Leach, L., Stuss, D., Ngo, L., & Black, S. E. (2004). Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia. *Neurology*, *62*, 556–562.

Cherry, K. E., Hawley, K. S., Jackson, E. M., Volaufova, J., Su, L. J., & Jazwinski, S. M. (2008). Pictorial superiority effects in oldest-old people. *Memory*, *16*, 728–741.

Cohen, J. D. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Di Paola, M., Macaluso, E., Carlesimo, G. A., Tomaiuolo, F., Worsley, K. J., Fadda, L., et al. (2007). Episodic memory impairment in patients with Alzheimer's disease is correlated with entorhinal cortex atrophy. *Journal of Neurology*, *254*, 774–781.

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., et al. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *The Lancet. Neurology*, *6*, 734–746.

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., et al. (2014). Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *The Lancet. Neurology*, *13*, 614–629.

Dubois, B., Slachevsky, A., Litvan, I., & Pillon, B. (2000). The FAB: A Frontal Assessment Battery at bedside. *Neurology*, *55*, 1621–1626.

Embree, L. M., Budson, A. E., & Ally, B. A. (2012). Memorial familiarity remains intact for pictures but not for words in amnestic mild cognitive impairment. *Neuropsychologia*, *50*, 2333–2340.

Fargo, X., & Bleiler, X. (2014). Alzheimer's Association 2014. Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *10 (2)*, e47–e92.

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.

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.

Grober, E., Hall, C., Sanders, A., & Lipton, R. B. (2008). Free and cued selective reminding distinguishes Alzheimer's disease from vascular dementia. *Journal of the American Geriatrics Society*, *56*, 944–946.

Grober, E., Lipton, R. B., Hall, C., & Crystal, H. (2000). Memory impairment on free and cued selective reminding predicts dementia. *Neurology*, *54*, 827–832.

Grober, E., Sanders, A. E., Hall, C., & Lipton, R. B. (2010). Free and cued selective reminding identifies very mild dementia in primary care. *Alzheimer's Disease and Associated Disorders*, *24*, 284–290.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*, 839–843.

Lemos, R., Duro, D., Simões, M. R., & Santana, I. (2014). The Free and Cued Selective Reminding Test distinguishes frontotemporal dementia from Alzheimer's disease. *Archives of Clinical Neuropsychology*, *29*, 670–679.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*, 263–269.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*, 153–157.

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): A brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, *21*, 1078–1085.

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, *43*, 2412–2414.

Muñoz-Neira, C., Henríquez, C. F., Ihnen, J. J., Sánchez, C. M., Flores, M. P., & Slachevsky, C. A. (2012). Psychometric properties and diagnostic usefulness of the Addenbrooke's Cognitive Examination-Revised in a Chilean elderly sample. *Revista medica de Chile*, *140*, 1006–1013.

Paivio, A. (1995). Imagery and memory. In M. S. Gazzaniga (Ed.), The cognitive neurosciences (pp. 977–986). Cambridge, MA: MIT Press.

Peña-Casanova, J., Gramunt-Fombuena, N., Quiñones-Ubeda, S., Sánchez-Benavides, G., Aguilar, M., Badenes, D., et al. (2009). Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for the Rey-Osterrieth complex figure (copy and memory), and free and cued selective reminding test. *Archives of Clinical Neuropsychology*, *24 (4)*, 371–393.

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.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimers & Dementia*, *9*, 63–75. e2.

Quiroga, P., Albala, C., & Klaasen, G. (2004). Validación de un test de tamizaje para el diagnóstico de demencia asociada a edad, en Chile [Validation of a screening text for age associated cognitive impairment, in Chile]. *Revista medica de Chile*, *132*, 467–478.

Sarazin, M., Berr, C., De Rotrou, J., Fabrigoule, C., Pasquier, F., Legrain, S., et al. (2007). Amnestic syndrome of the medial temporal type identifies prodromal AD: A longitudinal study. *Neurology*, *69*, 1859–1867.

Sarazin, M., Chauviré, V., Gerardin, E., Colliot, O., Kinkingnéhun, S., de Souza, L. C., et al. (2010). The amnestic syndrome of hippocampal type in Alzheimer's disease: An MRI study. *Journal of Alzheimer's Disease*, *22*, 285–294.

Swerdlow, R. H., & Jicha, G. A. (2012). Alzheimer disease: can the exam predict the pathology? *Neurology*, *78 (6)*, 374–375.

Tounsi, H., Deweer, B., Ergis, A. M., Van der Linden, M., Pillon, B., Michon, A., et al. (1999). Sensitivity to semantic cuing: an index of episodic memory dysfunction in early Alzheimer disease *Alzheimer Disease & Associated Disorders*, *13 (1)*, 38–46.

Zimmerman, M. E., Katz, M. J., Wang, C., Burns, L. C., Berman, R. M., Derby, C. A., et al. (2015). Comparison of "Word" vs. "Picture" version of the Free and Cued Selective Reminding Test (FCSRT) in older adults. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *1*, 94–100.