Psychology Science Quarterly, Volume 51, 2009 (3), pp. 266-282

The Free and Cued Selective Reminding Test: evidence of psychometric adequacy

ELLEN GROBER[1], KATJA OCEPEK-WELIKSON[2] & JEANNE A. TERESI[3]

Abstract

These analyses examine the psychometric properties of the Free and Cued Selective Reminding Test with Immediate Recall (FCSRT-IR). FCSRT-IR is a measure of memory under conditions that control attention and cognitive processing in order to obtain an assessment of memory unconfounded by normal age-related changes in cognition. FCSRT-IR performance has been associated with preclinical and early dementia in several longitudinal epidemiological studies. Factor and item response theory analyses were applied to FCSRT-IR data from patients at a geriatric primary care center who had independently established clinical diagnoses. The results provide supporting evidence for the psychometric adequacy of the FCSR-IR in terms of reliability, essential (sufficient) unidimensionality, information across the continuum of memory disability/ability, and classification accuracy. The psychometric adequacy of the FCSRT-IR adds further validity to its use as a case finding strategy for dementia.

Key words: Free and Cued Selective Reminding Test; item response theory; factor analyses; early dementia

[1] Correspondence concerning this article should be addressed to: Ellen Grober, PhD, Albert Einstein College of Medicine and Montefiore Medical Center, 1300 Morris Park Avenue, Flor. 125, Bronx, New York 10461, USA; email: egrober@montefiore.org

[2] Research Division, Hebrew Home at Riverdale

[3] Columbia University Stroud Center and Faculty of Medicine and New York State Psychiatric Institute, New York

Introduction

Memory testing is critical to identifying dementia because current criteria for the diagnosis of any dementia, irrespective of subtype, require that memory impairment be present (American Psychiatric Association, 1994; McKhann, Drachman, Folstein, et al, 1984). Impaired memory is one of the earliest manifestations of Alzheimer's Disease (AD), the most common form of late-life dementia. The availability of biomarkers of AD, including structural and molecular neuroimaging changes and cerebrospinal fluid, and an interest in measuring memory functioning in mild cognitive impairment (MCI: Petersen, 2004), the transitional state between normal cognition and AD, have led to the development of revised criteria for AD (Dubois et al., 2007). These criteria are "centered on a clinical core of early and significant episodic memory impairment" defined by controlled learning procedures modeled on those that have been developed in the Einstein Aging Study (EAS) over the past 20 years (Buschke, 1984; Buschke, Sliwinski, Kuslansky, & Lipton, 1995; Grober & Buschke, 1987; Grober et al, 1988; Grober, Lipton, Hall, & Crystal, 2000; Grober et al., 2008b; Lipton et al., 2003). By controlling the conditions of learning, a measure of memory is obtained that is not confounded by normal age-related changes in cognition.

Controlled learning begins with a study phase in which the patients identify words or pictures (grapes, vest) in response to category cues (fruit, clothing) that are used in the test phase to prompt recall of items not retrieved by free recall. Controlled learning induces specific semantic processing for effective encoding of the to-be remembered items and it also provides for maximum cued recall because the same cues used to study the items initially are used in the test phase to retrieve items that were not retrieved by free recall. Testing memory using controlled learning conditions was posited to distinguish the genuine deficits in encoding and storage that characterize AD from the memory deficits associated with normal aging that are secondary to impaired attention, inefficient information processing, or ineffective retrieval operations (Grober & Buschke, 1987). Retrieval deficits that occur in many healthy elderly individuals are remediated by controlled learning procedures (Buschke et al., 1995; Grober, Merling, Heimlich, & Lipton, 1997). In patients with dementia, these procedures have very modest benefits. As a consequence, controlled learning procedures should increase differences produced by normal aging and dementia, thereby improving discriminative validity. Because memory is the earliest manifestation of AD and other causes of acquired memory impairment in the elderly are rare (Cummings & Benson, 1992), we reasoned that the identification of impaired memory under controlled learning conditions should be highly predictive of prevalent and incident AD (Grober et al, 1988; Grober et al., 2008a; Grober, Lipton, Hall, & Crystal, 2000).

The 16-item version of controlled learning is called the Free and Cued Selective Reminding Test with Immediate Recall (FCSR-IR) (Grober & Buschke, 1987). It has been used in several other longitudinal aging studies in the North America and Europe to identify preclinical and early dementia (Grober et al., 2008a; Lindenberger & Reischies, 1999; Petersen, Smith, Ivnik, Kokmen, & Tangalos, 1994; Petersen et al., 1995; Tuokko, Kristjansson, & Miller, 1995), and is used to determine inclusion and trigger clinical evaluations in the Alzheimer's Disease Cooperative Study Instrumentation Protocol (Ferris et al., 2006). Most recently, FCSRT-IR performance distinguished patients with mild cognitive impairment who converted to AD from MCI nonconverters at levels that led the authors to define prodromal AD or the amnestic syndrome of the medial temporal lobe by FCSRT-IR performance

(Sarazin et al., 2007). Further validation of FCSRT-IR performance as an indicator of early AD comes from correlations with abnormalities in structural and functional imaging and with neurofibrillary lesions in parahippocampal regions that are the earliest targets of AD pathology (Grober et al., 1999; Habert et al., 2009; Lekeu et al., 2003; Zimmerman et al., 2008).

The purpose of these analyses were to examine: a) the internal consistency of the three test forms of FCSRT-IR, each consisting of 16 unique category-items pairs (fruit-grapes, bird-owl), b) the dimensionality of the FCSRT-IR using exploratory and confirmatory factor analysis, c) the information provided by the items and test, using item response theory, and d) the concurrent criterion validity by deriving receiver operating characteristic (ROC) curves separately for each FCSRT-IR form, and then for the combined forms. IRT-based methods have been used previously in the evaluation of cognitive tests (Crane, Gibbons, Jolley, & van Belle, 2006; Jones, 2006; Morales, Flowers, Gutiérrez, Kleinman, & Teresi, 2006; Mungas & Reed, 2000; Orlando-Edelen, Thissen, Teresi, Kleinman, & Ocepek-Welikson, 2006; Teresi et al., 1995; Teresi, Kleinman, & Ocepek-Welikson, 2000; Teresi, Kleinman, Ocepek-Welikson et al., 2000). Although the FCSRT-IR is a widely used memory test, it has never been analyzed with respect to IRT.

Materials and methods

Subjects

The subjects who provided FCSRT-IR data analyzed here were patients from the Geriatric Ambulatory Practice (GAP), an urban academic primary care practice staffed by geriatricians at Montefiore Medical Center in the Bronx, New York. Following procedures approved by the local institutional review board, patients were contacted by phone to determine eligibility and interest. Participants were 65 years or older; described themselves as White, not of Hispanic Origin, Black, not of Hispanic Origin, or Hispanic; provided the name of a family member or friend who had known them for at least 5 years; and had adequate vision and hearing to complete the neuropsychological tests. In addition, participants were required to score ≥ 18 on the Mini Mental State Exam (MMSE: Folstein et al, 1975), the most widely used screening test for cognitive impairment. We adopted a score of 18 or higher for two reasons: we wanted to capture all cases of mild dementia which often are missed in primary care (Callahan et al, 1995); and African American patients who score in the range of 18 to 23 on the MMSE are frequently misclassified as demented (false positives) when they are not independently diagnosed as having a clinical dementia (Bohnstedt et al, 1994). Patients with scores lower than 18 are likely to have a more advanced dementia and would not need to undergo dementia screening. Two illiterate patients with lower scores were included because several errors they made were on language-related items. Participants completed a comprehensive battery of neuropsychological tests and informants were interviewed by phone. Presence or absence of dementia was established by consensus of an expert panel using DSM IV criteria independent of FCSRT-IR scores to avoid diagnostic circularity. The methods are described elsewhere (Grober & Buschke, 1987). The sensitivity and specificity of the different test forms in correctly classifying patients were compared to each other using a previously established FCSRT-IR cut score (Grober et al., 2000).

Table 1:
Sample Demographic Characteristics by Form

Demographic Characteristic	Form A	Form B	Form C
Gender			
Male, n (percent)	38 (18.1%)	29 (15.3%)	27 (14.8%)
Female, n (percent)	172 (81.9%)	160 (84.7%)	155 (85.2%)
Race			
African-American, n (percent)	104 (49.5%)	95 (51.1%)	85 (46.7%)
Caucasian not Hispanic, n (percent)	101 (48.1%)	90 (48.4%)	94 (51.6%)
Hispanic, n (percent)	5 (2.4%)	1 (0.5%)	3 (1.6%)
Age at Interview, mean (stand. dev.)	78.6 (7.3)	80.0 (6.7)	79.2 (7.4)
Years of education, mean (stand. dev.)	12.6 (3.4)	12.4 (3.4)	12.6 (3.4)
MMSE, mean (stand. dev.)	26.6 (3.1)	26.6 (3.3)	26.6 (3.7)
Meets DSM IV Criteria for Dementia, n (percent)	50 (24.6%)	52 (28.1%)	51 (28.3%)
Incident Dementia 2 ½ year follow-up, n (percent)	20 (13.6)	16 (9.9%)	14 (9.0%)

Table 1 presents the demographic characteristics for the patients who provided the FCSRT-IR data by form. As shown, 210 subjects were tested with Form A, 189 with Form B and 182 with Form C. The distribution of gender and race did not differ as a function of form nor did age, education, or MMSE scores. Among the same subjects, 50 (24.6%), 52 (28.1%) and 51 (28.3%) met DSM-IV criteria for dementia at baseline and 20 (13.6%), 16 (9.9%) and 14 (9.0%) were diagnosed with incident dementia at the end of approximately 2 ½ years of follow-up when tested by Form A, Form B and Form C respectively. These differences by form were not statistically different.

Free and Cued Selective Reminding Test with Immediate Recall (FCSRT-IR)

The test begins with a study phase in which subjects were asked to examine a card containing line drawings of easily recognized objects (e.g., grapes) for an item that goes with a unique category cue (e.g., fruit). The 16 items to be learned were presented four at a time on a card, one picture in each quadrant. The subject was asked to search each card and point to and name aloud each item (e.g., grapes) after its cue (fruit) was aurally presented. After all four items were identified correctly, the card was removed, and immediate cued recall of just those four items was tested by presenting the cues again. The subject was reminded of any item he or she failed to retrieve by presenting the cue and the item together (e.g., the vehicle was a train).

Once immediate recall for a group of four items was completed, the next set of items was presented for study. The study phase was followed by the test phase that consisted of three recall trials, each preceded by 20 seconds of subjects counting backward to prevent recall from short-term memory. Each recall trial consisted of two parts. First, each subject had up to two minutes to freely recall as many items as possible. Next, aurally presented category cues were provided for items not retrieved by free recall. If subjects failed to retrieve the item with the category cue, they were reminded by presenting the cue and the item together. The sum of free and cued recall is total recall. The current analyses focused on free recall.

There were forms of the FCSRT-IR (A,B,& C) each containing 16 simple line drawings of basic level items (e.g., vest, chimney) that belong to different semantic categories (article of clothing, part of a building). The particular items were selected from category norms (Battig & Montague, 1969) because they were unlikely responses when persons were asked to generate the name of an item from the semantic category to which the item belonged, yet the item was still familiar enough to be named readily. This was necessary to avoid correct responses due to guessing when the category cues were presented during the test phases to prompt recall (Grober, Gitlin, Bang, & Buschke, 1992).

For this analysis, the items were created by summing across the three trials: if a subject recalled an item spontaneously in all three trials, the item score was three; two if recalled on two trials, one if recalled on one trial, or zero for no recall of that item on any trial. A sum score was produced as well as an IRT-based ability score (theta). A high score is indicative of higher memory ability.

Methods of analyses

The following analyses were conducted using the two parameter graded (for polytomous, ordered response category) item response models (Hambleton, Swaminathan, & Rogers, 1991; Lord & Novick, 1968; Lord, 1980; Samejima &, 1969). First steps in the analyses include examination of model assumptions (such as unidimensionality) and model fit.

"*Receiver operating characteristic (ROC) curves:* ROC curves were generated to visual differences in sensitivity and specificity across a range of cut-scores. The diagnostic accuracy of tests can by compared by examining the area under the ROC curve (AUC)."

Tests of model assumptions: exploratory and confirmatory factor analyses

Exploratory factor analyses (EFA), using the method of principal components were completed, followed by a test of scree. Eigenvalues and explained percent of variance for consecutive factors were evaluated to provide supporting evidence to confirm (or refute) the essential unidimensionality of the studied constructs. Because the FCSRT-IR test was assumed to be essentially unidimensional, and the second factor was assumed to be only a nuisance factor, the confirmatory factor analyses (CFA) compared only the one factor solutions to the two factor solutions to test whether the two factor solutions improved the model fit. EFA and CFA were conducted using the weighted least squares mean variance (WLSMV) estimator (weighted least square parameter estimates with standard errors and mean and variance adjusted chi-square test statistics that used a full weight matrix) in MPLUS (Multhen & Multhen, 2004).

Item response theory model: polytomous response model

In this model, ordered responses, x=k and k=1,2,...m, are assumed. The discrimination parameter or slope can be defined as a_i, and difficulty parameters for response k as b_{ik}.

$P(x=k) = P^*(k) - P^*(k+1) = 1 / [1 + \exp[-Da_i(\theta-b_{ik-1})] - 1 / [1 + \exp[-Da_i(\theta-b_{ik})]$, where $P^*(k)$ is the item characteristic curve describing the probability that a response is in category k or higher, for each value of θ (Samejima, 1969; Thissen, 1991, 2001). The model assumes an average discrimination across response categories.

Methods of estimation and software: Marginal maximum likelihood methods of parameter estimation applying the EM algorithm (Bock & Aitken, 1981), available in MULTILOG (Thissen, 1991, 2001), were used to fit polytomous item response models. MULTILOG results were used to calculate a standardized residual measure of goodness-of-fit which is defined as the difference between the observed and expected frequency divided by the square root of the expected frequency for each response pattern associated with a particular level of theta (ability estimate). The standardized residual is distributed approximately normally with mean of 0 and σ^2 of 1.

Information function: The information function for score x is defined as the square of the ratio of the slope of the regression of x on θ to the standard error of x for fixed θ (Lord & Novick, 1968). The standard error of theta can be expressed as: $Se(\theta) = 1 / \sqrt{I(\theta)}$. An item provides more information when the slope is steeper, the standard error is lower and when b (difficulty parameter) is close to the theta being evaluated.

Results

Test of model assumptions

Exploratory Factor Analyses (EFA)

The eigenvalues for the first factors ranged from 6.34 (Form C) to 5.50 (Form A), which was roughly four times that of the second eigenvalues (ranging from 1.10 to 1.34). The first factor accounted for between 34% and 40% of the total explained variance. These results provide evidence of essential unidimensionality for all three Forms (see Figure 1).

Confirmatory Factor Analyses (CFA)

The results of the CFA for the unidimensional model yielded non-significant χ^2's for the Form C indicating the best model fit for the construct (χ^2=58.1,d.f.=58, p=0.47). Although the χ^2 model fit statistics were significant for Forms A and B (Form A: χ^2=76.7, d.f.=58, p=0.05; Form B: χ^2=81.8, d.f.=54, p=0.01), other fit statistics showed an acceptable fit for the unidimensional models, with little improvement evidenced for the two-factor models. For example, test of the two-factor models improved the root mean square residual index of fit by 0.01 from 0.04 to 0.03 for Form A and from 0.05 to 0.04 for Form B. The Comparative Fit Index (CFI) ranged from 0.95 to 1.00 across forms for the one factor solution, and from 0.97 to 1.00 for the two factor solution. The standardized factor loadings ranged from 0.45 to 0.65 for Form A, from 0.40 to 0.72 for Form B and from 0.40 to 0.70 for Form C (see Table 2).

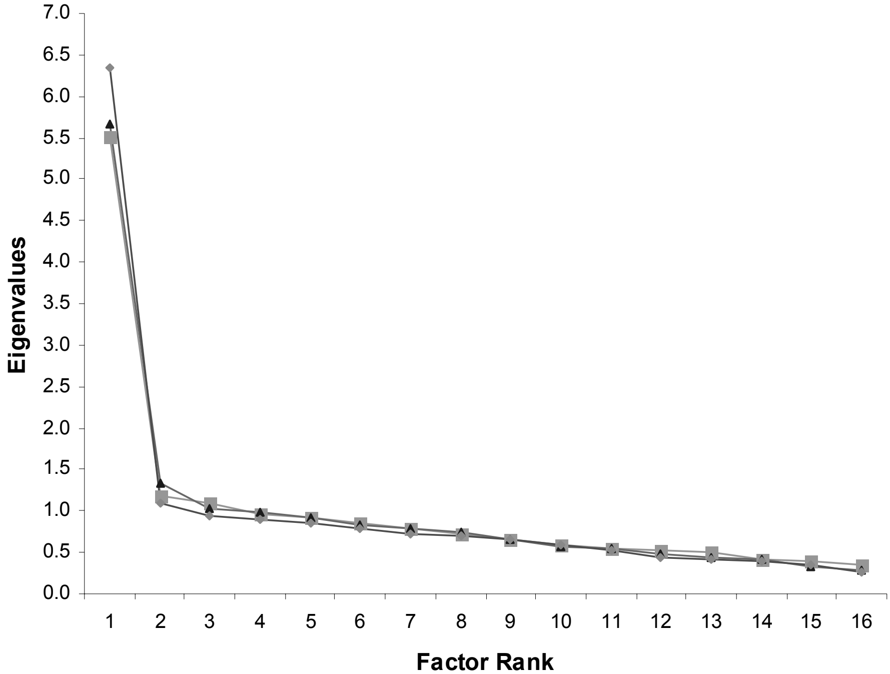

Figure 1:
Eigenvalues Scree Plots for All Forms

IRT-model fit and item parameter estimates

Examination of standardized residuals (not shown) showed that all items fit the model for all forms: ($z < 0.01$ to 0.78) for form A; ($z < 0.01$ to 0.67) for form B; and ($z < 0.01$ to 1.03) for Form C. Shown in Table 3 are the parameter estimates from IRT. Parameters are sorted from the easiest item to the most difficult. The items optimally discriminate at a very wide range of theta/ability spectrum from ability level -3.45 to 2.63 for Form A; -3.67 to 2.96 for Form B; and -2.57 to 3.03 for Form C. The discrimination parameters were adequate, ranging from 0.88 to 1.51, 0.75 to 1.85 and 0.75 to 1.86 for Forms A, B and C, respectively; the standard errors ranged from 0.18 to 0.25 for Form A, 0.19 to 0.28 for Form B and 0.19 to 0.31 for Form C. The range for the difficulty parameter standard errors was from 0.15 to 0.76 for Form A, 0.12 to 0.96 for Form B, and 0.13 to 0.90 for Form C.

IRT item and test information functions

The information functions were examined in order to determine item and total test discrimination ability at different levels of latent (memory) ability. Individual item information functions were examined, and are presented in Figure 2. The highest information of all the

Table 2:

Parameter Estimates and Their Standard Errors From the Unidimensional Confirmatory Factor Analysis Model: FCSRT-IR, Free Recall 16 Item Scales, Form A, Form B and Form C

	Form A Reliability Coefficient Alpha: 0.85			Form B Reliability Coefficient Alpha: 0.86			Form C Reliability Coefficient Alpha: 0.88		
	Item Description	**Mean (Std. Dev.)**	**Standardized Factor Loadings (s.e.)**	**Item Description**	**Mean (Std. Dev.)**	**Standardized Factor Loadings (s.e.)**	**Item Description**	**Mean (Std. Dev.)**	**Standardized Factor Loadings (s.e.)**
FR1	Grapes	1.9 (1.0)	0.57 (0.05)	Spider	1.4 (1.0)	0.72 (0.04)	Bear	1.9 (1.0)	0.70 (0.05)
FR2	Racquet	1.2 (1.0)	0.45 (0.06)	Foot	1.7 (0.9)	0.53 (0.06)	Skate	1.7 (1.0)	0.65 (0.05)
FR3	Owl	1.6 (1.0)	0.65 (0.05)	Pretzel	1.3 (1.0)	0.61 (0.05)	Onion	1.3 (1.0)	0.63 (0.05)
FR4	Desk	1.8 (1.0)	0.63 (0.05)	Wheelchair	1.8 (1.0)	0.63 (0.05)	Broom	1.2 (0.9)	0.59 (0.06)
FR5	Chimney	1.9 (1.0)	0.60 (0.05)	Clouds	2.0 (1.0)	0.63 (0.06)	Volcano	1.7 (1.1)	0.65 (0.05)
FR6	Canoe	1.9 (1.0)	0.62 (0.06)	Anchor	1.8 (1.0)	0.58 (0.06)	Train	1.9 (1.0)	0.65 (0.05)
FR7	Paper Clip	1.3 (1.0)	0.54 (0.05)	Watch	1.4 (1.0)	0.56 (0.06)	Dominoes	1.1 (1.0)	0.52 (0.06)
FR8	Shark	1.8 (1.0)	0.58 (0.05)	Rolling Pin	1.5 (0.9)	0.68 (0.05)	Cake	1.7 (1.1)	0.46 (0.07)
FR9	Ruler	1.5 (1.0)	0.48 (0.06)	Vest	1.7 (1.1)	0.49 (0.06)	Crown	1.8 (1.1)	0.71 (0.04)
FR10	Telescope	1.6 (1.0)	0.60 (0.05)	Balloons	1.7 (1.0)	0.57 (0.06)	Triangle	1.7 (1.0)	0.63 (0.05)
FR11	Cactus	1.5 (1.0)	0.49 (0.06)	Nine	1.5 (1.0)	0.57 (0.05)	Cabin	2.3 (0.9)	0.66 (0.05)
FR12	Rattle	1.7 (1.0)	0.64 (0.04)	Bench	2.1 (0.9)	0.59 (0.07)	Tulip	1.8 (1.0)	0.58 (0.05)
FR13	Steer Wheel	1.5 (1.1)	0.53 (0.06)	Ax	2.1 (0.9)	0.52 (0.07)	Candle	1.8 (0.9)	0.62 (0.05)
FR14	Pitcher	1.5 (0.9)	0.47 (0.06)	Pipe	1.5 (0.9)	0.40 (0.07)	Sword	1.8 (1.0)	0.55 (0.06)
FR15	Razor	1.7 (0.9)	0.49 (0.06)	Wreath	2.0 (0.9)	0.42 (0.08)	Thread	1.4 (0.9)	0.40 (0.07)
FR16	Guitar	1.9 (0.9)	0.45 (0.07)	Basket	1.6 (0.9)	0.41 (0.06)	Toaster	1.7 (0.9)	0.52 (0.06)

Table 3:
MULTILOG IRT Analysis, Item Parameter Estimates Sorted by Item Difficulty from Easiest: FCSRT-IR, Free Recall 16 Item Scales, Form A, Form B and Form C

Form A						Form B						Form C					
Item	Item Name	*a** (s.e.)	*b1*** (s.e.)	*b2* (s.e.)	*b3* (s.e.)	Item	Item Name	*a* (s.e.)	*b1* (s.e.)	*b2* (s.e.)	*b3* (s.e.)	Item	Item Name	*a* (s.e.)	*b1* (s.e.)	*b2* (s.e.)	*b3* (s.e.)
FR16	Guitar	0.88 (0.18)	-3.45 (0.76)	-1.04 (0.30)	1.23 (0.33)	FR15	Wreath	0.75 (0.19)	-3.67 (0.96)	-1.45 (0.44)	0.99 (0.37)	FR11	Cabin	1.48 (0.26)	-2.57 (0.47)	-1.39 (0.25)	0.19 (0.15)
FR15	Razor	0.94 (0.20)	-2.44 (0.54)	-0.51 (0.24)	1.84 (0.41)	FR13	Ax	1.02 (0.20)	-3.20 (0.68)	-1.39 (0.34)	0.68 (0.24)	FR16	Toaster	1.06 (0.21)	-2.50 (0.54)	-0.55 (0.25)	1.52 (0.36)
FR14	Pitcher	0.92 (0.19)	-2.23 (0.50)	-0.08 (0.22)	2.18 (0.50)	FR16	Basket	0.75 (0.20)	-3.16 (0.79)	-0.38 (0.29)	2.52 (0.72)	FR13	Candle	1.33 (0.24)	-2.42 (0.42)	-0.46 (0.19)	1.37 (0.27)
FR1	Grapes	1.18 (0.22)	-2.11 (0.38)	-0.85 (0.23)	0.84 (0.22)	FR14	Pipe	0.75 (0.19)	-2.57 (0.70)	-0.12 (0.28)	2.96 (0.80)	FR14	Sword	1.07 (0.23)	-2.31 (0.49)	-0.56 (0.25)	1.03 (0.27)
FR8	Shark	1.24 (0.22)	-2.07 (0.38)	-0.44 (0.19)	1.02 (0.22)	FR12	Bench	1.28 (0.21)	-2.45 (0.43)	-1.27 (0.26)	0.56 (0.19)	FR15	Thread	0.75 (0.20)	-2.25 (0.66)	0.46 (0.29)	3.03 (0.90)
FR6	Canoe	1.37 (0.21)	-1.92 (0.35)	-0.58 (0.18)	0.63 (0.17)	FR2	Foot	1.12 (0.21)	-2.11 (0.42)	-0.49 (0.21)	1.41 (0.30)	FR12	Tulip	1.25 (0.25)	-1.88 (0.40)	-0.38 (0.20)	1.00 (0.24)
FR5	Chimney	1.33 (0.24)	-1.85 (0.35)	-0.55 (0.19)	0.77 (0.19)	FR5	Clouds	1.27 (0.23)	-1.96 (0.35)	-0.75 (0.21)	0.55 (0.19)	FR8	Cake	0.92 (0.19)	-1.86 (0.45)	-0.29 (0.24)	1.22 (0.35)
FR4	Desk	1.40 (0.24)	-1.80 (0.32)	-0.51 (0.18)	0.97 (0.20)	FR9	Vest	0.87 (0.20)	-1.87 (0.46)	-0.78 (0.29)	1.38 (0.38)	FR6	Train	1.51 (0.25)	-1.59 (0.30)	-0.60 (0.18)	0.78 (0.18)
FR12	Rattle	1.47 (0.24)	-1.80 (0.30)	-0.27 (0.15)	1.00 (0.19)	FR4	Wheelchair	1.34 (0.24)	-1.84 (0.32)	-0.66 (0.19)	1.10 (0.22)	FR1	Bear	1.71 (0.26)	-1.48 (0.26)	-0.62 (0.19)	0.66 (0.15)
FR11	Cactus	0.99 (0.21)	-1.59 (0.38)	-0.03 (0.20)	1.77 (0.41)	FR6	Anchor	1.26 (0.22)	-1.77 (0.35)	-0.49 (0.19)	0.80 (0.22)	FR2	Skate	1.50 (0.26)	-1.45 (0.29)	-0.28 (0.16)	1.02 (0.21)
FR9	Ruler	0.98 (0.21)	-1.57 (0.39)	-0.01 (0.20)	1.50 (0.36)	FR10	Balloons	1.22 (0.22)	-1.61 (0.34)	-0.23 (0.19)	1.18 (0.27)	FR10	Triangle	1.41 (0.26)	-1.44 (0.32)	-0.37 (0.18)	1.10 (0.21)
FR10	Telescope	1.34 (0.24)	-1.55 (0.29)	-0.08 (0.16)	1.23 (0.25)	FR11	Nine	1.21 (0.23)	-1.40 (0.29)	-0.01 (0.19)	1.84 (0.38)	FR5	Volcano	1.43 (0.25)	-1.32 (0.27)	-0.29 (0.17)	0.84 (0.19)
FR3	Owl	1.51 (0.25)	-1.48 (0.26)	-0.32 (0.16)	1.24 (0.21)	FR8	Rolling Pin	1.60 (0.25)	-1.31 (0.23)	-0.11 (0.14)	1.57 (0.25)	FR9	Crown	1.86 (0.31)	-1.16 (0.24)	-0.40 (0.15)	0.62 (0.13)
FR13	Steer Wheel	1.15 (0.21)	-1.25 (0.30)	0.08 (0.18)	1.43 (0.30)	FR7	Watch	1.15 (0.23)	-1.22 (0.30)	0.29 (0.19)	1.92 (0.39)	FR3	Onion	1.46 (0.26)	-1.01 (0.22)	0.34 (0.16)	1.65 (0.29)
FR7	Paper Clip	1.13 (0.22)	-1.11 (0.28)	0.43 (0.18)	1.94 (0.40)	FR3	Pretzel	1.37 (0.24)	-1.02 (0.23)	0.36 (0.16)	1.64 (0.31)	FR4	Broom	1.43 (0.26)	-0.84 (0.23)	0.57 (0.18)	2.33 (0.46)
FR2	Racquet	0.93 (0.19)	-1.03 (0.31)	0.63 (0.25)	2.63 (0.60)	FR1	Spider	1.85 (0.28)	-0.99 (0.18)	0.12 (0.12)	1.44 (0.22)	FR7	Dominoes	1.10 (0.26)	-0.64 (0.26)	0.75 (0.24)	2.17 (0.52)

* The *a* parameter is the item discrimination; ** The *b* parameters are the item difficulty associated with *k+1* response categories

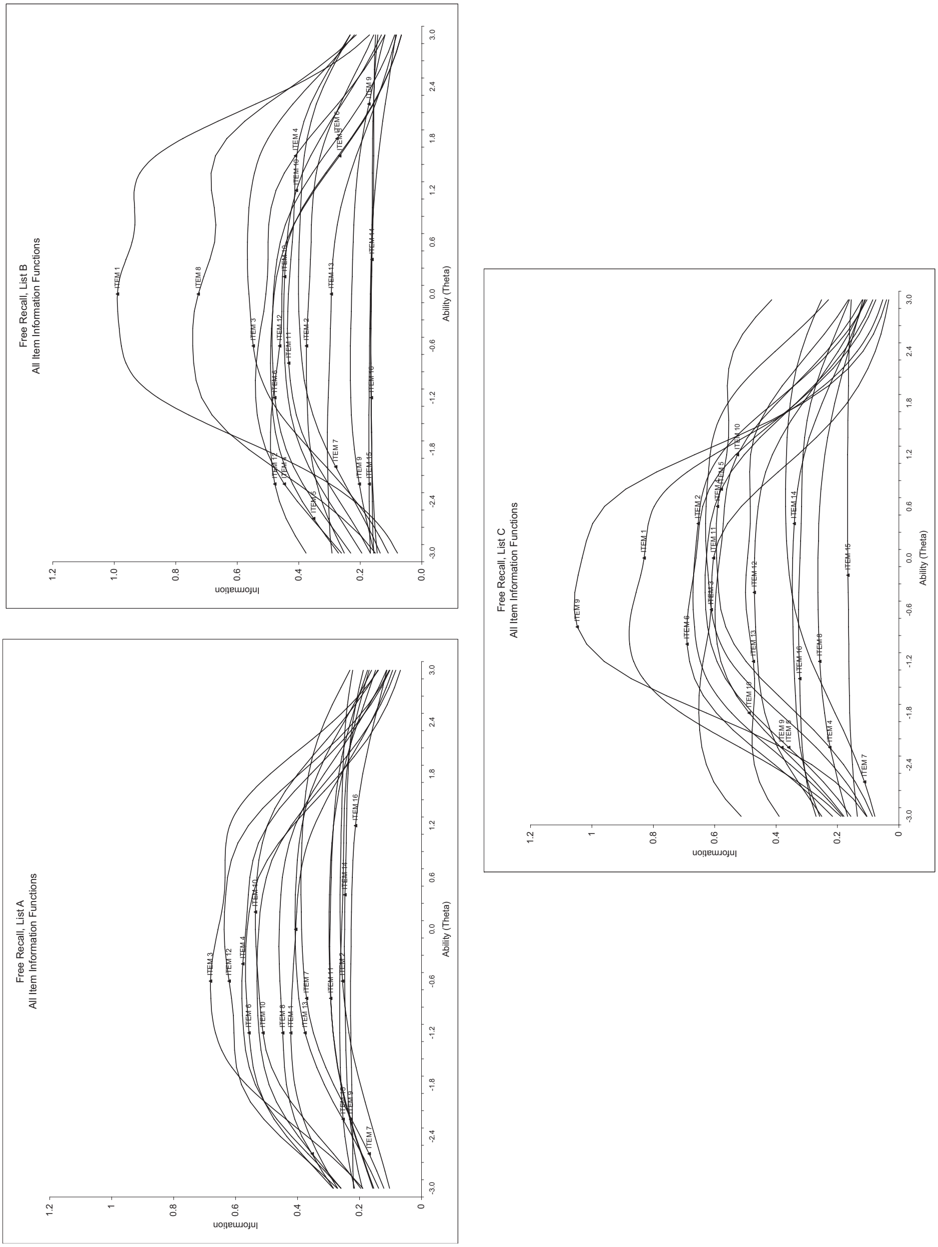

Figure 2: Item Information Functions by Form

items was on Form C and peaked at 1.06 for item FR9 – Crown at theta level -0.6 and 0.88 for item FR 1 – Bear at theta level -0.8. The next highest information was 0.99 on Form B and peaked at 0.99 for item FR1 – Spider at theta level 0.0 and 0.75 for item FR8 – Rolling pin at theta level -0.6 to -0.4. The highest item information on Form A was 0.68 at theta = -0.6 for item FR3 – Owl and 0.64 at theta 0.0 for item FR12 - Rattle. The following items provided the least information: items FR14 – Pipe, FR15 – Wreath and FR16 – Basket on Form B and item FR15 – Thread on Form C; these items never reached the level of 0.2 at any point along the theta continuum. Items on Form C provide peak information along the largest spread of thetas/ability levels.

The test information functions (Figure 3) show that the test covered the full range of levels of memory ability/theta, discriminating better in the middle range and slightly below the mean. The peak values for the information functions were 7.8, 7.9 and 9.7 for Form A, Form B and Form C respectively.

Classical test theory and IRT reliability analyses

For all forms, the overall Cronbach's alpha reliability coefficient was good: 0.85, 0.86 and 0.88. Corrected item-total correlations were moderate for all the items, ranging from 0.37 to 0.55 for Form A; 0.34 to 0.60 for Form B; 0.35 to 0.65 for Form C.

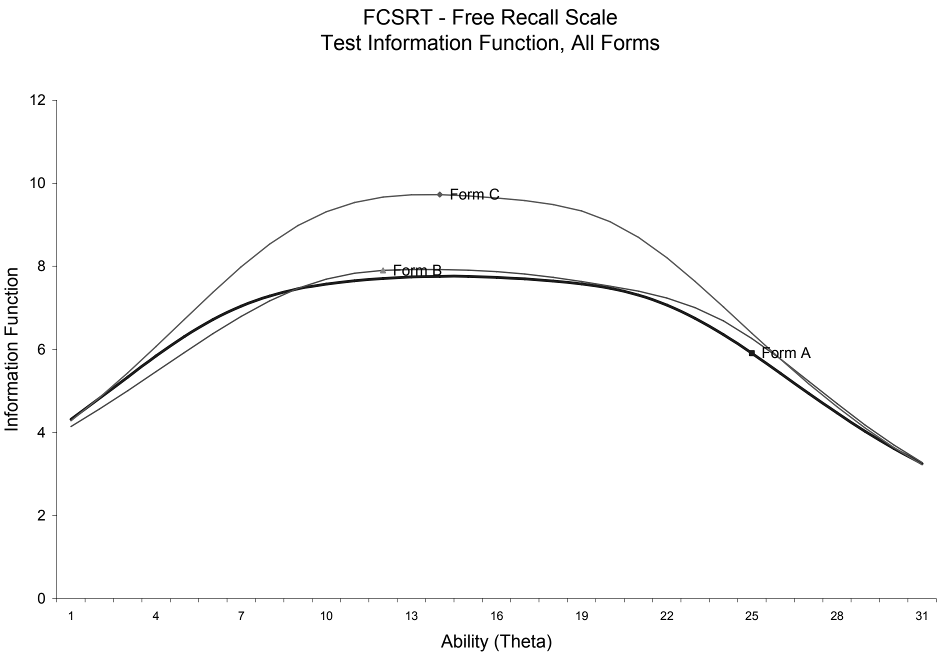

Figure 3:
Test Information Functions by Form

IRT analysis allows computation of test reliability estimates over the entire spectrum of the theta/cognitive ability continuum. The results show high reliability at all levels, for all three forms, with the highest value of 0.91 in the theta range from -0.8 to 0.2 for Form C, in the high 0.80's over most of the theta distribution for all forms, and in high 0.70's at the high ability end (2.8 – 3.0) of the theta distributions. The marginal reliabilities for the forms were: 0.86, 0.87 and 0.89 for A, B, and C respectively.

Examination of the distributions

The distributional properties for the sum and theta scores were examined for the three forms and are summarized in Table 4. Typically the empirical distributions for these variables are normal or close to normal, but either negatively or positively skewed. This was the case for the current data set. The Kolmogorov-Smirnov statistics, used to test for distributional normality, were usually significant but relatively small. Examination of normal and detrended normal plots showed that the distributions of the original sum scores were approximately normal although negatively skewed for all three forms; the negative skew was partially due to the very low performance of the diagnosed cases of dementia included in the sample.

Table 4:
Distributional Characteristics of the Summary Score and Ability/Theta* Estimates: FCSRT-IR, Free Recall 16 Item Scales, Form A, Form B and Form C

Statistic	**Form A**		**Form B**		**Form C**	
	Sum Score	**Theta/ Ability**	**Sum Score**	**Theta/ Ability**	**Sum Score**	**Theta/ Ability**
N	210		189		182	
Mean	26.4	0.00	27.0	0.01	26.7	0.01
St. Dev.	8.7	.92	8.7	.92	9.4	.93
Median	27.5	.13	29.0	.14	28.0	.15
Minimum	0.0	-3.20	2.0	-2.93	0.0	-3.09
Maximum	45	2.16	44	2.05	44	1.93
Interqartile Range	11.0	1.04	12.0	1.16	12.0	1.07
Skewness (Std. Error)	-0.73 (0.17)	-0.94 (0.17)	-0.56 (0.18)	-0.54 (0.18)	-0.84 (0.18)	-1.00 (0.18)
Kurtosis (Std. Error)	0.56 (0.33)	1.56 (0.33)	-0.04 (0.35)	0.33 (0.35)	0.34 (0.36)	1.09 (0.36)
Kolmogorov-Smirnov Test of Normality	0.11, d.f. 210, p <0.001	0.11, d.f. 210, p <0.001	0.11, d.f. 189, p <0.001	0.08, d.f. 189, p = 0.002	0.10, d.f. 182, p <0.001	0.10, d.f. 182, p <0.001

* A higher score is indicative of higher cognitive ability

Concurrent criterion validity

ROC curves were derived separately for each FCSRT-IR form and then for the combined forms. The area under the curve (AUC) was similar for the three forms ranging from 0.89 (95%CI: 0.84,0.94) for form A, 0.87 (95% CI: 0.80,0.94) for form B, and 0.91 (95% CI: 0.86,0.96) for form C. Using the previously established cut score of ≤24 to distinguish prevalent and incident cases of dementia from non-cases (Grober et al, 2000), the sensitivity and specificity of the three forms was similar: for form A, sensitivity = 0.75, specificity = 0.82; for form B, sensitivity = 0.75, specificity = 0.86; and for form C, sensitivity = 0.77, specificity = 0.87. Figure 4 shows the ROC curve for all forms combined. AUC was 0.88 (95% CI: 0.86, 0.92). Sensitivity and specificity were 0.75 and 0.85 respectively using the cut score of ≤ 24.

Discussion

Overall, the results provide support for the psychometric soundness of the FCSRT-IR. First, the factor analyses indicate that the 16 unique category-item pairs in each of the three forms assess a single construct or dimension which we presume to be memory ability. This

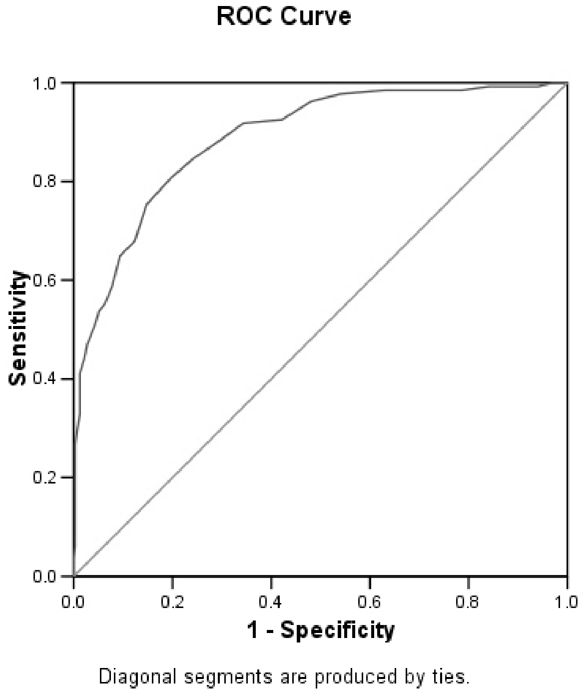

Figure 4:
ROC Curve for the Three Forms Combined

evidence of essential (sufficient) unidimensionality is important because it provides supporting evidence for our approach to identifying preclinical and early AD. By testing memory under controlled learning conditions, genuine deficits in encoding and storage that characterize AD are distinguished from the memory deficits associated with normal aging that are secondary to impaired attention, inefficient information processing, or ineffective retrieval operations (Light, 1991). The finding that a unidimensional model was the best fit for the data suggests that memory ability alone determines performance on the FCSRT-IR. Second, high reliability was evidenced in the IRT analyses for all three forms over the entire spectrum of the memory ability continuum. The test information functions indicate that the FCSRT-IR is reasonably discriminating over a relatively large range of the memory ability continuum. We view this as an advantage because the test is used not just for identifying incident and prevalent dementia but also for quantifying memory functioning in elderly persons without dementia at cross-section and longitudinally (Grober, Lipton, Katz, & Sliwinski, 1998; Ivnik, 1997). Furthermore, the test has been shown in other studies to have promise for use in clinical trials to quantify change in memory functioning in patients with mild cognitive impairment, the transitional state between normal cognition and Alzheimer's disease (Ferris et al., 2006; Sarazin et al., 2007).

Third, the FCSRT-IR was shown to have good concurrent criterion validity according to ROC curves derived separately for each FCSRT-IR form and then for the combined forms. The criterion was presence or absence of dementia at baseline or follow-up determined by expert consensus using DSM IV criteria independent of FCSRT-IR scores to avoid diagnostic circularity. Using the previously determined cut score of 24 or less over three trials of free recall (max recall = 48), sensitivity and specificity of the three forms combined for incident and very mild dementia (CDR=0.05) were 0.75 and 0.85 respectively. Classification accuracy was adequate (AUCs ranged from 0.87 to 0.91 across forms). That it was not higher reflects the reduction in accuracy that occurs when a cut score developed in one sample is applied to a different sample. In addition, the current cohort consists of primary care patients with co-morbid medical conditions that can complicate classification rather than patients from a memory disorders clinic that constitute an enriched sample.

Finally, the psychometric properties of the three forms of the FCSRT-IR were similar including their factor structure, internal consistency, information functions, and accuracy of classification, although Form C performs slightly better, in terms of test information. Limitations of the analyses include the fact that the statistical equivalence of the three forms could not be evaluated because the forms do not include common category-item pairs; however, summary statistics such as means, medians and interquartile ranges were similar, as were reliability and validity estimates. Overall, each form appears to be a reliable and valid measure of memory ability.

The present results may also be limited by the relatively small, but adequate sample sizes (n=about 200 across forms). The factor analyses resulted in acceptable properties, although slight misfit was observed. It is likely that with larger sample sizes the accuracy of the estimates from the IRT, exploratory and confirmatory factor analyses, which are sensitive to small sample sizes, will improve. Additionally, future work will involve examining possible differential item functioning of the measure with respect to education and racial/ethnic groups; relatively small sample sizes did not permit such analyses.

Acknowledgments

The Free and Cued Selective Reminding Test with Immediate Recall (FCSRT-IR) is copyrighted by the Albert Einstein College of Medicine (AECOM) and is made freely available from the Albert Einstein College of Medicine for teaching and academic research purposes. To request a copy of the FCSRT-IR materials, contact Dr. David Silva at dsilva@aecom.yu.edu. EG receives a small royalty for commercial use of the FCSRT.

References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, D.C.: American Psychiatric Association Press.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80* (3), Pt.2.

Bock, R. D., & Aitken, M. (1981). 'Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm'. *Psychometrika, 46*, 443-445.

Bohnstedt, M., Fox, P., & Kohatsu, N. (1994). Correlates of Mini Mental Status examination among elderly demented patients: the influence of race-ethnicity. *Journal of Clinical Epidemiology, 47*, 1381-1387.

Buschke, H. (1984). Cued recall in Amnesia. *Journal of Clinical and Experimental Neuropsychology, 6*, 433-440.

Buschke, H., Sliwinski, M., Kuslansky, G., & Lipton, R. B. (1995). Aging, encoding specificity, and memory change in the Double Memory Test. *Neurology, 1*, 483-493.

Callahan, C. M., Hendrie, H. C., & Tierney, W. M. (1995). Documentation of evaluation of cognitive impairment in elderly primary care patients, *Annals of Internal Medicine, 122*, 422-429.

Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). DIF analysis with ordinal logistic regression techniques: DIFDETECT. *Medical Care, 44,* S115-123.

Cummings, J. L., & Benson, D. F. (1992). *Dementia: A clinical approach*. Boston: Butterworth-Henemann.

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.

Ferris, S. H., Aisen, P. S., Cummings, J., Galasko, D., Salmon, D., Schneider, L., et al. (2006). ADCS Prevention Instrument Project: Overview and initial results. *Alzheimer's Disease and Associated Disorders, Supplement*.

Folstein, M. F., Folstein, S., & McHugh, P. R. (1975). "Mini Mental State": A practical methods of grading cognitive status of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

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., Dickson, D., Sliwinski, M. J., Buschke, H., Katz, M., Crystal, H., et al. (1999). Memory and mental status correlates of modified Braak staging. *Neurobiology of Aging, 20*(6), 573.

Grober, E., Gitlin, H., Bang, S., & Buschke, H. (1992). Implicit and explicit memory in young, old, and demented adults. *Journal of Clinical and Experimental Neuropsychology, 14*(2), 298-316.

Grober, E., Hall, C., Lipton, R. B., Zonderman, A., Resnick, S., & Kawas, C. (2008a). Memory Impairment, Executive Dysfunction, and Intellectual Decline. *Journal of the International Neuropsychological Association, 14*, 266-278.

Grober, E., Hall, C., McGinn, M., Nicholls, T., Stanford, S., Ehrilich, A., et al. (2008b). Neuropsychological strategies for detecting early dementia. *Journal of the International Neuropsychological Society, 14*, 130-142.

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., Lipton, R. B., Katz, M., & Sliwinski, M. (1998). Demographic Influences on Free and Cued Selective Reminding Performance in Older Persons. *Journal of Clinical and Experimental Neuropsychology, 20*, 221-226.

Grober, E., Merling, A., Heimlich, T., & Lipton, R. B. (1997). Comparison of selective reminding and free and cued selective reminding in the elderly. *J Clinical and Experimental Neuropsychology, 19*, 643-654.

Habert, M.-O., Horn, J.-F., Sarazin, M., Lotterie, J.-A., Puel, M., Onen, F., et al., (2009). Brain perfusion SPECT with an automated quantitative tool can identify prodromal Alzheimer's disease among patients with mild cognitive impairment. *Neurobiology of Aging, In Press, Corrected Proof.*

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory.

Ivnik, R. J., Smith, G. E., & Lucas, J. A. (1997). Free and cued selective reminding test: MOANS norms. *Journal of Clinical and Experimental Neuropsychology, 19*, 676-691.

Jones, R. N. (2006). Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination: Detecting differential item functioning using MIMIC modeling. *Medical Care, 44*, S124-133.

Lekeu, F., Van der Linden, M., Chicherio, C., Collette, F., Degueldre, C., Franck, G., et al. (2003). Brain Correlates of Performance in a Free/Cued Recall Task With Semantic Encoding in Alzheimer Disease. *Alzheimer Disease and Assocaited Disorders, 17*, 35-45.

Light, L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42*, 333-376.

Lindenberger, U., & Reischies, F. M. (1999). *Limits and potentials of intellectual functioning in old age*. New York: Cambridge University Press.

Lipton, R. B., Katz, M., Kuslansky, G., Sliwinski, M., Stewart, W., Verghese, J., et al. (2003). Screening for Dementia by Telephone Using the Memory Impairment Screen. *Journal of the American Geriatrics Society, 51*, 1382-1390.

Lord, F. M. & Novick, M. R. (1968). Statistical theories of mental test scores (with contributions by A. Birnbaum). Reading, MA: Addison-Wesley Publishing Co.

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Lawrence Erlbaum.

McKhann, G., Drachman, D., Folstein, M., et al. (1984). Clinical diagnosis of Alzheimer's Disease: Report of the NINCDS-ADRA work group under the auspices of Department of Health and Human Services Task force on Alzheimer's Disease. *Neurology, 34*, 939-944.

Morales, L. S., Flowers, C., Gutiérrez, P., Kleinman, M., & Teresi, J. A. (2006). Item and scale differential functioning of the Mini-Mental Status Exam assessed using the DFIT methodology. *Medical Care, 44*, S143-151.

Multhen, L. K., & Multhen, B. O. (2004). MPLUS.

Mungas, D., & Reed, B. R. (2000). Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. *Statistics in Medicine, 19*, 1631-1644.

Orlando-Edelen, M., Thissen, D., Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the Mini-mental status examination. *Medical Care ,44*, S134-142.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine, 256*, 183-194.

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.

Petersen, R. C., Smith, G. E., Ivnik, R. J., Tangalos, E. G., Schaid, D. J., Thibodeau, S. N., et al. (1995). Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals. *Journal of the American Medical Association, 273*, 1274-1278.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 17*.

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.

Teresi, J. A., Golden, R. R., Cross, P., Gurland, B., Kleinman, M., & Wilder, D. (1995). Item bias in cognitive screening measures: Comparisons of elderly white, Afro-American, Hispanic and high and low education subgroups. *Journal of Clinical Epidemiology, 48*, 473.

Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2000). Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Statistics in Medicine, 19*, 1651-1683.

Teresi, J. A., Kleinman, M., Ocepek-Welikson, K., Ramirez, M., Gurland, B., Lantigua, R., et al. (2000). Applications of Item Response Theory to the Examination of the Psychometric Properties and Differential Item Functioning of the Comprehensive Assessment and Referral Evaluation Dementia Diagnostic Scale Among Samples of Latino, African American, and Non-Latino White Elderly. *Research on Aging, 22*, 738-773.

Thissen, D. (1991, 2001). *MULTILOG™ User's Guide: Multiple Categorical Item Analysis and Test Scoring Using Item Response Theory* (Version 6.0). Chicago, Il: Scientific Software International, Inc.

Tuokko, H., Kristjansson, E., & Miller, J. A. (1995). The neuropsychological detection of dementia: An overview of the neuropsychological component of the Canadian Study of Health and Aging. *Journal of Clinical and Experimental Neuropsychology, 13*, 871-879.

Zimmerman, M. E., Pan, J. W., Hetherington, H. P., Katz, M. J., Verghese, J., Buschke, H., et al. (2008). Hippocampal neurochemistry, neuromorphometry, and verbal memory in nondemented older adults. *Neurology, 70*, 1594-1600.