Neuropsychology Review, Vol. 5, No. 3, 1995

The California Verbal Learning Test: Psychometric Characteristics and Clinical Application

Richard W. Elwood[1,2]

The California Verbal Learning Test (CVLT) is a popular clinical and research test that claims to measure key constructs in cognitive psychology such as repetition learning, serial position effects, semantic organization, intrusions, and proactive interference. The psychometric characteristics of the CVLT are reviewed and related to the test's clinical utility. The utility of the CVLT is shown to be limited by its poor standardization and inflated norms. Further, the validity is limited because the CVLT uses multiple trials whereas the constructs it purports to measure are based on single-trial paradigms. The review proposes modifications to the CVLT and guidelines for its clinical use. It concludes that if the limitations of the CVLT are recognized, it can still make a useful contribution to the clinical assessment of verbal learning and memory.

KEY WORDS: California Verbal Learning Test; neuropsychological tests, memory, learning.

INTRODUCTION

The California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987) is part of an ambitious project to develop neuropsychological tests based on the methods and constructs of cognitive science. Delis and his colleagues designed the CVLT to measure not only how much a subject learned but also to show *how* they learned, that is, what strategies or processes they used and what kinds of errors they made (Delis, Kramer, Fridland, & Kaplan, 1988). The CVLT attempts to measure a broad range of theoretical constructs from experimental cognitive psychology. These in-

[1]Veterans Affairs Medical Center, 500 East Veterans Street, Tomah, Wisconsin 54660.
[2]To whom correspondence should be addressed.

1040-7308/95/0900-0173$07.50/0

clude free and cued recall, serial position effects (including primacy and recency), semantic clustering, intrusions, interference, and recognition. Despite its popularity as a clinical and research measure, the CVLT has not been critically reviewed. This article evaluates the psychometric properties of the CVLT and relates them to the test's practical utility as a clinical instrument. The inordinate number of presented papers included in this review reflects the large, unpublished literature on the CVLT.

DESCRIPTION, ADMINISTRATION, AND SCORING

The CVLT is a multitrial recall and recognition word list learning test based on the popular Auditory Verbal Learning Test (AVLT; Rey, 1964). The examiner reads a list of 16 nouns aloud at one-second intervals in fixed order over five learning trials. After each trial, the subject is asked to recall as many words as s/he can in any order (i.e., free recall). The major feature of the CVLT is that the words are drawn from four semantic categories (tools, fruits, clothing, spices and herbs), with no consecutive words from the same category. If subjects "cluster" words from each category together, they are presumably using semantic organization to aid their recall. An interference list (list B) is then presented. List B shares two categories from List A (fruits, spices and herbs) and has two unshared categories (fish and kitchen utensils). Neither list uses the most common exemplars of each category in order to minimize ceiling effects and to be more sensitive to intrusion errors (Delis et al., 1987). Thus, plums are used rather than apples, nutmeg rather than salt. Free and cued recall of list A are tested immediately (short-delay) and again after 20 minutes (long-delay). On the cued recall trials, subjects must recall the list A words as the examiner specifies each category in turn. The CVLT ends with a recognition task. As each word on a 44-word list is read aloud, the subject must indicate whether it is a target word or a distractor. Some distractors share semantic categories with the target words while others sound alike. Both word lists on the CVLT are introduced as shopping lists. McCarthy, Ferris, Clark, and Crook (1981) earlier recommended a shopping-list format precisely because it is the kind of task that people often face in their everyday activities.

Unlike the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987), the CVLT instructions do not warn subjects of the delayed recall and recognition tasks. Hess, Colliver, Antonacci, and Sutton (1991) found no significant differences in CVLT scores between college students who were given the standard (no warning) instructions and those informed of the later delayed tasks. However, they also found a clear ceiling effect;

most of the subjects learned all the words. When Hess et al. (1991) replicated the experiment with a longer word list the informed group performed significantly better on both delayed recall and recognition measures. It thus seems reasonable to expect that advance warning would increase the delayed recall and recognition scores in clinical settings where there is less ceiling effect.

Most measures on the CVLT are easily scored by hand, although some calculations are so complicated that manual scoring is discouraged. A computer administration and scoring system (Fridlund & Delis, 1987) generates scores for every measure, graphs a learning curve, and provides learning parameters, response errors, and interference effects. Examiners can enter the subject's responses directly while administering the test but only after considerable practice and with exceptional visual-motor speed. I suspect most examiners are more cautious and prefer to transcribe the responses from the completed record form.

Several variations on the CVLT have been devised. Delis, McKee, et al. (1991) developed an alternate form, a worthwhile contribution because total recall shows a large practice effect on repeated clinical assessments (McCaffrey et al., 1995). Libon et al. (1993) devised a dementia version (CVLT-D) that uses a nine-word list and three semantic categories to avoid floor effects with more impaired individuals. The CVLT-D has been used to compare normal subjects with demented patients (Scott, Duff, Moczynski, Norton, & Gansler, 1995) and those with traumatic head injuries (Gansler, McGrath, Kaplan, Moczynski, & Norton, 1995). Heinrichs and Bury (1991) devised a word stem completion task using list A and 16 other words to test implicit memory among schizophrenic patients. In a study of early Parkinson's (PD) patients, Taylor, Saint-Cyr, and Lang (1990) replaced the usual aural recognition with a computer-based visual presentation. The 16 words from list A, 8 from list B, and 20 distractors are gradually unmasked. As soon as subjects recognize each word, they are asked to identify its source (list A, B, or neither). The CVLT has also been modified with Italian (Zappalá, Measo, Romani, & Merlin, 1992) and German (Köhler, 1994) language word lists. A childrens' version of the CVLT is also available (CVLT-C; Delis, Kramer, Kaplan, & Ober, 1993). The CVLT is included in the proposed National Institutes of Mental Health (NIMH) battery for the assessment of cognitive functioning in HIV+ patients (Butters et al., 1990).

NORMS AND RELIABILITY

Norms for seven groups covering the ages of 17 to 80 are provided in the manual (Delis et al., 1987) and scoring software (Fridlund & Delis, 1987). Pope (1986, 1987) collected data from older subjects (55 to 91), although these tentative norms have not been published or widely disseminated. The CVLT includes separate norms for males and females and women consistently show better recall on the CVLT (Delis et al., 1987; Kramer, Delis, & Daniel, 1988; Otto et al., 1994; Reite, Cullum, Stocker, Teale, & Kozora, 1993; Wiens, Tindall, & Crosson, 1994). The gender effect on recognition is less clear. Women outperformed men in one study (Wiens et al., 1994) but equalled them in another (Kramer, Delis, & Daniel, 1988). Performance on the CVLT is related to estimates of verbal IQ (Tindall & Bradley, 1991) and full-scale IQ (Wiens et al., 1994). Not surprisingly, recall and recognition on the CVLT are related to education (Otto et al., 1994). Wiens et al. (1994) found significant examiner effects on numerous CVLT variables. Because the word lists are read aloud, this variability may reflect differences in speech characteristics.

The original CVLT norms are clearly inadequate. The non-clinical reference group consists of only 273 subjects who volunteered for research projects in four U.S. cities. With no effort toward unbiased sampling and an average educational level of nearly 14 years, the reference group does not represent the national population. The publisher rightly describes the CVLT norms as "preliminary" and refers to the test as the "research edition" (Psychological Corporation, 1994). Moreover, the computer scoring software (Fridlund & Delis, 1987) refers to the "limited" reference group and cautions against drawing normative inferences from CVLT scores. The manual anticipates a standardization of the CVLT at a later data.

Recent studies show that the standard scores in the manual are grossly inflated. Wiens et al. (1994) administered the test to 700 civil service job applicants, all of whom had been screened for alcohol and other substance abuse and passed tests of academic skills and physical agility. Even with an average IQ of 106 and 14.5 years of education, the group still performed slightly *below* the norms in the CVLT manual. Randolph et al. (1994) compared the CVLT with the Wechsler Memory Scale-Revised (WMS-R) among older, well-educated volunteers. Judged by the respective test norms, subjects performed more than a standard deviation above average on the WMS-R but were only average on the CVLT. The discrepancy was even greater in a clinical sample, where mean scores were a standard deviation below normal on the WMS-R but *three* standard deviations low on the CVLT. Stallings, Boake, and Sherer (1993) compared the CVLT with the Rey AVLT among head injured patients. Although the raw scores of the two tests were almost equivalent, the CVLT standard scores were significantly lower. D'Elia, Boone, and Mitrushina (1995) summarize the available CVLT norms.

The only reliability data provided in the CVLT are internal consistency coefficients for total recall and test-retest correlations for selected indices based on a subsample of only *21* subjects. The reliability of the other CVLT indices is simply unknown. The manual provides standard errors of measurement error only for total recall of List A. I have argued elsewhere (Elwood, 1991) that due to their inherent measurement error, scores on neuropsychological tests should be described (and interpreted) in terms of confidence intervals rather than absolute values. The CVLT manual does not discuss confidence intervals, though it does provide the information needed to calculate intervals for total recall. Take the case of a 67 year old man who obtains an average agewise score of 46 (Delis et al., 1987, Table A-1, *T* score = 50). Confidence intervals are usually established by adding and subtracting a *z* score multiple of the standard error of measurement to the subject's score. According to the manual, the standard error of measurement (SE_M) for total recall over the five learning trials is 5.07 or 5.21, depending on whether reliability is based on odd-even correlation or Cronbach's Alpha (Table 3). The standard (*z*) score multiple for a 90% confidence level is 1.64. The interval is 46 ± (1.64 SE_M), or 37 to 55, which is equivalent to a *T*-score range of 39 to 60, which corresponds to the 14th to 84th percentiles. With a 90% confidence level, the subject's true score could be considered to be in the low-average to high-average ranges.

Although this method of establishing confidence intervals of test scores is commonly used, it is technically incorrect. Confidence intervals should be centered not on the score a subject actually obtains on the test but on an estimated *true* score that accounts for regression toward the test mean (Glutting, McDermott, & Stanley, 1987). The effect of regression is more pronounced as the reliability of the test declines and as the actual score deviates further from the mean. Thus sizeable differences between actual and estimated true scores can be expected on the CVLT; total recall has only modest reliability and clinical scores are often well below normal. In general, true score estimates and their corresponding confidence intervals can be easily calculated from a test's reliability and mean (Elwood, 1991). However, the manual gives reliability and SE_M values only for total recall. The reliability of the other CVLT measures is unknown.

FACTOR STRUCTURE

The original factor study of the CVLT analyzed 19 separate age-corrected scores and identified six principal components among normal subjects and five in a mixed neurologic sample (Delis, Freeland, Kramer, & Kaplan, 1988). Subsequent replications using the same variables identified five to seven major components based on normal subjects (Wiens et al., 1994) and clinical samples (Schear & Craft, 1989b; Vanderploeg, Schinka, & Retzlaff, 1994; Wilde & Boake, 1993). However, most of the studies used Kaiser's K1 rule (Kaiser, 1960) to determine the number of principal components. The K1 rule, which retains each component whose eigenvalue is greater than one, overestimates the number of population components (Zwick & Velicer, 1986). Applying the more conservative scree test (Cattell, 1966) to their clinical sample, Vanderploeg et al. (1994) identified five CVLT components. The minimum average partial correlation (MAP) method (Velicer, 1976) is more accurate and less variable than either the K1 or scree criteria (Zwick & Velicer, 1986). When applied to the correlation matrix from Schear and Craft (1989b), the MAP method retains only a single component.[3] The number of components may also be inflated by using experimentally dependent variables (Comrey & Lee, 1992). Although Delis, Freeland, et al. (1988) claim the variables in their (and subsequent) factor analyses are not "monotonic transformations of one another" (p. 41), several variables are clearly dependent (e.g., recall on list B and the difference between recall on list B and recall on the last trial of list A). Wilde and Boake (1993) found that only the first (general verbal learning) component correlated with other neuropsychological measures (e.g., WMS-R, GOAT, AVLT, WCST) and suggest that six-factor solutions may be artifacts of variable selection. That argument is bolstered by a recent factor analysis of the AVLT. When Vakil and Blachstein (1993) selected only those variables that were not linear composites of one another, they found that even the liberal K-1 rule retained only one component.

Discounting redundant variables and applying more rigorous factor retention criteria suggests that the CVLT can be represented by one to three major components and perhaps two or three minor components. A general verbal learning component consistently accounts for about 35-40% of the total variance and consists of total free recall over the five trials of list A, semantic clustering, free and cued recall (both short- and long-delays), and recognition hits. A second, "response discrimination," component has also been found in most studies (as the third component in Schear & Craft, 1989b). It accounts for about 8-10% of the variance with loadings from free and cued recall intrusions and recognition false positives. The remaining components, learning strategy (semantic and serial clustering), serial position (primacy and recency), and proactive effect (List B recall) are inconsistent and account for little additional variance.

[3]Using the Principal Components Analysis Package (PCA_PACK; Reddon, 1992). The output from the analysis can be obtained from the author.

LEARNING

The primary and most studied measure of learning on the CVLT is the total recall over the five learning trials of list A. Recall on the CVLT declines with age (Delis et al., 1987; Taylor et al., 1990) and is lower in men (Kramer, Delis, & Daniel, 1988; Reite et al., 1993). Lowered recall has also found across a wide range of clinical groups, including Alzheimer's disease (Bondi et al., 1994; Cullum, Filley, & Kozora, 1995; Kohler, 1994; Massman, Delis, & Butters, 1993; Kramer, Delis, Blusewicz, et al., 1988; Kramer, Levin, Brandt, & Delis, 1989; Simon, Leach, Winocur, & Moscovitch, 1994; Stout et al., 1995), Huntington's disease (Johnson et al., 1994; Kramer, Delis, Blusewicz, et al., 1988; Massman, Delis, Butters, Levin, & Salmon, 1990), and Parkinson's disease (Massman et al., 1990; Taylor et al., 1990), closed head injury (Crosson, Novack, Trennery, & Craig, 1988; Novack, Kofoed, & Crosson, 1995; Numan & Sweet, 1992; Stallings et al., 1993, Tibby, Schmitter-Edgecombe, & Long, 1994; Zappalá & Trexler, 1992), open head injury (Millis & Ricker, 1994b), temporal lobectomy (Herman, Wyler, Bush, & Tabatabai, 1992; Herman, Wyler, & Somes, 1993), lucunar infarcts (Wolfe, Linn, Babikian, Knoefel, & Alberte, 1990), right hemisphere CVA (Welte, 1993), multiple sclerosis (Kessler, Cohen, Lauer, & Kausch, 1992), alcoholism (Bondi, Drake, Butters, & Griffith; Kramer, Blusewicz, & Preston, 1989), Korsakoff's syndrome (Heinrichs, 1994), cocaine dependence (Mittenberg & Motta, 1993), left side complex partial seizures (Hermann, Wyler, Richey, & Rea, 1987; Hermann, Wyler, Steenman, & Richey, 1988), anterior communicating artery aneurysm (Diamond & DeLuca, 1993), schizophrenia (Cullum et al., 1989; Heinrichs, 1994; Paulsen et al., 1995; Saykin et al., 1994), and encephalopathy from Lyme disease (Kaplan, Meadows, Vincent, Logigian, & Steere, 1992). Lower CVLT scores among subjects with active AIDS have been found in one study (Peavy et al., 1994) while differences in another were not significant (Saykin et al., 1988). Total recall has been found to parallel the recovery of patients with severe traumatic brain injuries (Novack, Kofoed, & Crosson, 1995).

The various recall measures on the CVLT correlate highly with each other (Delis et al., 1987; Delis, Freeland, et al., 1988) and with the general memory index on the Wechsler Memory Scale-Revised (Schear & Craft, 1989a). In fact, recall on trial 5 alone has actually been found to correlate higher with the WMS-R than total recall over all five trials (Delis, Cullum, Butters, & Cairns, 1988). Perrine, Devinsky, Donofrio, and Luciano (1994) compared recall on the CVLT and the Selective Reminding Test (SRT; Buschke & Fuld, 1974) with the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983). They found that the Boston Naming Test correlated much more with the CVLT (r = .58) than with the SRT and suggest the CVLT requires more word finding ability because of its less common words. Among subjects with closed head injuries, recall on the CVLT is related to language measures like confrontation naming and verbal fluency (Crosson, Cooper, Lincoln, Bauer, & Velozo, 1993).

Two other learning parameters are discussed in the manual and calculated by the computer scoring system. The learning slope (of the regression line of the five recall scores) is generally considered a measure of acquisition rate. However, learning slope has been largely ignored in studies of the CVLT, although lower scores were reported in one sample of Huntington's patients (Massman et al., 1990). Recall consistency is the percent of words recalled on one trial that are also recalled on the next. Few studies have compared consistency on the CVLT, although lower scores have been found among Parkinson's, Huntingtons's (Massman et al., 1990), and schizophrenic subjects (Cullum et al., 1989; Paulsen et al., 1995). Consistency is apparently unrelated to gender (Kramer, Delis, & Daniel, 1988) or IQ (Tindall & Bradley, 1991).

SERIAL POSITION EFFECTS

Primacy and Recency

Severely reduced primacy and increased recency on the CVLT have been found in Huntington's (HD) (Massman et al., 1990) and Alzheimer's (AD) disease (Massman et al., 1993). However, Massman et. al. (1993) found that HD (but not AD) patients *recognized* primacy words as well as the controls, suggesting that their impaired primacy was not due to poor long-term storage. The authors propose that deficient primacy recall on the CVLT may indicate impaired long-term storage only with some patients (e.g., AD) but not with others (e.g., HD). Kramer, Delis, and Daniel (1988) found that among normal subjects, although women had better recall, men had higher recency *and* primacy scores. Also, Schear and Craft (1989b) found that while recency was moderately and inversely related (r = −.47) to total recall in their neuropsychiatric sample, primacy was unrelated. Greater recency has been found among schizophrenic patients (Cullum et al., 1989) and older subjects with a family history of progressive dementia (Bondi et al., 1994). Among depressed patients, those who complain of cognitive problems have shown lower primacy on the CVLT but equivalent recency (Hill, Stourdemire, Morris, Martino-Saltzman, & Markwatler, 1992). Reduced recency (but not primacy) on the CVLT has also been found in patients with right hemisphere CVAs (Welte, 1993).

Kozora, Cullum, Smernoff, and Allensworth (1992) compared serial position effects on the CVLT and the Logical Memory (paragraph recall) from the Wechsler Memory Scale-Revised (WMS-R) in a sample of older subjects. They found more primacy effect on the WMS-R than on the CVLT and speculate that stories like those on the WMS-R have more semantic organization than word lists like those on the CVLT. A more parsimonious explanation is the difference in test instructions. The WMS-R manual directs subjects to "begin at the beginning" (Wechsler, 1987, p. 19) while the CVLT asks them to say the words "in any order" (Delis et al., 1987, p. 7).

Primacy and recency on the CVLT are confusing because the same terms are used to refer to both the traditional single-trial paradigm and the multitrial format used in the CVLT (Delis et al., 1987; Massman, Delis, & Butters, 1993). The CVLT manual acknowledges the ambiguity and in one section refers to "pooled recency" and "pooled primacy," although this convention is not widely used. Several studies have found that the robust primacy and recency effects observed in the traditional single-trial paradigm do not extend to the "pooled" scores on the CVLT. For example, although primacy generally decreases with age (Davis et al., 1990), primacy on the CVLT does not show a corresponding agewise decline (Delis et al., 1987). Further, primacy effect should predict overall recall (Delis, Freeland, et al., 1988) but Kozora, Cullum, Smernoff, and Allensworth (1992) found that neither recency or primacy were related to total recall among older normal controls. Further, Kramer, Delis, and Daniel (1988) found that although women consistently showed better recall, men had higher recency *and* primacy scores.

Serial Clustering

Serial clustering on the CVLT is scored over the five trials of List A whenever the subject recalls two words in the same order as they are presented. Because clustering may occur by chance alone, the clustering measure is based on the ratio of observed to expected scores (Shuell, 1969). Frankel and Cole (1971) review various clustering statistics and recommend a standard score ratio so that clustering can be compared between different tests. However, the method used on the CVLT seems adequate for clinical applications. Clustering ratios are difficult to calculate by hand but are provided by the computer scoring system. Only a few studies have examined serial clustering scores and these have not found significant differences between normal subjects and those with dementia (Bondi et al., 1994; Hill et al., 1993), schizophrenia (Cullum et al., 1989; Cullum et al., 1995; Paulsen et al., 1995), or AIDS (Peavy et al., 1994). Although Novack et al.

(1995) found that the recall scores of patients with traumatic brain injuries improved when tested several months after their injury, they found no significant gains in primacy, recency, or serial clustering scores.

SEMANTIC ORGANIZATION

Semantic Clustering

We know that material that is easy to organize is also easy to learn and that people who organize better also learn better. Semantic clustering in free recall has long been the experimental paradigm for semantic organization (Blousfield, 1953; Tulving, 1962). Semantic clustering on the CVLT is scored each time a subject recalls two successive words from the same category. Clustering is totalled over all free recall trials of both the target and interference word lists (it cannot be scored on cued recall because the cues are the categories themselves). Because semantic clustering can occur by chance alone, depending on how many words a subject generates on a given trial, the overall score employs the same observed/expected ratio used to measure serial clustering. The expected semantic clustering score (and thus the clustering ratio) is hard to calculate by hand but is provided by the scoring software. The CVLT standard scores in the manual reflect the expected agewise decline in semantic clustering among normal individuals (Craik, 1984). Semantic clustering is lower in men and parallels lower recall in Alzheimer's disease (Bondi et al., 1994; Massman et al., 1993; Simon et al., 1994), Parkinsons's disease (Taylor et al., 1990), AIDS (Peavy et al., 1994), complex partial seizures (Hermann et al., 1987; Hermann et al., 1988), closed head injury (Crosson et al., 1988; Crosson, Sartor, Jenney, Nabors, & Moberg, 1993; Novack et al., 1995), penetrating head injury (Millis & Ricker, 1994b), and schizophrenia (Cullum et al., 1989; Heinrichs, 1994; Paulsen et al., 1995). The one comparison between alcoholics and normals found no significant difference (Kramer, Blusewicz, & Preston, 1989). Likewise, Hill et al. (1993) reported that semantic clustering did not discriminate subgroups of depressed patients with depression alone, mood-related cognitive dysfunction, or primary dementia. Novack et al. (1995) found that semantic clustering scores among head-injured patients showed significant gain several months after their injury.

If semantic organization is so helpful, recall should be better on the CVLT, which uses semantically related words, than on the similar Rey AVLT which does not. However, two studies that directly compared the tests showed no advantage. Crossen and Wiens (1994) found that normal subjects recalled an average of 55.1 words over the five learning trials on the CVLT versus 51.7 on the Rey. The authors attribute the difference to the lengths of the respective word lists, 16 words on the CVLT versus 15 on the AVLT. This explanation seems plausible because the mean recall scores are almost exactly proportional to the corresponding number of words (i.e., $51.7/55.1 \approx 15/16$). Stallings et al. (1993) also found no difference in raw scores between the CVLT and AVLT in their sample of head-injured patients. Because the CVLT and AVLT are administered identically (at least over the learning trials) the two tests only differ in their word lists. The AVLT uses common, high-frequency words whereas the CVLT avoids the most common exemplars of each category. Familiar words are easier to recall (though not to recognize) (Crowder, 1976) and thus a more parsimonious account for equivalent recall on the AVLT and CVLT is that the benefit of semantic categories is offset by a more difficult word list. An alternative explanation for the apparent equivalence between the AVLT and CVLT is that semantic clustering on the CVLT is attenuated over repeated learning trials. Mandler and Dean (1969) showed that if words are presented in a fixed order (as on the CVLT) subjects tend to recall them in the same order each time. Thus, if subjects are to organize the words by category, they must presumably overcome this natural propensity to repeat the words in the order they are presented. Scrambling the word order on each trial may minimize this effect and provide a more sensitive measure of semantic clustering over repeated trials. Tulving (1962) himself employed mixed word order in his original paradigm, a practice that is also used on the popular CERAD (Consortium to Establish a Registry for Alzheimer's Disease) assessment battery (Morris et al., 1989).

Cued Recall

In general, the rationale for comparing free and cued recall is to discriminate retrieval from encoding deficits. Crosson et al. (1988) found that categorical cues on the CVLT improved recall by head-injured patients but not by normal controls. Of course, this finding could suggest either poor retrieval among the head-injured subjects or a ceiling effect in free recall among the control subjects. Crosson and his colleagues then divided patients into groups of presumed encoding and retrieval deficits on the basis of their recognition scores (Crosson, Novack, Trenerry, & Craig, 1989). They found that semantic cues improved recall only in the encoding deficit group, just the opposite of the predicted result. Using a similar procedure with head-injured patients, Wilde, Boake, and Sherer (1984) defined encoding and retrieval deficits by the difference between recognition discriminability and delayed free recall. In this case, the encoding and retrieval groups gained equally from category cues.

Another technique used to judge if differences between cued and free recall distinguish encoding and retrieval deficits has been to compare Alzheimer's and Huntington's patients. It is widely held that although both groups have impaired recall, recognition memory is relatively preserved in Huntington's disease (Brandt, Folstein, & Folstein, 1988), at least in its early stages.[4] This pattern implies encoding of Huntington's patients is intact and thus they will benefit more from retrieval cues than patients with Alzheimer's disease. However, several studies compared free and cued recall on the CVLT (Delis, Massman, et al., 1991; Johnson et al., 1994; Massman et al., 1990) and none found significant group differences, regardless of the stage of dementia. Delis et al. (1987) propose that cued recall may help to choose cognitive rehabilitation strategies by showing whether patients can use semantic mnemonics. This is an intriguing and clinically relevant hypothesis but it has not been empirically evaluated. Clearly, the promise of cued recall scores to distinguish encoding and retrieval deficits has not been realized.

RETENTION

Delayed Recall

Delayed recall on verbal learning tasks is generally thought to be more sensitive to early Alzheimer's disease than immediate recall (Welsh et al., 1989). Studies of the CVLT have found reduced long-delay free recall in subjects with closed head injuries (Crosson et al., 1988; Haut & Shutty, 1992) and subcortical dementias (Massman et al., 1990). One problem with delayed recall on the CVLT is that the number of words recalled on long-delay depends on how many words were originally learned. In other words, long-delay recall confounds retention with acquisition. For that reason, several investigators have calculated retention *rate*, or the proportion of words retained over the 20 minute delay. Bondi et al. (1994) reported that long-delay "savings" scores[5] (but not short-delay recall itself) identified all of a subset of elderly subjects with family histories of dementia who later met criteria for Alzheimer's disease. Cullum et al. (1995) compared older normal controls with younger early Alzheimer's patients and found that the

[4]However, this common assumption is challenged by Brandt, Corwin, and Krafft (1992).
[5]"Savings," which Butters et al. (1988) popularized as a term for percent retention, should not be confused with its traditional meaning in experimental psychology as time or trials to a learning criterion.

greatest difference between the groups was in long delay savings. Kramer, Levin, et al. (1989) compared groups of severely demented Alzheimer's (AD), Parkinson's (PD) and Huntington's (HD) patients and found significant differences in "rate of forgetting" not only between the PD and HD groups from the AD group but between the PD and HD groups themselves. Unfortunately, neither of these studies is entirely clear on whether the savings or forgetting rates were referenced to the last learning trial before the distractor list or the short-delay free recall just after. Nonetheless, the studies do suggest that retention *rate* may be a more sensitive measure of long-term retention than the absolute long-delay free recall score itself.

Recognition

The CVLT uses four measures of recognition: (1) recognition hits, or how many of the 16 target words are correctly identified, (2) false positives, the number of foils that are wrongly identified, (3) response bias, an index of yea or nea tendency, and (4) discriminability, which considers the total number of responses and thus accounts for correct endorsements by chance alone. The recognition hit rate is related to general learning while false positives are associated with response discrimination (Schear & Craft, 1989b; Vanderploeg, Schinka, & Retzlaff, 1994; Wiens et al., 1994; Wilde & Boake, 1993). Studies of recognition on the CVLT have yielded modest results. Crosson et al. (1988) found lower discriminability and greater false positives among patients with severe closed head injuries. On the other hand, Millis and Ricker (1994b) reported near average discriminability scores in a small sample of patients with penetrating head wounds. Massman et al. (1990) found that both Alzheimer's and Parkinson's patients had lower discriminability and more false positives than normals but the difference between the two groups was not significantly different. Cullum et al. (1995) found that discriminability scores of early Alzheimer's subjects were significantly lower than those of healthy older subjects. In a series of studies, Hermann and his colleagues used the CVLT to assess patients with dominant temporal lobe complex partial seizures (CPS). They found that while the left side focus group recalled less than the right side CPS or control groups, recognition hits and false positives were not significantly different (Hermann et al., 1987). In a later study (Hermann et al., 1988) they devised a measure of "retrieval difficulty" by comparing recognition hits with the number of words recalled on the last trial of list A. The left side CPS group scored lower on this index than either the right side CPS group or the control group. The authors argued that retrieval difficulty discriminated left side from right side seizures. However, by combining recall (which did discriminate the two groups) and recognition (which did not), differences in "retrieval difficulty" could simply be an artifact of the differences in recall. Hermann et al. (1988) found that retrieval difficulty was related to word generation and reading comprehension, though they did not separate the correlations between recall and recognition and those language measures. Hermann et al. (1992) found that none of the recognition measures discriminated CPS patients after left or right temporal lobectomy.

One of the clinical appeals of recognition tests is their supposed ability to distinguish retrieval from encoding deficits. Although the manual makes that claim for the CVLT, it is challenged by several studies of head injured patients. Crosson et al. (1989) defined subjects as having encoding, consolidation, or retrieval deficits by their recognition hits and false positive scores. The retrieval deficit group actually recalled more words on long-delay than they did on the short-delay trials and the encoding group benefitted most from category cues. Wilde et al. (1994) formed their retrieval group with subjects who scored better on recognition discriminability than long delay free recall score. Surprisingly, both groups benefitted equally from semantic cues and the supposed retrieval deficit group actually showed more consistent recall. Wilde, Boake, and Sherer (1995) divided their subjects groups with presumed retrieval and encoding deficits by the same criteria that Crosson et al. (1989) used earlier (viz., recognition hits and false positives). Again, both groups benefitted equally from semantic cues (the differences between free- and cued-recall scores were similar).

Several studies have compared CVLT recognition scores by subjects with Alzheimer's (AD) Huntington's (HD) dementia. Kramer, Delis, Blusewicz, et al. (1988) found no significant difference in recognition hits between the two groups, although mild HD subjects had less response bias and fewer false positives than did AD subjects. Discriminability scores did discriminate Alzheimer's patients from mild, but not moderate to severe, Huntington's patients. However, later studies have rather consistently found that while recall on the CVLT is impaired in both AD and HD groups, recognition discriminability is relatively spared in Huntington's disease (Delis et al., 1991; Massman, Delis, Butters, DuPont, & Gillin, 1992; Johnson et al., 1994).

Some studies suggest that recognition tasks on the CVLT may help to indicate whether subjects are feigning or exaggerating memory problems. Trueblood (1994; Trueblood & Schmidt, 1993) identified two subgroups of patients referred for evaluation of mild head injury. One group was presumed to be malingering because their performance on another forced-choice recognition task was below chance. The other group was thought to be questionable because of inconsistent test results. Both "questionable" groups scored lower than their matched patient controls on recognition hits and total recall. Millis and Ricker (1994a) compared subjects who were known to have moderate to severe traumatic brain injuries with those suspected of malingering. The malingering group consisted of patients who were seeking compensation for alleged head injuries but who reported only brief or no loss of consciousness, and showed no focal neurologic signs or abnormal CT or MRI scans. Those suspected of malingering had fewer recognition hits and more false positives than those with unequivocal head injuries. A discriminant function based largely on recognition hits correctly classified the entire sample. A recent study suggests that electrophysiologic measures may be helpful in detecting feigned memory deficits by corroborating recognition performance. Retzlaff, Morris, and Duff (1993) administered the CVLT to normal subjects and measured evoked potentials while *visually* presenting words on the recognition list. Target words evoked greater P1 response at frontal sites and P3 at posterior sites than did the distractor words. However, the robust modality effect on recall implies that visual and auditory material are processed by separate systems (Penney, 1989). Thus evoked potentials may not discriminate target words on the CVLT recognition task because it uses an auditory presentation. Otto et al. (1994) found that recognition hits were lower among patients with major depression, although scores were not related to severity of the mood disorder.

INTRUSIONS AND PERSEVERATIONS

Intrusive errors on the CVLT are broadly defined as any word not in the appropriate list or category. Perseverations are defined as repetitions on the same trial. Intrusions are scored separately on free and cued recall trials whereas perseverations are tallied over all trials. Intrusions are generally thought to be useful in diagnosing Alzheimer's disease, although they are not unique to that disorder (Fuld, 1981; Kramer, Delis, Blusewicz, et al., 1988). Both types of response errors vary with the task and the clinical population. For example, Alzheimer's and Korsakoff's patients tend to make intrusive errors when recalling paragraphs and perseverative errors when generating words (Butters et al., 1987). Response errors do not appear to be related to either age (Kozora et al., 1992) or gender (Delis et al., 1987; Kramer, Delis, & Daniel, 1988).

Alzheimer's (AD) patients make far more intrusive errors on the CVLT than do subjects with Huntington's disease (HD) (Kramer, Delis, Blusewicz, et al., 1988; Kramer, Levin, et al., 1989), Parkinson's (PD) dementia (Kramer, Levin, et al., 1989), and depression-related pseudodementia (Hill et al., 1993; Massman et al., 1992). AD patients also make more intrusive errors on cued than free recall (Kramer, Delis, Blusewicz, et al., 1988; Massman, Delis, Butters, DuPont, & Gillin, 1992). Bondi et al. (1994) tested normal elderly subjects on the CVLT at 12 month intervals. Those with a positive family history of degenerative dementia made more intrusive errors on cued (but not free) recall. Further, subjects who were later classified as demented made more intrusive errors on cued recall. Thus cued recall appears to be more sensitive than free recall in the early detection of degenerative dementia.

The CVLT manual suggests that repetitions that occur close together (proximal perseverations) may indicate poor response inhibition while those far apart (distal perseverations) suggest poor short-term memory, although this distinction has not been empirically evaluated.

Intrusion and perseveration scores in the CVLT manual reflect their absolute frequency. Crosson et al. (1988) argue that because the wide variability in recall prevents meaningful comparisons, response errors should be measured by their *rate,* or proportion to the total number of responses on each trial. They found higher rates of intrusions (but not perseverations) by head injured patients, especially on cued recall. Kramer, Levin, et al. (1989) directly compared the two methods and found that intrusion rate was better at discriminating Alzheimer's from control subjects than the absolute number of intrusions. More recently, however, Cullum et al. (1995) did not find a significant difference in intrusion rates between early Alzheimer's and older normal control subjects.

Intrusions on the CVLT have also been found with traumatic brain injuries, although corresponding increases in perseverations have not been found (Crosson et al., 1988; Numan & Sweet, 1992). Head injured patients with left temporal lesions show more intrusions, usually from the same semantic category, during learning and delayed recall (Crosson, Sartor, et al., 1993). Cullum et al. (1989) found more intrusions (but not perseverations) among schizophrenic patients. Heinrichs and Awad (1993) used cluster analysis to identify five subgroups of schizophrenic patients. The only group with abnormally high intrusions had other test scores that were characteristic of dementia. However, Heinrichs (1994) found that intrusions among schizophrenics were within a standard deviation of normal standard scores in the CVLT manual. Hermann et al. (1992) evaluated patients who underwent anterior temporal lobectomies for intractable complex partial seizures. They found significant postoperative increases in free recall intrusions after left (but not right) surgeries. Diamond and DeLuca (1992) found that amnesics with anterior communicating artery aneurysms had more intrusions than either healthy controls or subjects with multiple sclerosis. Preliminary evidence suggests that the rate of intrusions (to the total number of responses) may be a more sensitive index than their absolute number. While perseverations on the CVLT are generally less common, they are more frequent in Huntingtons's dementia (Kramer, Levin, et al., 1989).

INTERFERENCE EFFECTS

Proactive interference (also called proactive inhibition, or PI) occurs when old memories disrupt new learning. Retroactive interference or inhibition (RI) is just the opposite: the disruption of old memories by new learning. The few studies of interference on the CVLT have yielded mixed results. The manual discusses three ways of measuring proactive interference, each involving a comparison between other recall scores: (1) The first is the difference between recall of list B and recall on the first trial of list A. The rationale is that if recall on the first (and only) trial of list B is less than that on the first trial of list A, then learning of list B must be disrupted by memories of list A. According the manual, the average recall is somewhat higher on list B than on the first trial of list A. However, other studies have found that recall is actually *better* on list B than on the first trial of list A among normal and head injured subjects (Crosson et al., 1988; Crosson & Wiens, 1994; Vanderploeg & Eichler, 1990). (2) The second measure of PI is the difference in recall between the shared and unshared categories on list B and the first trial of list A. Normal subjects in the CVLT reference group recalled fewer shared category words, implying a build-up of PI; they also recalled more nonshared words, which suggests a release from PI (Kramer & Delis, 1991). However, Numan and Sweet (1992) found that subjects with mild head injuries recalled fewer shared words from both A *and* B. (3) The last measure of proactive interference discussed in the CVLT manual is the difference between the number shared and nonshared words recalled on list B and the corresponding weighted averages over the five trials of list A. This index takes into account variations in the relative proportions of shared and nonshared words on list A. Numan and Sweet (1992) found a significant PI effect using this measure in a head injured sample. Massman et al. (1990) proposed measuring PI by words from list A that intrude on the recall of list B, although their suggestion has not been formally evaluated. Some investigators question whether the format of the CVLT is sensitive enough to measure Proactive interference under clinical conditions (Numan & Sweet, 1992). Both types of interference are measured by comparing recall scores. The scoring software calculates the recall scores but does not give base rates or other normative data for the contrast measures themselves.

According to the manual, retroactive interference can only be measured by revising the test procedures and using a third word list whose four semantic categories are not shared by either lists A or B. This method, which renders the standard scores unusable, has not attracted clinical interest nor has it been formally evaluated. However, several measures of RI have been proposed that can be scored from conventional administrations of the CVLT. Massman et al. (1990) measured RI by the number of list B words that intrude on the following free- and cued-recall trials but found no difference between Alzheimer's, Huntington's and Parkinson's patients. However, two recent studies found apparent evidence of retroactive interference by comparing recall on list A just before and immediately after list B. Zappalá and Trexler (1992) reported that subjects who had sustained a mild head injury recalled fewer words on short-delay free recall than on the last learning trial. Using the same criteria, Yehuda et al. (1995) found substantial retroactive interference on both short- and long-delay recall among combat veterans with posttraumatic stress disorder (PTSD). Kramer and Delis (1991) found that normal subjects recalled fewer shared category words on list B and also endorsed more recognition foils from shared than unshared categories (the CVLT scoring software distinguishes false positives that are semantically or acoustically similar to words on list B). These results suggest that it may be feasible to measure retroactive interference by conventional clinical administrations of the CVLT.

We recall that the recognition word list on the CVLT combines the 16 target words with 28 distractors or foils. Some distractors belong to one of the categories from list A (e.g., tools), some sound like one of the target words (drill-GRILL), while others are unrelated. High rates of false positives in each category presumably suggest that a subject is vulnerable to interference from semantic or acoustic similarity. The administration and scoring software counts recognition errors from each category but does not provide norms or advice for their interpretation. Although categorical errors are potentially useful, they have been largely ignored in studies of the CVLT and their clinical contribution has not been established.

CONCLUSIONS

The CVLT is based on the well-known Rey Auditory Verbal Learning Test (AVLT). It differs from the AVLT by using categorized word lists, adding a cued recall task, and testing recognition by a word list rather than a story. The fundamental problem with the CVLT is that it has not been properly standardized. Its small reference group is clearly atypical; as a result the norms are woefully inadequate and grossly inflated. The CVLT yields many other scores and comparisons, like proactive interference and recall versus recognition, for which it provides no normative data at all.

The Weins et al. (1994) norms are better than those in the CVLT manual but even they are inflated for clinical applications. The only reliability and standard error of measurement (SE_m) values in the manual are for total recall on trials 1-5. Despite a modest reliability that results in a sizeable SE_m, measurement error and confidence intervals are ignored altogether. Further, confidence intervals cannot be calculated because the reliability of most CVLT measures is simply unknown. The CVLT is perhaps best characterized as a clinical *procedure* that has not been developed into a sound psychometric instrument. To their credit, the authors and publisher concede that its norms are preliminary and discourage normative interpretations of CVLT scores. Any critique of the CVLT must acknowledge that few memory tests are adequately normed (Malec, Ivnik, & Smith, 1993; Naugle, Cullum, & Bigler, 1990; Zec, 1993). In his 1992 presidential address to the International Neuropsychological Society, Charles (Chuck) Matthews recalls a recurring discussion about a certain clinical test,

> ...the question is posed: "How many of you always or almost always view the [test] as a critical part of [your] evaluation?"; and all the hands go up. The typical follow-up question is "How many of you are satisfied with the reliability and validity of the procedure?", and in a sort of collective nervous flutter, all the hands go down. (Matthews, 1992, p. 137).

Adequate standardization of neuropsychological tests is not a utopian goal. The Kaufman Short Neuropsychological Assessment Procedure (K-SNAP; Kaufman & Kaufman, 1994) and the computer-administered Microcog: Assessment of Cognitive Function (Powell et al., 1993) are both based on large, stratified random samples that reflect recent U.S. census. Both tests provide separate norms by age, gender, and education, and include confidence intervals. Whatever validity these tests may possess, this level of standardization should be the "norm" for all neuropsychological tests, including the CVLT.

Delis et al. (1987) developed the CVLT to measure key concepts in cognitive science. However, the CVLT involves multiple trials while most of those concepts are based on single-trial paradigms and some of the effects found in the laboratory paradigms do not extend to the CVLT. For example, primacy on the CVLT does not show the usual agewise decline and the categorized word list has not been shown to improve recall by either normal or clinical subjects. Factor analyses of the CVLT find only two unequivocal major components, general verbal learning, which accounts for most of the variance, and response discrimination. Components that reflect learning strategy and serial position effects are less consistent and account for relatively little variance. Additional components have been identified in some studies but these are hard to justify on statistical grounds.

They appear to be unrelated to other neuropsychological measures and may simply be artifacts of analyzing redundant measures.

Empirically, the recall measures on the CVLT are consistently lower across a broad range of clinical groups. Total free recall appears sensitive to subtle impairments in the prodromal stage of Alzheimer's disease. When judged by the traditional psychometric criterion, total recall on the CVLT appears to have good discriminative validity. However, few studies have calculated the sensitivity or specificity of the CVLT and no study has established its predictive validity. I have argued elsewhere (Elwood, 1993) that statistical between-group comparisons (e.g., *t*-test, ANOVA) by themselves do not demonstrate that a test can discriminate individual subjects with enough accuracy for clinical use. Rather, the discriminative validity of a clinical test must be evaluated by its predictive powers (i.e., the rates of test positives to true positives and test negatives to true negatives). Because predictive power varies with the base rate of the target disorder(s), it should be based on the prevalence of the disorder in the population being assessed.

The various recall measures on the CVLT are all highly correlated. For example, total recall over all five trials of list A can be predicted by either trial 1 or 5 alone. This redundancy can be deceiving. Normally, agreement between independent measures implies good reliability and inspires confidence in their collective interpretation. However, because recall measures on the CVLT are highly dependent, agreement between observed scores may simply be a statistical artifact. Learning slope (acquisition rate) and recall consistency on the CVLT have been largely ignored, although lower consistency is found among Parkinson's, Huntingtons's, and schizophrenic subjects. Measures of serial position effects on the CVLT are inconsistent; some studies find that increased recency is associated with poor recall while others do not. The expected association between primacy and greater recall has not been found. No study has found that serial clustering predicts performance on other CVLT measures or that it can discriminate memory disorders.

The clinical contribution of semantic organization on the CVLT is not clear. Semantic clustering is so highly associated with recall that it may be redundant. Certainly, semantic clustering does not *explain* poor recall. Lezak (1995) finds the semantic categories a mixed blessing: on one hand, they provide valuable information about a subject's learning strategy while on the other hand, they confound verbal learning with conceptual ability. Neither semantic clustering or cued recall on the CVLT has been shown to consistently discriminate between encoding and retrieval deficits. Cued recall on the CVLT appears to be helpful in discriminating Alzheimer's disease from Huntington's dementia, not by the cued recall score itself but by the frequent intrusive errors these patients make. However, intrusions on the CVLT are more sensitive when they reflect the proportion of total responses rather than their absolute number. Likewise, proportional retention scores (saving or forgetting rates) appear to be more sensitive than the absolute number of List A words on long-delay free recall. The utility of recognition measures on the CVLT is somewhat mixed. Discriminability (but not recognition hits) may be useful in discriminating Alzheimer's from Huntington's disease, especially in their mild to moderate stages. However, studies of CVLT with head injured and demented subjects have not shown that encoding and retrieval deficits can be identified by discrepancies between recall and recognition measures.

RECOMMENDATIONS

Development of the CVLT

A comprehensive standardization of the CVLT is urgently needed. Representative norms are needed, along with reliability indices, standard errors of measurement, and confidence intervals. These data are needed not just for certain raw scores but also for the various contrast measures. Standardization of the CVLT would also provide the opportunity to revise the test according to studies of the existing CVLT. Several revisions should be considered. (1) Scrambling the word order on each trial may make the CVLT a more sensitive measure of semantic organization. (2) The rate of response errors to total responses should replace (or supplement) the absolute number of intrusions and perseverations. (3) Likewise, retention (delayed recall) should be expressed as the proportion of words that were earlier recalled on the short-delay trial. (4) An optional word-stem completion test (Heinrichs & Bury, 1991) should be added to the CVLT. This simple priming task would provide a good measure of incidental learning. (5) In view of the significant examiner effects on CVLT performance (Wiens et al., 1994), the auditory presentation of the word lists should be standardized. An audiotape administration would eliminate variability in examiners' speech rate and vocal characteristics. Better still, now that many personal computers can play CD-quality sound, the word list could be presented on a computer by modifying the CVLT software. Digitized voice feedback has already been adapted to a computer-based version of the Wisconsin Card Sort Test (Heaton, 1993). I do *not* recommend visually presenting the words on screen, as several new computer tests have adopted, because the change in modality fundamentally alters the nature of the test. (6) A nine word, three category dementia version of the CVLT should also be developed. The present word list is intimidating to patients with significant memory problems and several studies suggest that a shortened version is feasible. The word list could be a subset of the one used in the full CVLT so that the same test form and software could be used for both the 16- and nine-word versions.

Clinical Application

Until the CVLT is standardized, clinicians who use the test in their own practice should recognize that the norms in the manual and scoring system are grossly inflated and that *in most clinical populations those norms will misclassify many individuals as impaired and exaggerate the impairments that are found.* Clinicians who want to evaluate verbal learning essentially have a choice between the CVLT, the AVLT, and SRT. As a measure of learning and retention the CVLT does not appear to be inherently better than the AVLT. The Mayo's Older American norms (Ivnik et al., 1992) may make the AVLT a better choice for elderly patients, though these norms appear to be inflated as well (Vangel, Lichtenberg, & Ross, 1995).

Clinicians should interpret the CVLT in terms of confidence intervals rather than absolute scores while remembering that reliability and standard errors of measurement have been established only for total recall over trials 1-5 of list A. *All* Scores on the CVLT should be interpreted cautiously. I suggest at a minimum that scores should not judged to be abnormal unless they are 1-2 z-scores below the normal range (total recall < T-score of 15-25, other scores < –2 standard scores). Although this caution is warranted, it certainly limits the clinical utility of the CVLT. The reason for clinical tests is to disclose subtle impairments that might not apparent on a neurologic or mental status examination. The time and cost to administer the test can hardly be justified if it only corroborates the most glaring deficits. Perhaps the soundest and most useful approach is to consider *profiles* of scores. Impaired recall, combined with poor recognition *and* cued recall intrusion scores would predict Alzheimer's with more confidence than poor recall and recognition alone. However, profiles of test scores do not simply translate into respective diagnoses because the pattern of scores itself often depends on the degree of dementia. Here again, the statistical techniques used in studies of the CVLT are less suited to clinical applications. Discriminant function analyses may reveal which scores best predict a given diagnosis but they do not show clinicians how to integrate those scores in order to evaluate an individual patient. Clinicians must also remember that the recall measures on the CVLT are highly correlated. The agreement between scores does not imply they are more reliable or bestow a greater confidence on clinical judgements made from them. Clinicians should avoid the temptation (and common practice) of using semantic clustering scores to *explain* poor recall. Semantic clustering is largely redundant to recall itself and its use to account for low recall scores is circular. Clinicians should evaluate the proportion of intrusion rates to total responses as well as their absolute number. Although there are no established norms for error rates on the CVLT, several studies discussed earlier may serve as interpretive guidelines. Recognition performance should be judged by the discriminability index, not merely the absolute number of recognition hits. Finally, the discriminant validity of CVLT scores should be judged by their predictive power based on realistic base rates in the clinical population being assessed. Although these procedures will not overcome the limitations of the CVLT, they will hopefully minimize errors in its interpretation.

ACKNOWLEDGMENTS

The author gratefully acknowledges Glen Salter for his invaluable assistance, the many investigators who shared their unpublished materials, and two anonymous reviewers for their helpful comments on earlier drafts of this article.

REFERENCES

Bondi, M. W., Drake, A. I., Butters, N. M., and Griffith, T. (1993, February). Verbal memory impairments in alcoholics and polysubstance abusers with concurrent alcohol abuse. Paper presented at the meeting of the International Neuropsychological Society, Galveston, TX.

Bondi, M. W., Monsch, A., Galasko, D., Butters, N., Salmon, D. P., and Delis, D. C. (1994). Preclinical cognitive markers of dementia of the Alzheimer type. *Neuropsychology* 8: 374-384.

Brandt, J., Corwin, J., and Krafft, L. (1992). Is verbal recognition really different in Huntington's and Alzheimer's disease? *Journal of Clinical and Experimental Neuropsychology* 14: 773-784.

Brandt, J., Folstein, S. E., and Folstein, M. F. (1988). Differential cognitive impairment in Alzheimer's disease and Huntington's disease. *Annals of Neurology* 35: 585-589.

Buschke, H., and Fuld, P. A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology* 24: 1019-1025.

Butters, N., Granholm, E., Salmon, D. P., and Grant, I. (1987). Episodic and semantic memory: A comparison of amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology* 9: 479-497.

Butters, N., Grant, I., Haxby, J., Judd, L. L., Martin, A., McClelland, J., Pequegnat, W., Schacter, D., and Stover, E. (1990). Assessment of AIDS-related cognitive changes: Recommendations of the NIMH Workgroup on neuropsychological assessment approaches. *Journal of Clinical and Experimental Neuropsychology* 12: 963-978.

Butters, N., Salmon, D. P., Cullum, C. M., Cairns, P., Tröster, A. I., Jacobs, D., Moss, M., and Cermak, L. S. (1988). Differentiation of amnesic and demented patients with the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist* 2: 133-148.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research* 1: 245-276.

Comrey, A. L., and Lee, H. B. (1992). *A First Course in Factor Analysis* (2nd ed.), Erlbaum, Hillsdale, NJ.

Craik, F. I. M. (1984). Age difference in remembering. In Squire, L. R., and Butters, N. (Eds.), *Neuropsychology of Memory,* Guilford Press, New York.

Crosson, B., Cooper, P. V., Lincoln, R. K., Bauer, R. M., and Velozo, C. A. (1993). Relationship between verbal memory and language performance after blunt head injury. *The Clinical Neuropsychologist* 7: 250-267.

Crosson, B., Novack, T. A., Trenerry, M. R., and Craig, P. L. (1988). California Verbal Learning Test (CVLT) performance in severely head-injured and neurologically normal adult males. *Journal of Clinical and Experimental Neuropsychology* 10: 754-768.

Crosson, B., Novack, T. A., Trenerry, M. R., and Craig, P. L. (1989). Differentiation of verbal memory deficits in blunt head injury using the recognition trial of the California Verbal Learning Test: An exploratory study. *The Clinical Neuropsychologist* 3: 29-44.

Crosson, B., Sartor, K. J., Jenny, A. B., Nabors, N. A., and Moberg, P. J. (1993). Increased intrusions during verbal recall in traumatic and nontraumatic lesions of the temporal lobe. *Neuropsychology* 7: 193-208.

Crosson, J. R., and Wiens, A. N. (1994). Comparison of the Rey Auditory-Verbal Learning (AVLT) and California Verbal Learning Test (CVLT) in a sample of normal subjects. *Journal of Clinical and Experimental Neuropsychology* 16: 75-90.

Cullum, C. M., Filley, C. M., & Kozora, E. (1995). Episodic memory function in advanced aging and early Alzheimer's disease. *Journal of the International Neuropsychological Society* 1: 100-103.

Cullum, C. M., Kuck, J., Delis, D. C., Heaton, R., Zisook, S., and Grant, I. (1989, February). *Verbal learning characteristics of Schizophrenia.* Paper presented at the meeting of the International Neuropsychological Society, Orlando, FL.

Davis, H. P., Cohen, A., Gandy, M., Colombo, P., VanDusseldorp, G., Simolke, N., and Romano, J. (1990). Lexical priming deficits as a function of age. *Behavioral Neuroscience* 104: 288-297.

D'Elia, L. F., Boone, K. B., & Mitrushina, A. M. (1995). *Handbook of Normative Data for Neuropsychological Assessment.* New York: Oxford University Press.

Delis, D. C., Cullum, C. M., Butters, N., and Cairns P. (1988). Wechsler Memory Scale-Revised and California Verbal Learning Test: Convergence and divergence. *The Clinical Neuropsychologist* 2: 188-196.

Delis, D. C., Freeland, J., Kramer, J. H., and Kaplan, E. (1988). Integrating clinical assessment with cognitive neuroscience: Construct validation of the California Verbal Learning Test. *Journal of Consulting and Clinical Psychology* 56: 123-130.

Delis, D. C., Kramer, J. H., Kaplan, E., and Ober, B. A. (1987). *The California Verbal Learning Test: Research Edition, Adult Version,* The Psychological Corporation, San Antonio, TX.

Delis, D. C., Kramer, J. H., Kaplan, E., and Ober, B. A. (1993). *The California Verbal Learning Test for Children,* The Psychological Corporation, San Antonio, TX.

Delis, D. C., Massman, P. J., Butters, N., Salmon, D. P., Cermak, L. S., and Kramer, J. H. (1991). Profiles of demented and amnesic patients on the California Verbal Learning Test: Implications for the assessment of memory disorders. *Psychological Assessment* 3: 19-26.5

Delis, D. C., McKee, R., Massman, P. J., Kramer, J., Kaplan, E., and Gettmen, D. (1991). Alternate form of the California Verbal Learning Test: Development and reliability. *The Clinical Neuropsychologist* 5: 154-162.

Diamond, B. J., and DeLuca, J. (1993, November). Performance of anterior communicating artery amnesics and multiple sclerosis patients on the California Verbal Learning Test. Paper presented at the meeting of the National Academy of Neuropsychology, Phoenix, AZ.

Elwood, R. W. (1991). The Wechsler Memory Scale-Revised: Psychometric characteristics and clinical application. *Neuropsychology Review* 2: 179-201.

Elwood, R. W. (1993). Clinical discriminations and neuropsychological tests: An appeal to Bayes' theorem. *The Clinical Neuropsychologist* 7: 224-233.

Frankel, F., and Cole, M. (1971). Measures of category clustering in free recall. *Psychological Bulletin* 76: 39-44.

Fridlund, A. J., & Delis, D. C. (1987). *CVLT administration and scoring system* (version 1.2) [computer program], The Psychological Corporation, San Antonio, TX.

Fuld, P. A., Katzman, R., Davies, P., and Terry, R. D. (1981). Intrusions as a sign of Alzheimer dementia: Chemical and pathological verification. *Annals of Neurology* 11: 155-159.

Gansler, D. A., McGrath, N., Kaplan, E., Moczynksi, N., and Norton, J. (1995, February). The utility of the Dementia Version of the California Verbal Learning Test (CVLT-D) in a traumatic brain injury population. Paper presented at the meeting of the International Neuropsychological Society, Seattle, WA.

Glutting, J. J., McDermott, P. A., and Stanley, J. C. (1987). Resolving differences among methods of establishing confidence limits for test scores. *Educational and Psychological Measurement* 47: 607-614.

Haut, M. W., and Shutty, M. S. (1992). Patterns of verbal learning after closed head injury. *Neuropsychology* 6: 51-58.

Heaton, R. K. (1993). *WCST: Computer Version-2 Research Edition,* Psychological Assessment Resources, Odessa, FL.

Heinrichs, R. W. (1994). Performance of tests of diencephalic-hippocampal verbal memory function in schizophrenia, Korsakoff's syndrome and personality disorder. *Schizophrenia Research* 13: 127-132.

Heinrichs, R. W., & Awad, A. G. (1993). Neurocognitive subtypes of chronic schizophrenia. *Schizophrenia Research* 9: 49-48.

Heinrichs, R. W., and Bury, A. S. (1991, February). Verbal learning in schizophrenia: Evidence of a selective deficit in implicit memory. Paper Presented at the meeting of the International Neuropsychological Society, San Antonio, TX.

Hermann, B. P., Wyler, A. R., Bush, A. J., and Tabatabai, F. R. (1992). Differential effects of left and right anterior temporal lobectomy on verbal learning and memory performance. *Epilepsia* 33: 289-297.

Hermann, B. P., Wyler, A. R., Richey, E. T., and Rea, J. M. (1987). Memory function and verbal learning ability in patients with complex partial seizures of temporal lobe origin. *Epilepsia* 28: 547-554.

Hermann, B. P., Wyler, A. R., and Somes, G. (1993, February). Memory loss following left anterior temporal lobectomy is associated with hippocampal pathology and not extend of hippocampal resection. Paper presented at the meeting of the International Neuropsychological Society, Galveston, TX.

Hermann, B. P., Wyler, A. R., Steenman, H., and Richey, E. (1988). The relationship between language function and verbal memory performance in patients with complex partial seizures. *Cortex* 24: 245-253.

Hess, J. L., Colliver, J. A., Antonacci, N. J., and Sutton, M. L. (1991, February). Informed vs. uninformed delayed recall following interference on the California Verbal Learning Test. Paper presented at the meeting of the International Neuropsychological Society, San Antonio, TX.

Hill, D. D., Stoudemire, A., Morris, R., Martino-Saltzman, D., and Markwalter, H. R. (1993). Similarities and differences in memory deficits in patients with primary dementia and depression-related cognitive dysfunction. *Journal of Neuropsychiatry* 5: 277-282.

Ivnik, R. J., Malec, J. F., Tangalos, E. G., Peterson, R. C., Kokmen, E., and Swenson, M. R. (1992). Mayo's older Americans normative studies: Updated AVLT norms for ages 56 to 94. *The Clinical Neuropsychologist* 6: (Suppl.): 49-82.

Johnson, S. A., Sadek, J. R., Paulsen, J. S., Butters, N., Salmon, D. P., Delis, D. C., and Swenson, M. R. (1994, February). *Verbal memory in late stage Huntington's and Alzheimer's disease.* OH.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20: 141-151.
Kaplan, E. F., Goodglass, H., and Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.), Lea & Febiger, Philadelphia, PA.
Kaplan, R. F., Meadows, M. F., Vincent, I. C., Logigian, F. I., and Steere, A. C. (1992). Memory impairment and depression in patients with lyme encephalopathy: Comparison with fibromyalgia and nonpsychotically depressed patients. *Neurology* 42: 1263-1267.
Kaufman, A. S., and Kaufman, N. (1994). *K-SNAP: Kaufman Short Neuropsychological Test Procedure,* American Guidance Service, Circle Pines, MN.
Kessler, H. R., Cohen, R. A., Lauer, K., & Kausch, D. F. (1992). The relationship between disability and memory dysfunction in multiple sclerosis. *International Journal of Neuroscience* 62: 17-34.
Kibby, M. Y., Schmitter-Edgecome, M. E., and Long C. J. (1994, February). The California Verbal Learning Test (CVLT) as a measure of memory functioning in severely head-injured adults and neurologically normal controls. Paper presented at the meeting of the National Academy of Neuropsychology, Orlando, FL.
Köhler, S. (1994). Quantitative characterization of verbal learning deficits in patients with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology* 16: 749-753.
Kozora, E., Cullum, C. M., Smernoff, E. N., Allensworth, C. A. (1992, February). Qualitative features of story and word list recall in healthy older adults: Intrusion errors and serial position effects. Paper presented at the meeting of the International Neuropsychological Society, San Diego, CA.
Kramer, J. H., Blusewicz, M. J., and Preston, K. A. (1989). The premature aging hypothesis: Old before its time? *Journal of Consulting and Clinical Psychology* 57: 257-262.
Kramer, J. H., and Delis, D. C. (1991). Interference effects on the California Verbal Learning Test. *Psychological Assessment* 3: 299-302.
Kramer, J. H., Delis, D. C., Blusewicz, M. J., Brandt, J., Ober, B. A., and Strauss, M. (1988). Verbal errors in Alzheimer's and Huntington's dementias. *Developmental Neuropsychology* 4: 1-15.
Kramer, J. H., Delis, D. C., and Daniel, M. H. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology* 44: 907-915.
Kramer, J. H., Levin, B. E., Brandt, J., and Delis, D. C. (1989). Differentiation of Alzheimer's, Huntington's, and Parkinson's disease patients on the basis of verbal learning characteristics. *Neurology* 3: 111-120.
Lezak, M. (1995). *Neuropsychological assessment* (3rd ed.). Oxford University Press, New York.
Libon, D. J., Mattson, R. E., Glosser, G., Kaplan, E., Malamut, B. L., Sands, L., Swenson, R., and Cloud, B. S. (1993). A nine word dementia version of the California Verbal Learning Test. Manuscript submitted for publication.
McCarthy, M., Ferris, S. H., Clark, E., and Crook, T. (1981). Acquisition and retention of categorized material in normal aging and senile dementia. *Experimental Aging Research* 7: 127-135.
Malec, J. F., Ivnik, R. J., and Smith, G. (1993). Neuropsychology and normal aging: The clinician's perspective. In Parks, R. W., Zec, R. F., and Wilson, R. S. (Eds.), *Neuropsychology of Alzheimer's Disease and Other Dementias* (pp. 3-80), Oxford University Press, New York.
Massman, P. J., Delis, D. C., Butters, N., Levin, B. E., and Salmon, D. P. (1990). Are all subcortical dementias alike?: Verbal learning and memory in Parkinson's and Huntington's diseased patients. *Journal of Clinical and Experimental Neuropsychology* 12: 729-744.
Massman, P. J., Delis, D. C., Butters, N., Dupont, R. M., and Gillin, J. C. (1992). The subcortical dysfunction hypothesis of memory: Neuropsychological validation in a subgroup of patients. *Journal of Clinical and Experimental Neuropsychology* 14: 687-706.
Massman, P. J., Delis, D. C., and Butters, N. (1993). Does impaired primacy recall equal impaired long-term storage? Serial position effects in Huntington's disease and Alzheimer's disease. *Developmental Neuropsychology* 9: 1-15.

Matthews, C. G. (1992). The neuropsychology of epilepsy: An overview. *Journal of Clinical and Experimental Neuropsychology* 14: 133-143.
McCaffrey, R. J., Cousins, J. P., Westervelt, H. J., Martynowicz, M., Remick, S. C., Szebenyi, S., Wagle, W. A., Bottomley, P. A., Hardy, C. J., and Haase, R. F. (1995). Practice effects with the NIMH AIDS Abbreviated Neuropsychological Battery. *Archives of Clinical Neuropsychology* 10: 241-250.
Millis, S. R., and Ricker, J. H. (1994a, February). The California Verbal Learning Test in the detection of feigned memory impairment. Paper presented at the meeting of the International Neuropsychological Society, Cincinnati, OH.
Millis, S. R., and Ricker, J. H. (1994b, June). Verbal learning impairment in adult civilians following penetrating missile wounds: Performance patterns on the California Verbal Learning Test. Paper presented at the European Conference of the International Neuropsychological Society, Angers, France.
Mittenberg, W., and Motta, S. (1993). Effects of chronic cocaine abuse on memory and learning. *Archives of Clinical Neuropsychology* 8: 477-483.
Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D., and Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 39: 1159-1165.
Naugle, R I., Cullum, C. M., and Bigler, E. D. (1990). Evaluation of intellectual and memory function among dementia patients who were intellectually superior. *The Clinical Neuropsychologist* 4: 355-374.
Novack, T. A., Kofoed, B. A., and Crosson, B. (1995). Sequential performance on the California Verbal Learning Test following traumatic brain injury. *The Clinical Neuropsychologist* 9: 38-43.
Numan, B., and Sweet, J. J. (1992, November). Use of the California Verbal Learning Test to detect proactive interference in a traumatically brain injured sample. Paper presented at the meeting of the National Academy of Neuropsychology, Pittsburgh, PA.
Otto, M. W., Bruder, G. E., Fava, M., Delis, D. C., Quitkin, F. M., and Rosenbaum, J. F. (1994). Norms for depressed patients for the California Verbal Learning Test: Associations with depression severity and self-report of cognitive difficulties. *Archives of Clinical Neuropsychology* 9: 81-88.
Paulsen, J. S., Heaton, R. K., Sadek, J. R., Perry, W., Delis, D. C., Braff, D., Kuck, J., Zisook, S., and Jeste, D. V. (1995). The nature of learning and memory impairments in schizophrenia. *Journal of the International Neuropsychological Society* 1: 88-89.
Peavy, G., Jacobs, D., Salmon, D. P., Butters, N., Delis, D. C., Taylor, M., Massman, P., Stout, J., Heindel, W. C., Kirson, D., Atkinson, J. H., Chandler, J. L., and Grant, I. (1994). Verbal memory performance of patients with human immunodeficiency virus infection: Evidence of subcortical dysfunction. *Journal of Clinical and Experimental Neuropsychology* 16: 508-523.
Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition* 17: 398-422.
Perrine, K., Devinsky, O., Donofrio, N., and Luciano, D. J. (1994, February). Relationship of the California Verbal Learning Test to other measures of memory language and frontal lobe functions. Paper presented at the meeting of the International Neuropsychological Society, Cincinnati, OH.
Pope, D. (1986). The California Verbal Learning Test: Preliminary investigation of its clinical usefulness with older adults. *Dissertation Abstracts International* 47: 1738B.
Powell, D. H., Kaplan, E. F., Whitla, D., Weintraub, S., Catlin, R., & Funkenstein, H. S. (1993). *Microcog: Assessment of Cognitive Functioning,* The Psychological Corporation, San Antonio, TX.
The Psychological Corporation. (1994). *Tests and Products for Psychological Assessment.* Author, San Antonio, TX.
Randolph, C., Gold, J. M., Kozora, E., Cullum, C. M., Hermann, B. P., and Wyler, A. R. (1994). Estimating memory function: Disparity of Wechsler Memory Scale-Revised and

California Verbal Learning Test indices in clinical and normal samples. *The Clinical Neuropsychologist* 8: 99-108.

Reddon, J. R. (1992). *PCA*_PACK: Principal components analysis package (version 2.7) [computer program]. Author, Edmonton, Alberta, Canada.

Reite, M., Cullum, C. M., Stocker, J., and Kozora, E. (1993). Neuropsychological test performance and MEG-based brain lateralization: Sex differences. *Brain Research Bulletin* 32: 325-328.

Retzlaff, P., Morris, G. L., and Duff, K. (1993, February). Evoked-potential correlates of the California Verbal Learning Test. Paper presented at the meeting of the International Neuropsychological Society, Galveston, TX.

Rey, A. (1964). *L'examen Clinique en Psychologie*. Presses Universitaries de France, Paris.

Saykin, A. J., Janssen, R. S., Sprehn, C. C., Kaplan, E., Spira, T. J., and Weller, P. (1988). Neuropsychological dysfunction in HIV infection: Characterization of a lymphadenopathy cohort. *International Journal of Clinical Neuropsychology* 10: 81-95.

Saykin, A. J., Shtasel, D. L., Gur, R. E., Kester, D. B., Mozley, L. H., Stafiniak, P., Gur, R. C. (1994). Neuropsychological deficits in neuroleptic naive patients with first-episode schizophrenia. *Archives of General Psychiatry* 51: 124-131.

Schear, J. M., and Craft, R. B. (1989a). Examination of the concurrent validity of the California Verbal Learning Test. *The Clinical Neuropsychologist* 3: 162-168.

Schear, J. M., and Craft, R. B. (1989b, February). A replication of the factor structure of the California Verbal Learning Test. Paper presented at the meeting of the International Neuropsychological Society, Vancouver, BC, Canada.

Schuell, T J. (1969). Clustering and organization in free recall. *Psychological Bulletin* 72: 353-374.

Scott, T. M., Duff, K., Moczynski, N., Norton, J., and Gansler, D. A. (1995, February). The utility of the Dementia Version of the California Verbal Learning Test (CVLT-D) in discriminating between demented and non-demented groups. Paper presented at the meeting of the International Neuropsychological Society, Seattle, WA.

Simon, E., Leach, L., Winocur, G., and Moscovitch, M. (1994). Intact primary memory in mild to moderate Alzheimer disease: Indices from the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology* 16: 414-422.

Stallings, G. A., Boake, C., & Sherer, M. (1993, February). Correspondence between the California Verbal Learning Test and the Rey Auditory Verbal Learning Test in closed-head injury patients. Paper presented at the meeting of the International Neuropsychological Society, Galveston, TX.

Stout, J. C., Bondi, M. W., Jernigan, T. L., Archibald, S. L., Salmon, D. P., and Butters, N. (1995, February). Brain morphometric predictors of California Verbal Learning Test (CVLT) performance in Alzheimer's disease. Paper presented at the meeting of the International Neuropsychological Society, Seattle, WA.

Taylor, A. E., Saint-Cyr, J. A., and Lang, A. E. (1990). Memory and learning in Parkinson's disease: Evidence for a "frontal lobe syndrome." *Brain and Cognition* 13: 211-232.

Tindall, A. G., and Bradley, J. R. (1991, February). Relationships between California Verbal Learning Test (CVLT) and Wechsler Adult Intelligence Scale-Revised (WAIS-R) Information and Vocabulary subtests in a college sample screened for neuromedical risk factors. Paper presented at the meeting of the International Neuropsychological Society, San Antonio, TX.

Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered and other invalid WAIS-R and clinical memory data. *Journal of Clinical and Experimental Neuropsychology* 16: 597-607.

Trueblood, W., and Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology* 5: 578-590.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review* 69: 344-354.

Vakil, E., and Blachstein, H. (1993). Rey Auditory-Verbal Learning Test: Structure analysis. *Journal of Clinical Psychology* 49: 883-890.

Vanderploeg, R. D., and Eichler, S. R. (1990, August). Performance of the severely traumatic brain injured on the CVLT. Paper presented at the meeting the American Psychological Association, Boston, MA.

Vanderploeg, R. D., Schinka, J. A., and Retzlaff, P. (1994). Relationships between measures of auditory verbal learning and executive functioning. *Journal of Clinical and Experimental Neuropsychology* 16: 243-252.

Vangel, S. J., Lichtenberg, P. A., & Ross, T. P. (1995). Clinical utility of the Logical Memory subtests and the relationship of demographic factors to test performance. *Journal of Clinical Geropsychology* 1: 67-77.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised*. The Psychological Corporation, San Antonio, TX.

Welsh, K. A., Butters, N., Hughes, J. P., Mohs, R. C., and Heymand, A. (1989). Detection and staging of dementia in Alzheimer's disease. *Neurology* 49: 448-452.

Welte, P. O. (1993). Indices of verbal learning and memory deficits after right hemisphere stroke. *Archives of Physical Medicine and Rehabilitation* 74: 631-636.

Wiens, A. N., Tindall, A. G., and Crossen, J. R. (1994). California Verbal Learning Test: A normative study. *The Clinical Neuropsychologist* 8: 75-90.

Wilde, M. C., and Boake, C. (1993, October). Factorial validity of the California Verbal Learning Test in head injury. Paper presented at the meeting of the National Academy of Neuropsychology, Phoenix, AZ.

Wilde, M C., Boake, C., and Sherer, M. (1994, February). Recognition versus recall discrepancies in closed head injury. Paper presented at the meeting of the International Neuropsychological Society, Cincinnati, OH.

Wilde, M. C., Boake, C., and Sherer (1995, February). A reevaluation of memory deficit subtypes in closed head injury using the California Verbal Learning Test. Paper presented at the meeting of the International Neuropsychological Society, Seattle, WA.

Wolfe, N., Linn, R., Babikian, V. L., Knoefel, J. E., and Albert, M. L. (1990). Frontal systems impairment following multiple lacunar infarcts. *Archives of Neurology* 47: 129-132.

Yehuda, R., Keefe, R. S. E., Harvey, P. D., Levengood, R. A., Gerber, D. K., Geni, J., and Siever, L. J. (1995). Learning and memory in combat veterans with posttraumatic stress disorder. *American Journal of Psychiatry* 152: 137-139.

Zappalá, G., Measo, G., Romani, L., & Merlin, P. (1992, July). The California Verbal Learning Test (CVLT): Preliminary report on the development of the Italian version. Paper presented at the European conference of the International Neuropsychological Society, Durham, United Kingdom.

Zappalá, G., and Trexler, L. E. (1992). Quantitative and qualitative aspects of memory performance after minor head injury. *Archives of Clinical Neuropsychology* 7: 145-154.

Zec, R. F. (1993). Neuropsychological functioning in Alzheimer's disease. In Parks, R. W., Zec, R. F., and Wilson, R. S. (Eds.), *Neuropsychology of Alzheimer's Disease and Other Dementias* (pp. 3-80), Oxford University Press, New York.

Zwick, W. R., and Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin* 99: 432-443.