Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination

Gail Kuslansky [a,*], Mindy Katz [a], Joe Verghese [a], Charles B. Hall [a,c],
Pablo Lapuerta [b], Gia LaRuffa [a], Richard B. Lipton [a,c,d]

[a] *Einstein Aging Study, Department of Neurology, Albert Einstein College of Medicine,
1300 Morris Park Avenue, Bronx, New York City, NY 10461, USA*
[b] *Bristol-Myers-Squibb Company, New York, NY, USA*
[c] *Department of Epidemiology and Social Medicine, Albert Einstein College of Medicine,
Bronx, New York City, NY, USA*
[d] *Institute for Medical Effectiveness Research, Albert Einstein College of Medicine,
Bronx, New York City, NY, USA*

Abstract

The Hopkins Verbal Learning Test (HVLT) and the Mini-Mental State Examination (MMSE) were administered to 323 non-demented elderly and 70 individuals who meet DSM-IV criteria for dementia in order to compare the validity of these two measures for detecting mild dementia and for the two most common dementia subtypes, Alzheimer's disease (AD) and vascular dementia (VaD). The study was conducted in an elderly, ethnically diverse community-dwelling population. Sensitivity, specificity, positive and negative predictive values were calculated over a range of clinically relevant cut scores for each test. We analyzed the influence of age, education, reading ability and sex on test performance using logistic regression models.

When sensitivity is held constant at 0.69, the specificity for the HVLT total recall was 0.89 and the MMSE 0.82 for all dementias ($P = .10$). Age, sex and education did not significantly influence test performance for either test in this sample. Results were similar for AD and VaD. However, while adding a measure of reading ability to the regression models did not affect the overall dementia model, it resulted in improved specificities when combined with the MMSE for AD and combined with the HVLT for VaD. Additional tests such as reading ability can improve discrimination of dementia subtypes. The modest

* Corresponding author. Tel.: +1-718-430-4272; fax: +1-718-430-3870.
E-mail address: Kuslansk@aecom.yu.edu (G. Kuslansky).

sensitivity of either the HVLT or the MMSE alone suggests that further neuropsychological evaluation is required to confirm dementia diagnosis.

1. Introduction

Since memory impairment is the hallmark of dementia and frequently the first symptom (American Psychiatric Association, 1997), accurate testing for memory deficits is an essential part of detecting early dementia. Unfortunately, the diagnosis of dementia of all types and Alzheimer's disease (AD) in particular, is frequently delayed or missed in primary care settings (Callahan, Hendrie, & Tierney, 1995; Liston, 1978). Patients' cognitive complaints are often attributed to the normal aging process (Ross et al., 1997). AD is an increasingly treatable disorder with several approved pharmacological treatments (donezepil, rivastigmine, and galantamine), and new treatments in development (Cesura et al., 1995; Gelmacher, 1997; Petit et al., 1998; Raskind, Peskind, Wessel, & Yuan, 2000). Treatments may also be more effective early in disease before there is extensive neuronal damage and loss (Gauthier, Thal, & Rossor, 1996; Khatchaturian, Phelps, & Buchholz., 1994).

In addition, as candidate interventions are developed for treating dementia, effective approaches for dementia screening are required for application to clinical trials and clinical practice. In this context, the ideal screening test to detect memory impairment should be sensitive to early disease so that incident cases of dementia are efficiently detected, and specific, so that most patients with dementia are identified and referred for definitive diagnostic evaluations (U.S. Agency for Health Care Policy and Research, 1996).

Memory test and mental status tests have been used to screen for dementia in clinical trials and clinical practice. Medical students and residents are often taught to present a simple three- or four-word list to remember, and assess free recall after 3–5 min (Mesulam, 2000; Petersen, 1991; Strub & Black, 1985; Trzepacz & Baker, 1993). These three- or four-word delayed free recall tests are also often part of mental status tests (Blessed, Tomlinson, & Roth, 1968; Cammermeyer & Evans, 1988; Folstein, Folstein, & McHugh, 1975; Kokmen, Smith, Petersen, Tangalos, & Ivnik, 1991; Mattis, 1976; Pfeiffer, 1975; White, Bauer, Bowers, Crosson, & Kessler, 1995). The available evidence suggests that these three- and four-word free recall tests may generate unacceptably high "false positive" rates for dementia (Beardsall & Huppert, 1991; Cullum, Thompson, & Smernoff, 1993; Jenkyn et al., 1985; Kuslansky et al., 2002). Other brief memory tests with utility in dementia screening include the Memory Impairment Screen (MIS; Buschke et al., 1999), the East Boston Memory Test (Albert et al., 1991), and the 10-item free recall with enhanced learning (Knopman & Ryberg, 1989).

A promising screening test for memory impairment is the Hopkins Verbal Learning Test (HVLT; Brandt, 1991). It is a three-trial list learning and free recall task comprising 12 words, 4 words from each of three semantic categories. Because the HVLT has six equivalent alternate forms, it is particularly appropriate for serial testing as part of longitudinal studies; alternative forms can be used to circumvent practice effects due to item familiarity (Fugita et al., 1999;

Harris, Heaton, Schalz, Bailey, & Patterson, 1997). Test–retest correlations of the HVLT are similar to those of other verbal memory tests such as the Logical Memory subtest of the Wechsler Memory Scale—Revised and the California Verbal Learning Test (Rasmusson, Bylsma, & Brandt, 1995). Other studies of the HVLT support its alternate form and test–retest reliability (Benedict, Schretlen, Groninger, & Brandt, 1998) and its construct and content validity (Shapiro, Benedict, Schretlen, & Brandt, 1999).

The reliability and validity of the HVLT has been demonstrated in patients with head injury (Guskiewicz, Riemann, Perrin, & Nashner, 1997; Lovell, Iverson, Collins, McKeag, & Maroon, 1999), schizophrenia (Bryson, Bell, & Lysaker, 1997), and dementia (Frank & Byrne, 2000; Shapiro et al., 1999). These dementia studies were conducted in patients referred to a geriatric psychiatry practice (Frank & Byrne, 2000) and in patients referred for neuropsychological evaluation (Shapiro et al., 1999), but not in community-based samples. These studies are limited by small samples (Frank & Byrne, 2000; Guskiewicz et al., 1997), and selection bias (Beardsall & Huppert, 1991; Bryson et al., 1997; Shapiro et al., 1999). Hogervorst and associates (Hogervorst, Combrinck, & Lapuerta, 2001) found the HVLT total recall score to have 87% sensitivity and 98% specificity for discriminating "demented patients" from "healthy controls", but they eliminated all cases with "questionable" diagnoses from their analyses. This result may overestimate test characteristics in a community sample with a continuous range of symptom severity. The HVLT has been revised to include delayed recall; the revised version adds considerably to testing time and may not be practical in all clinical settings.

Mental status tests are an important and widely used alternatives to memory tests. The Mini-Mental State Examination (MMSE; Folstein et al., 1975) has been used for detecting dementia for over 25 years (Morris et al., 1989; Schmitt, Ramseen, & DeKosky, 1989). The MMSE includes measures of memory, attention, formation and other cognitive domains. The memory task consists of three words, which are repeated immediately after presentation and are recalled after two additional tasks (five serial subtractions and backward spelling). The MMSE also includes orientation items, figure copying, reading and writing. The MMSE has been recommended as a screen for early dementia (Knopman et al., 2001). Discriminative validity for the MMSE as been reported to improve with adjustment for age and education (Chum, Anthony, Bassett, & Folstein, 1993; Kittner et al., 1986).

The aim of this study was to directly compare the performance of the HVLT and the MMSE as screening tools for dementia in a community-based sample. Participants were recruited from the Einstein Aging Study (EAS), a longitudinal study of normal aging and dementia conducted at the Albert Einstein College of Medicine in the Bronx County of New York City. Because memory deficits may not be as severe in vascular dementias (VaDs) as they are in AD, we also investigated the discriminative validity of the HVLT and the MMSE in the two most common dementia subtypes, AD and VaD.

2. Method

2.1. Participants

The sample comprised 393 participants in the EAS, a longitudinal study of cognitive aging and dementia, conducted in a multi-ethnic, community-dwelling population. All competent

participants gave informed consent as specified by the Committee on Clinical Investigations at Albert Einstein College of Medicine. Others gave assent with informed consent obtained from their next of kin. Of the 393 participants included in this study, 372 individuals were systematically sampled from the Medicare enrollment lists for the area adjacent to the our clinical research center. To supplement the 49 individuals with dementia in the systematic sample, we recruited 21 additional community volunteers with dementia. These individuals did not differ significantly from the Medicare recruits with respect to age, sex, ethnicity and education. Eligible individuals were aged 70 years or older, ambulatory, and able to understand task instructions and respond in English. The control group of elderly individuals without dementia did not exclude 71 individuals who reported memory complaints and received a Clinical Dementia Rating Scale score of 0.5 for memory (Hughes et al., 1982).

2.2. Procedures

All participants were administered the EAS clinical neuropsychological test battery and medical history, epidemiological, social and behavioral questionnaires as part of the EAS. The study neurologist performed a neurological examination and ordered additional diagnostic testing, as clinically indicated, including neuroimaging and blood tests. A diagnosis of dementia was made according to the DSM-IV criteria (American Psychiatric Association, 1997). A diagnosis of AD was based on the NINCDS/ADRDA criteria (McKhann et al., 1984) and VaD was diagnosed according to Chui et al. (Chui et al., 1992). The study sample comprised 323 non-demented elderly and 70 elderly with dementia. Forty-eight participants in the dementia group were diagnosed with possible or probable AD, 10 were diagnosed with possible or probable VaD, 9 were diagnosed with mixed dementia, and 3 had other subtypes (i.e., two fronto-temporal dementia and one dementia with Lewy bodies). After the EAS clinic visit, each subject was asked to return for a brief second day of testing, during which the HVLT and the MMSE were administered.

2.3. Materials

The HVLT was administered according to authors' instructions (Brandt, 1991). Briefly, the examiner read the 12 words aloud and the subject was asked to freely recall them immediately. The list was read a second time followed by a second free recall trial. This was followed by a third reading and third free recall. The words recalled for each trial were recorded and a total recall score tallied (range: 0–36). The free recall trials were followed by a yes/no recognition trial, which consisted of 24 words: 12 were the target list words; 6 were categorically related non-target words; and 6 were unrelated words. As the examiner read each word, the subject answered 'yes' if s/he thought it was one of the target words and 'no' if s/he thought it was not a target. Recognition was scored two ways: (1) total number of correct responses and (2) an adjusted score that subtracted false alarms from the correct responses. The entire HVLT requires less than 10 min to administer. Although a revised version of the HVLT (Shapiro et al., 1999) added delayed recall to improve discrimination, the additional administration time was impractical in this context.

The six alternate forms of the HVLT were counterbalanced for administration to the dementia and no-dementia groups, so that each form was used equally often.

The MMSE was administered according to authors' instructions (Folstein et al., 1975). Total scores could range from 0 to 30. Although there have been modifications to the MMSE (Teng, Chiu, Schneider, & Metzger, 1987), they take longer to administer and may not meet the "brevity" criteria for a screening measure. Tombaugh et al. (Tombaugh, McDowell, Krisjansson, & Hubley, 1996) found that the MMSE and the modified-MMSE (3MS), which added fluency, similarities and delayed recall, did not differ in sensitivity to AD.

2.4. Statistical analyses

The various groups were compared with respect to demographic variables using parametric and non-parametric measures. Discriminative validity of the HVLT recall and recognition scores and the MMSE was assessed by calculating the sensitivity and specificity of these tests for detecting dementia and for detecting AD and VaD for various cut scores. Logistic regression and receiver operating characteristic (ROC) analysis were used to examine the various sensitivity–specificity trade-offs of the HVLT recall and recognition scores and the MMSE scores for detecting dementia. Sensitivity, specificity and their confidence intervals were estimated for different HVLT and MMSE cut scores. Examining 70 dementia cases and 323 non-demented controls yielded a 95% confidence interval of less than 12% for sensitivity and less than 7% for specificity. Sensitivities and specificities for the HVLT and the MMSE for discriminating dementia from no dementia were determined over a range of cut scores.

Positive predictive value (PPV), an important index of screening efficiency, is the proportion of individuals that test positive who have dementia and were determined over a range of base rates, as described below (Streiner, Norman, & Blum, 1989). PPV varies with the prevalence of the disease in the screened population as well as the specificity of the test. Negative predictive value (NPV), the proportion of non-demented persons who screen negative, tells us how effectively a test identifies non-demented persons as unimpaired and varies with the disease prevalence and the sensitivity of the test measure.

Subset analyses were conducted to assess whether the performance of the HVLT and the MMSE are comparable for detecting AD and VaD. Although normative data are often presented in the form of percentiles or means and standard deviations, we present norms in the form of the probability of dementia (or AD) given different HVLT or MMSE cut scores. To calculate the probability of dementia, one must know the test sensitivity and specificity at each cut score, and the base rate of dementia (Altman, 1991). The probability of all dementia, and AD and VaD in particular, was calculated according to Bayes' Theorem (Elwood, 1993).

We used two methods to compare the performance of the HVLT and MMSE for detecting dementia. First, the area under the ROC curves was compared using an algorithm proposed by Metz and Pan (1999). However, because this procedure provides an omnibus test of area under the ROC curve, it can be influenced by differences in test performance that are in ranges of sensitivity and specificity that are not clinically or practically relevant. Therefore, we used the McNemar test to contrast the specificities of the HVLT and MMSE for fixed and clinically important levels of sensitivity (Fleiss, 1981).

We used two approaches to examine the influence of age, sex, education and reading ability on HVLT performance as it pertains to the detection of dementia. Logistic regression analyses indicate whether overall HVLT or MMSE performance is influenced by age, sex, education and reading ability. These analyses do *not* assess the influence of these variables on the optimal cut scores of the HVLT or MMSE for detecting dementia. A second logistic regression analysis tested whether cut scores for the HVLT or MMSE need to be modified according to patient characteristics (age, sex, education or reading ability) to optimize dementia discrimination.

Logistic regression and ROC analyses were used to determine whether combining standard or adjusted recognition scores with recall scores improve discriminative validity of the HVLT for dementia. The same analytic methods described above (testing area under ROC curves and McNemar's tests) were used to compare the dementia discrimination of recall scores with discrimination of recall scores combined with both standard and adjusted recognition scores.

3. Results

3.1. Demographic and neuropsychological variables

The demographic and neuropsychological characteristics of the 323 individuals without dementia and the 70 individuals with dementia are shown in Table 1.

The Dementia group includes the AD only group ($n = 48$), the VaD only group ($n = 10$), the mixed AD/VaD group ($n = 9$), fronto-temporal dementias ($n = 2$) and dementia with Lewy bodies ($n = 1$). The dementia groups did not differ significantly from the no-dementia group with respect to sex or ethnicity. The dementia group (82.0 years) was significantly older

Table 1
Demographic and neuropsychological characteristics of participant groups

	Control participants	Particpants with all dementias[a]	AD only	VaD only
Sample size	323	70	48	10
Sex (% male)	40	40	33	30
Age	78.6 ± 5.3	82.0 ± 5.5	81.7 ± 5.3	82.2 ± 6.8
Education	12.9 ± 3.3	11.8 ± 3.4	11.8 ± 3.2	10.6 ± 3.0
Ethnicity (% caucasian)	65	63	65	50
BIMC	3.3 ± 2.6	10.4 ± 5.2	10.9 ± 4.3	7.9 ± 4.5
HVLT free recall total	20.7 ± 5.7	12.0 ± 4.6	12.5 ± 4.1	12.0 ± 4.1
MMSE	26.1 ± 2.1	22.3 ± 4.0	22.4 ± 3.4	21.4 ± 3.2
GDS	2.8 ± 2.4	4.6 ± 3.2	4.5 ± 3.4	5.1 ± 2.4
WAIS-R Verbal IQ	105.1 ± 14.9	90.6 ± 11.3	91.9 ± 12.7	83.3 ± 4.2
WRAT-R reading	66.6 ± 15.1	59.5 ± 18.3	61.1 ± 18.2	44.0 ± 16.5
WAIS-R Digit Symbol Score	34.5 ± 12.3	19.3 ± 12.2	20.2 ± 12.2	15.9 ± 7.2
WAIS-R Block Design Score	17.3 ± 8.4	7.9 ± 7.5	8.8 ± 8.0	4.7 ± 4.0

Means (S.D.) are presented except for sample size, sex, and ethnicity.

[a] Includes persons with AD only, VaD only, mixed AD/VaD, fronto-temporal dementia and dementia with Lewy bodies.

($P < .001$) than the group without dementia (78.6 years) and had fewer years of education ($P = .011$). The no-dementia group outperformed the all-dementia group on the Blessed test of mental status (BIMC, $P < .001$), reading level (WRAT-R, $P = .001$), WAIS-R Verbal IQ ($P < .001$), cognitive-motor speed (WAIS-R Digit Symbol Substitution, $P < .001$), memory (HVLT, $P < .001$), and problem solving (WAIS-R Block Design (53) subtest, $P < .001$).

The AD group did not differ significantly from the VaD group with respect to age, sex, education, ethnicity, mental status or IQ variables. The AD group scored significantly higher with respect to WRAT-R reading scores ($P = .008$).

3.2. All dementias

We examined the effects of age, sex, education and reading ability in the logistic regression models for the HVLT and the MMSE. None of these variables influenced either the HVLT or MMSE performance or the cut scores for detecting dementia.

There were significant mean differences in HVLT free recall score and MMSE performance between participants with dementia and non-cases (Table 1, HVLT free recall, $P < .001$; MMSE, $P < .001$). Focusing first on the HVLT, Figure 1 shows the sensitivity–specificity trade-offs of different cut scores on the free recall portion of the HVLT for discriminating the all-dementia group from the group with no dementia.

The total area under the curve (AUC) for the HVLT is 0.89. At a cut score of <16, the HVLT has a sensitivity of 0.83 and specificity of 0.83 (see Table 2). While examination of Figure 1 shows that while the HVLT appears to be more effective (i.e., greater specificity at relevant cut scores), these differences were not statistically significant.
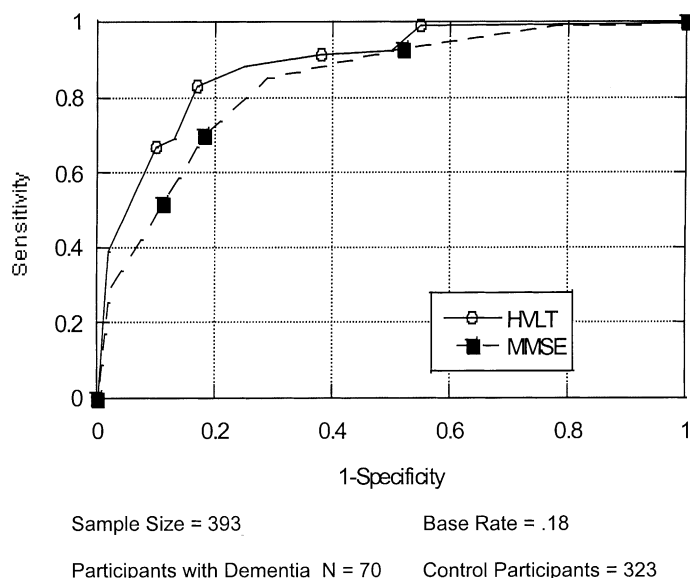

Sample Size = 393   Base Rate = .18

Participants with Dementia N = 70   Control Participants = 323

Fig. 1. HVLT free recall and the MMSE for participants with all dementias versus no dementia.

Table 2
Dementia sensitivity and specificity of HVLT free recall scores with corresponding probabilities of dementia (PPV) and probabilities of no dementia (NPV) at different base rates

All dementia			PPV/NPV at different base rates			
HVLT	Sensitivity	Specificity	5%	10%	15%	20%
<3	0.02	1.00	1.00/.95	1.00/.90	1.00/.85	1.00/.80
<10	0.39	0.99	.67/.97	.81/.94	.87/.90	.91/.87
<11	0.44	0.99	.70/.97	.83/.94	.89/.91	.92/.88
<12	0.50	0.93	.27/.97	.44/.94	.56/.91	.64/.88
<13	0.57	0.91	.25/.98	.41/.95	.53/.92	.61/.89
<14	0.62	0.91	.27/.98	.43/.96	.55/.93	.63/.91
<15	0.68	0.90	.26/.98	.43/.96	.55/.94	.63/.92
<16	0.83	0.83	.20/.99	.35/.98	.46/.97	.55/.95
<17	0.88	0.69	.13/.99	.24/.98	.33/.97	.42/.96
<18	0.90	0.66	.12/.99	.23/.98	.32/.97	.40/.96
<19	0.94	0.54	.10/.99	.19/.99	.27/.98	.34/.97
<20	0.97	0.46	.09/1.0	.17/.99	.24/.99	.31/.98
<21	0.99	0.39	.08/1.0	.15/1.0	.22/1.0	.29/.99
<22	0.99	0.30	.07/1.0	.14/1.0	.20/1.0	.26/1.0
<25	1.00	0.06	.05/1.0	.11/1.0	.16/1.0	.21/1.0
<29	1.00	5.00	.05/1.0	.10/1.0	.16/1.0	.21/1.0
<31	1.00	3.00	.05/1.0	.10/1.0	.15/1.0	.20/1.0
<35	1.00	0.00	.05/1.0	.10/1.0	.15/1.0	.20/1.0

As the cut score is raised sensitivity rises while specificity falls The sensitivity, specificity, PPV and NPV of each HVLT free recall cut score for different prevalence rates of dementia are shown in Table 2. At a cut score of <16, the HVLT has a sensitivity of 0.83 and specificity of 0.83.

We also examined the HVLT recognition scores for discriminating dementia. The AUC is 0.69 and sensitivity is 0.50 when the specificity is 0.80. When the HVLT true positive recognition score is entered into the logistic regression with the HVLT recall score, the area under the ROC curve does not change. When we applied an adjusted recognition score by subtracting the related errors or total errors from the true positive score, neither the AUC nor sensitivity or specificity changed.

We conclude that the HVLT recognition score does not improve the identification of dementia above the free recall score from the HVLT in our sample.

Figure 1 also shows the sensitivity–specificity trade-offs of different cut scores on the MMSE for discriminating the group with all dementias from those with no dementia. The total AUC for the MMSE is 0.83. The sensitivity, specificity, PPV and NPV of each MMSE cut score for different prevalence rates of dementia are shown in Table 3. For example, at a cut score of <25, the MMSE has a sensitivity of 0.70 and specificity of 0.82. Sensitivity rises to 0.86 at a cut score of <26 and specificity increases to 0.89 when the cut score is <24 (see Table 3).

We compared the AUC for the HVLT (0.89) recall and the MMSE (0.83) and found no statistically significant differences using the method proposed by Metz and Pan (1999). Using

Table 3
Dementia sensitivity and specificity of MMSE scores with corresponding probabilities of dementia (PPV) and probabilities of no dementia (NPV) at different base rates

All dementia			PPV/NPV at different base rates			
MMSE	Sensitivity	Specificity	5%	10%	15%	20%
<19	0.14	1.00	1.00/.96	1.00/.91	1.00/.87	1.00/.82
<20	0.17	0.99	.47/.96	.65/.91	.75/.87	.81/.83
<21	0.29	0.98	.43/.96	.62/.93	.72/.89	.78/.85
<22	0.33	0.97	.37/.96	.55/.93	.66/.89	.73/.85
<23	0.43	0.93	.24/.97	.41/.94	.52/.90	.61/.87
<24	0.53	0.89	.20/.97	.35/.94	.46/.91	.55/.89
<25	0.70	0.82	.17/.98	.30/.96	.41/.94	.49/.92
<26	0.86	0.70	.13/.99	.24/.98	.33/.97	.42/.95
<27	0.93	0.48	.09/.99	.17/.98	.24/.97	.31/.96
<28	0.99	0.23	.06/1.0	.13/1.0	.18/.99	.24/.99
<29	1.00	0.10	.06/1.0	.11/1.0	.16/1.0	.22/1.0
<30	1.00	0.04	.05/1.0	.10/1.0	.16/1.0	.21/1.0
=30	1.00	0.04	.05/1.0	.10/1.0	.15/1.0	.20/1.0

the alternative method, the McNemar test (Fleiss, 1981) demonstrated that when various sensitivity rates were held constant, there were no statistical differences in specificity.

3.3. Alzheimer's disease (AD)

As shown in the all-dementia analyses, age, sex and education were entered into all of the above logistic regression models and they did not significantly influence the performance of either the HVLT or the MMSE, optimal cut scores for detecting AD were not modified. However, WRAT-R reading ability was significant ($P < .02$) when added to the MMSE model for discriminating those with AD from those with no dementia; the ROC curve was shifted slightly to the left (not shown). When reading ability is examined for those with an eighth grade reading level or less, the ROC curve for the poor readers shows a modest increase in specificity for a given value of sensitivity.

Figure 2 shows the sensitivity–specificity trade-offs of different cut scores on the free recall portion of the HVLT and the MMSE for discriminating those with AD only from those with no dementia. The total AUC for the HVLT is 0.89.

At a cut score of <16, the HVLT has a sensitivity of 0.83 and specificity of 0.83. Sensitivity rises to 0.88 at a cut score of <17 but specificity drops to 0.67. When the cut score is <15, specificity increases to 0.92 but sensitivity decreases to 0.75. The ROC curve that describes the discrimination of those with AD only from those with no dementia by HVLT recognition has an AUC of 0.70; at a sensitivity of about 80% when the specificity is about 0.30. The area under the ROC curve for total recall does not change by adding any of the other adjusted recognition scores to the model.

Figure 2 also shows the sensitivity–specificity trade-offs of different cut scores on the MMSE for discriminating those with AD only from those with no dementia. The total AUC is 0.85.

Sample Size = 380   Base Rate = .15

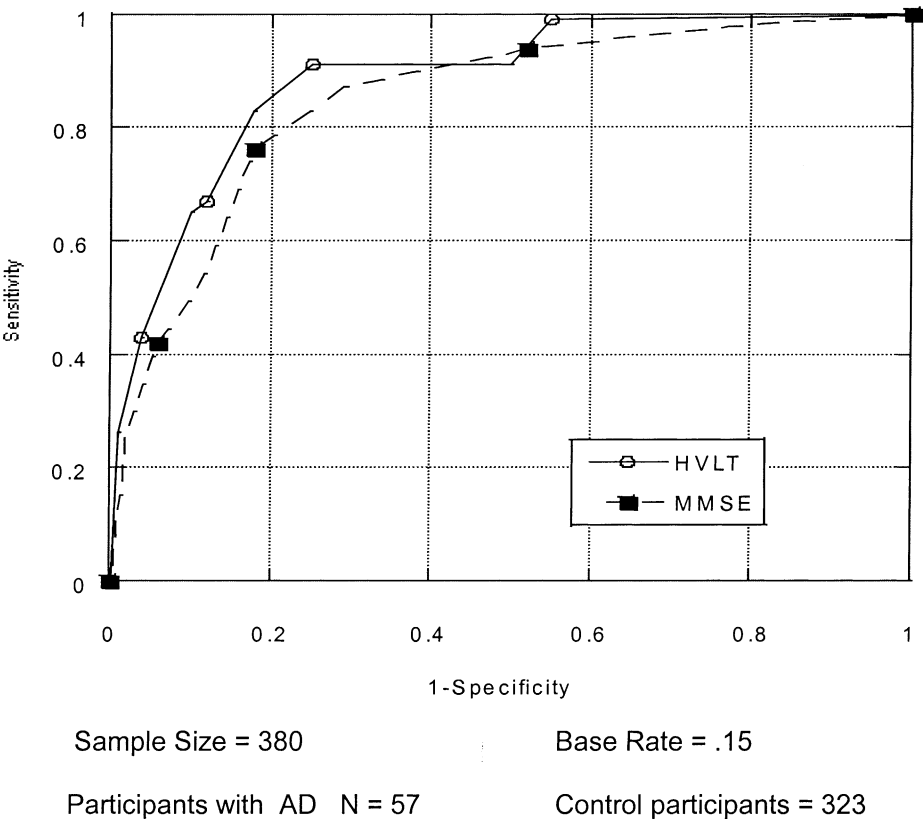Participants with AD   N = 57   Control participants = 323

Fig. 2. HVLT free recall and the MMSE for AD versus no dementia.

At a cut score of <25, the MMSE has a sensitivity of 0.75 and specificity of 0.82. Sensitivity rises to 0.88 at a cut score of <26 but specificity falls to 0.70. At a cut score of <24, specificity rises to 0.89 but sensitivity drops to 0.52. Adding reading ability to the MMSE model slightly improves discrimination with modest increases in specificity at the relevant cut scores of 24 and 25.

We compared the AUCs for the HVLT (0.89) recall and the MMSE (0.85) for participants with AD only and found no statistically significant differences.

3.4. Vascular dementia (VaD)

Age, sex and education were entered into logistic regression models for discriminating persons with VaD from persons with no dementia. They did not influence HVLT or MMSE performance significantly and they did not modify cut scores for detecting VaD. Unlike the results for AD, adding reading ability to the MMSE model did not affect the results or the cut scores. However, reading ability did enter significantly into the logistic regression for discriminating those with VaD from those with no dementia and while not changing the optimal cut score, it significantly improved specificity. When reading ability is dichotomized above

Sample Size = 342   Base Rate = .06
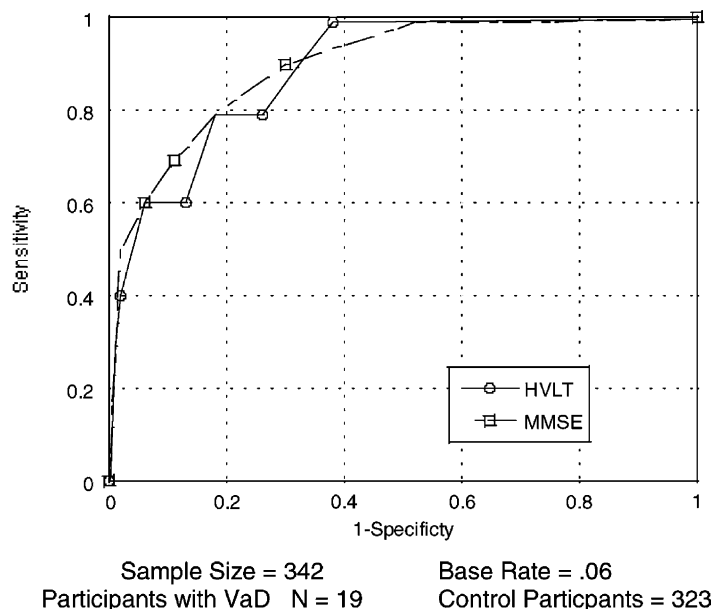Participants with VaD   N = 19   Control Particpants = 323

Fig. 3. HVLT free recall and the MMSE for VaD versus no dementia.

and below eighth grade, the ROC curve for the poorer readers was shifted considerably right when compared to the ROC curve of the better readers (i.e., greater sensitivity and specificity for better readers).

Figure 3 shows the sensitivity–specificity trade-offs of different cut scores on the free recall portion of the HVLT and the MMSE for discriminating those with VaD from those with no dementia. The total AUC is 0.90 for HVLT and 0.95 for the combined HVLT and reading ability (WRAT-R).

At an optimal cut score of <16, the HVLT alone has a sensitivity of 0.80 and specificity of 0.84. Holding sensitivity constant at 0.80, the combined HVLT and WRAT-R reading score have a specificity of 0.94. For HVLT alone, sensitivity rises to 0.90 at a cut score of <18 but specificity decreases to 0.68; however, when the HVLT and WRAT-R reading are combined, the specificity increases to 0.89 at a sensitivity of 0.90.

HVLT recognition has an area under the ROC curve of 0.74 and sensitivity is only 0.50 when the specificity is 0.90. When the HVLT true positive recognition score is entered into the logistic regression with the HVLT recall score, the area under the ROC curve does not change from the total recall score. However, adding true positive recognition maintains sensitivity at 0.90 and increases specificity modestly to 0.73. When we applied an adjusted recognition score by subtracting the related errors or total errors from the true positive score, neither the AUC nor sensitivity or specificity improved.

Figure 3 also shows the sensitivity–specificity trade-offs of different cut scores on the MMSE for discriminating those with VaD from those with no dementia. The total AUC is 0.91. At a cut score of <26, the MMSE has a sensitivity of 0.89 and specificity of 0.70. At a cut score of <25 for VaD, sensitivity is 0.75 and specificity increases to 0.82.

We compared the AUCs for the HVLT (0.89) recall alone and the MMSE (0.91) for those with VaD and found no statistically significant differences. When the ROC curves of the HVLT in combination with the WRAT-R reading scores was compared to the MMSE ROC curve, the McNemar's test (Fleiss, 1981) indicated that the HVLT–WRAT-R combination was significantly better than the MMSE ($P < .01$) at discriminating persons with VaD from persons without dementia.

4. Discussion

These results indicate that the HVLT and the MMSE are effective tests for detecting dementia overall as well as the AD and VaD subgroups, in an ethnically diverse community-based sample. For these study samples, logistic regression analyses found all results to be independent of the effects of demographic variables, i.e., sex, age, ethnicity and education, suggesting that no age- or education-corrections are needed for either the HVLT or the MMSE.

We determined that the optimal cut score in our sample with a dementia base rate of 18% for the HVLT is 15, 16, or 17 depending on the application and whether sensitivity or specificity is of paramount importance. Our results also suggest that the HVLT alone works equally well for both AD and VaD. When the dementia group is limited to 63 individuals with "mild dementia" defined as MMSE >18 (37), the optimal cut is 15, with sensitivity of 0.81 and specificity of 0.83, lower than the 18/19 cut obtained by Frank and Byrne (2000) with a small sample of 15 mildly demented patients and 15 normal controls. Using the mild dementia cuts specified by Frank and Byrne (2000), we obtained an optimal cut of 25 for the HVLT (with sensitivity of 0.84 and specificity of 0.70) comparable to the 25/26 cut obtained by Frank and Byrne (2000).

The sum of three free recall trials outperforms the recognition on the HVLT for the discrimination of the all-dementia groups from the no-dementia control sample. This remains true for the discrimination of the two specific dementia subtypes, AD and VaD. The ROC curves in Figures 1 and 2 suggest that, at high values of specificity, the HVLT free recall provided slightly higher values of sensitivity than the MMSE for participants with all dementias and for those with AD, though the differences were small and not statistically significant. In a larger sample the modest differences may reach statistical significance. Further research in other community settings may clarify the relationship of the HVLT and MMSE in dementia subtypes. Combining HVLT free recall with either HVLT recognition did not significantly improve discrimination of those with dementia from those without dementia.

HVLT performance is compromised in persons with low reading ability and clinicians will have to take reading ability into account when interpreting HVLT scores. In our sample, it appeared that the reading ability level of many of the individuals with VaD was compromised. Additional studies are needed to clarify the relationship between locus of vascular lesion, reading and semantic memory tests.

While the sensitivity, specificity, PPV and NPV of the HVLT in this sample appear modest compared with previous studies, it is important to note that the no-dementia control group includes persons with mild memory impairments and individuals with other mild cognitive deficits who may be in the preclinical stage of dementia, although they do not meet clinical criteria for a diagnosis of full-blown dementia. We deliberately did not exclude these individ-

uals since any community-based population will include these persons and they are representative of the elderly population at large.

The effect of depression on free recall has been reported (Blau & Ober, 1988; Breslow, Kocsis, & Belkin, 1981; Davis & Unruh, 1980; Massman, Delis, Butters, Dupont, & Gillin, 1992). While the participants with dementia had slightly higher Geriatric Depression Scale (GDS) scores, depression did not significantly enter into the regression models for this study sample. However, depression must be always examined as it may mimic dementia (pseudodementia (Caine, 1981; Gainotti & Marra, 1994), may be a precursor to dementia (Godwin-Austen & Bendall, 1990; Liston, 1978; Roth, 1980), or be co-morbid with dementia (Mendez, Martin, Smyth & Whitehouse, 1990; Teri & Wagner, 1992). As depression is treatable, some of the cognitive deficits noted may be alleviated when depression is treated (Hoch & Reynolds, 1990).

The choice of an appropriate screening measure for identifying dementia depends on the question being asked and the sample studied. In geriatric or neurological clinic samples with high prevalence of patients with stroke, Parkinson's disease or other muscular and/or neurological disorders, the MMSE may not be ideal as it requires reading, following commands, and writing ability that may be compromised in these diseases. For these populations, the HVLT may be more appropriate for identifying memory deficits associated with dementing processes. However, for screening and identifying impairments in any of several cognitive domains and to monitor progression of dementia, the multi-task MMSE may be more appropriate.

We have provided the PPV and NPV of a range of scores on the MMSE and HVLT at different base rates of dementia, to be used depending on the clinician's purposes. For screening and identifying elderly individuals for further clinical and neuropsychological evaluation, a test or a cut score on the test that maximizes sensitivity may be chosen. On the other hand, in order to identify patients with dementia for pharmacological interventions with potential side effects, a test or a cut score on the test with high specificity may be picked. The NPV of the HVLT and the MMSE are very high, suggesting that we can be confident in reassuring those persons who screen negative on these measures; and further neuropsychological evaluation may be avoided. However, the modest sensitivities and PPV's of the HVLT and MMSE suggest that many individuals with significant memory deficits would be missed on either test. For example, a cut score of 25 on the MMSE would misclassify 30% of impaired individuals as unimpaired. Based on our findings, a high index of suspicion with a passing score on a single measure would require additional testing as we have determined that combining two memory tests with a logical 'OR' increases sensitivity with modest loss of specificity (Kuslansky et al., 2002). While revised versions of the HVLT have been described, the additional time commitment makes it impractical to use in busy clinic or high volume community screening. Further research is needed to develop shorter, efficient versions or combination with other tests to improve the sensitivity and PPV of the HVLT.

References

Albert, M., Smith, L. A., Scherr, P. A., Taylor, J. O., Evans, D. A., & Funkenstein, H. H. (1991). Use of brief cognitive tests to identify individuals in the community with clinically diagnosed Alzheiemer's disease. *International Journal of Neuroscience*, *57*, 167–178.

Altman, D. A. (1991). *Practical statistics for medical research*. London: Chapman & Hall.

American Psychiatric Association. (1997). *The diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Beardsall, L., & Huppert, F. A. (1991). A comparison of clinical, psychometric and behavioural memory tests: Findings from a community study of the early detection of dementia. *International Journal of Geriatric Psychiatry*, *6*, 295–306.

Benedict, R. H. B., Schretlen, D. S., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test—Revised: Normative data and analysis of inter-form and test–retest reliability. *The Clinical Neuropsychologist*, *12*(1), 43–55.

Blau, E., & Ober, B. A. (1988). The effect of depression on verbal memory in older adults. *Journal of Clinical and Experimental Neuropsychology*, *10*, 81.

Blessed, G., Tomlinson, R., & Roth, M. (1968). The association between quantitative measures of dementia and of senile changes in the cerebral gray matter of elderly subjects. *British Journal of Psychiatry*, *114*, 797–811.

Breslow, R., Kicsis, J., & Belkin, B. (1981). The contribution of the depressive perspective to memory function in depression. *American Journal of Psychiatry*, *138*, 227–230.

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The Clinical Neuropsychologist*, *5*(2), 125–142.

Bryson, G., Bell, S., & Lysaker, P. (1997). Affect recognition in schizophrenia: A function of global impairment or a specific cognitive deficit. *Psychiatric Research*, *71*(2), 105–113.

Buschke, H., Kuslansky, G., Katz, M., Stewart, W. F., Sliwinski, M., & Eckholdt, H. M. (1999). Screening for dementia with the memory impairment screen (MIS). *Neurology*, *2*, 231–238.

Caine, E. D. (1981). Pseudodementia. *Archives of General Psychiatry*, *38*, 1359–1364.

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

Cammermeyer, M., & Evans, J. E. (1988). A brief neurobehavioral exam useful for early detection of postoperative complications in neurosurgical patients. *Journal of Neuroscience Nursing*, *20*(5), 314–323.

Cesura, A. M., Borroni, E., Gottowik, J., Kuhn, C., Malherbe, P., Martin, J., & Richards, J. G. (1995). Lazabemide for the treatment of Alzheimer's disease. *Advances in Neurology*, *80*, 521–528.

Chui, H. C., Victoroff, J. I., Margolin, D., Jagust, W., Shankle, R., & Katzman, R. (1992). Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers. *Neurology*, *54*, 473–480.

Chum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *Journal of the American Medical Association*, *269*, 2386–2421.

Cullum, C., Thompson, L., & Smernoff, E. (1993). Three-word recall as a measure of memory. *Journal of Clinical and Experimental Neuropsychology*, *15*, 321–329.

Davis, H., & Unruh, W. R. (1980). Word memory in non-psychotic depression. *Perceptual and Motor Skills*, *51*, 699–705.

Elwood, R. (1993). Clinical discriminations and neuropsychological tests, an appeal to Bayes' theorem. *The Clinical Neuropsychologist*, *7*, 224–233.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Folstein, M., Folstein, S., & McHugh, P. (1975). A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Frank, R. M., & Byrne, G. J. (2000). The clinical utility of the Hopkins Verbal Learning Test as a screening test for mild dementia. *Journal of Geriatric Psychiatry*, *15*(4), 317–324.

Fugita, M., Woods, S. W., Verhoeff, N. P., Abi-Dargham, A., Baldwin, R. M., Zobhbi, et al. (1999). Changes of benzodiazepine administration in humans. *European Journal of Pharmacology*, *368*(2/3), 161–172.

Gainotti, G., & Marra, C. (1994). Some aspects of memory disorders clearly distinguish dementia of the Alzheimer's type from depressive pseudo-dementia. *Journal of Clinical and Experimental Neuropsychology*, *16*(1), 66–78.

Gauthier, S., Thal, L. J., & Rossor, M. (1996). The future diagnosis and treatment of Alzheimer's disease. In S. Gauthier (Ed.), *Clinical diagnosis and management of Alzheimer's disease*. London: Martin Dunitz.

Gelmacher, D. S. (1997). Donepezil (Aricept) therapy for Alzheimer's disease. *Comprehensive Therapy*, *23*(7), 492–493.

Godwin-Austen, R., & Bendall, J. (1990). *The neuropsychology of the elderly*. New York: Springer.

Guskiewicz, K. M., Riemann, B. L., Perrin, D. H., & Nashner, L. M. (1997). Alternative approaches to the assessment of mild head injury in athletes. *Medicine & Science in Sports & Exercise*, *29*(Suppl. 7), S213–S221.

Harris, M. J., Heaton, R. K., Schalz, A., Bailey, A., & Patterson, T. L. (1997). Neuroleptic dose reduction in older psychotic patients. *Schizophrenia Research*, *27*(2/3), 241–248.

Hoch, C. C., & Reynolds, C. F. (1990). Psychiatric symptoms in dementia: Interaction of affect and cognition. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 4). Amsterdam: Elsevier.

Hogervorst, E., Combrinck, M., Lapuerta, P., Rue, J., Swale, K., & Budge, M. (2001). The Hopkins Verbal Learning Test and screening for dementia. *Dementia and Geriatric Cognitive Disorders*, *13*, 213–220.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Marten, R. L. I. (1982). A clinical scale for the staging of dementia. *British Journal of Psychiatry*, *140*, 566–572.

Jenkyn, L. R., Reeves, A. G., Warren, T., Whiting, R. K., Clayton, R. J., Moore, W. W., Rizzo, A., Tuzun, I. M., Bonnett, J. C., & Culpepper, B. W. (1985). Neurologic signs in senescence. *Archives of Neurology*, *42*, 1154–1157.

Khatchaturian, Z., Phelps, C. H., Buchholz, N. S. (1994). The prospect of developing treatments for Alzheimer's disease. In R. D. Terry, R. Katzman, & K. L. Bick (Eds.), *Alzheimer's disease*. New York: Raven Press.

Kittner, S. J., White, L. R., Farmer, M. E., Wolz, M., Kaplan, E., Moes, E., Brody, J. A., & Feinleib, M. (1986). Methodological issues in screening for dementia: The problem of education adjustment. *Journal of Chronic Disease*, *39*, 163–170.

Knopman, D. S., DeKosky, St. T., Cummings, J. L., Chui, H., Corey-Bloom, N., Relkin, N., Small, G. W., Miller, B., & Stevens, J. C. (2001). Practice parameter: Diagnosis of dementia (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, *56*, 1143–1153.

Knopman, D. S., & Ryberg, S. (1989). A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Archives of Neurology*, *46*, 141–145.

Kokmen, E., Smith, G., Petersen, R., Tangalos, E., & Ivnik, R. (1991). The short test of mental status: Correlations with standardized psychometric testing. *Archives of Neurology*, *48*, 728.

Kuslansky, G., Buschke, H., Verghese, J., Katz, M., Hall, C., & Lipton, R. (2002). Mild Cognitive Impairment (MCI): 1 Test or 2? *Annals of Neurology*, in press.

Liston, E. H. (1978). Diagnostic delay in presenile dementia. *Journal of Clinical Psychiatry*, *39*, 599–603.

Lovell, M. R., Iverson, G. L., Collins, M. W., McKeag, D., & Maroon, J. C. (1999). Does loss of consciousness predict neuropsychological decrements after concussion? *Clinical Journal of Sports Medicine*, *9*(4), 193–198.

Massman, P. J., Delis, D. C., Butters, N., Dupont, R. M., & Gillian, J. C. (1992). The subcortical dysfunction model of memory deficits in depression: Neuropsychological validation in a subgroup of patients. *Journal of Clinical and Experimental Neuropsychology*, *14*, 687–706.

Mattis S. (1976). Mental status examination for organic mental syndrome in the elderly patient. In L. Bellak & T. B. Karasu (Eds.), *Geriatric psychiatry*. New York: Grune & Stratton.

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

Mendez, M. F., Martin, R. J., Smyth, K. A., & Whitehouse, P. J. (1990). Psychiatric symptoms in Alzheimer's disease. *Journal of Neuropsychiatry and Clinical Neuroscience*, *2*, 28–33.

Mesulam, M.-M. (2000). *Principles of behavioral and cognitive neurology*. New York: Oxford University Press.

Metz, C. E., & Pan, X. (1999). 'Proper' binormal ROC curves: Theory and maximum likelihood estimation. *Journal of Mathematical Psychology*, *43*, 1–33.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., vanBelle, G., Fillenbaum, G., Mellits, E. D., & Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer' Disease (CERAD) Part I. Clinical and neuropsycholigical assessment of Alzheimer's disease. *Neurology*, *39*, 1159–1165.

Petersen, R. C. (1991). Memory assessment at the bedside. In T. Yanagihara & R. Petersen (Eds.), *Memory disorders*. New York: Dekker.

Petit, H., Bakchine, S., Dubois, B., Laurent, B., Montagne, B., Touchon, J., Robert, Plk., Vellas, J. M., Cogneau, J., Marin La Meslee, R., & Sorbe, G. (1998). Concensus statement of an interdisciplinary group of French

experts on modalities of diagnosis and medical treatment of Alzheimer's disease at a treatable stage. *Revue Neurologique (Paris)*, *154*, 432–438.

Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society*, *23*, 433–444.

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

Rasmusson, D. X., Bylsma, F. W., & Brandt, J. (1995). Stability of performance on the Hopkins Verbal Learning Test. *Archives of Clinical Neuropsychology*, *10*(1), 21–26.

Ross, G. W., Abbott, R. D., Petrovitch, H., Masali, K. H., Murdaugh, C., Trockman, C., Curb, J. D., & White, L. R. (1997). Frequency and characteristics of silent dementia among elderly Japanese–American men. *Journal of the American Medical Association*, *277*, 800–805.

Roth, M. (1980). Aging of the brain and dementia: An overview. In L. Amaducci, A. N. Davison, & P. Antuono (Eds.), *Aging of the brain and dementia*. New York: Raven Press.

Schmitt, F. A., Ramseen, J. D., & DeKosky, S. T. (1989). Cognitive mental status examinations. In Pirozzolo, F. J. (Ed.), *Clinics in geriatric medicine* (Vol. 5, No. 3). Philadelphia: WB Saunders.

Shapiro, A. M., Benedict, R. N., Schretlen, D., & Brandt, J. (1999). Construct and concurrent validity of the Hopkins Verbal Learning Test—revised. *The Clinical Neuropsychologist*, *13*(3), 348–358.

Streiner, D. L., Norman, G. R., & Blum, H. M. (1989). *PDQ epidemiology*. Ontario, Canada: Decker, Inc.

Strub, R., & Black, F. (1985). *The mental status examination in neurology*. Philadelphia: F.A. Davis.

Teng, E., Chiu, H., Sschneider, L., & Metzger, L. (1987). Alzheimer's dementia: Performance on the Mini-Mental State Examination. *Journal of Consulting and Clinical Psychology*, *55*, 96–100.

Teri, L., & Wagner, A. (1992). Alzheimer's disease and depression. *Journal of Consulting and Clinical Psychology*, *60*, 379–391.

Tombaugh, T., McDowell, I., Krisjansson, B., & Hubley, A. (1996). Mini-Mental State Examination (MMSE) and the modified-MMSE (3MS): A psychometric comparison and normative data. *Psychological Assessment*, *8*, 48–59.

Trzepacz, P. T., & Baker, R. W. (1993). *The psychiatric mental status examination*. New York: Oxford University Press.

U.S. Agency for Health Care Policy and Research. Recognition and initial assessment of Alzheimer's disease and related dementias. Silver Springs, MD: Clinical Practice Guideline No. 19, 1996.

White, T., Bauer, R., Bowers, D., Crosson, B., & Kessler, H. (1995). Recall of three words after five minutes: Its relationship to performance on neuropsychological memory tests. *Applied Neuropsychology*, *2*, 130–138.