Recall discriminability: Utility of a new CVLT–II measure in the differential diagnosis of dementia

DEAN C. DELIS,[1,2] SPENCER R. WETTER,[1] MARK W. JACOBSON,[1,2] GUERRY PEAVY,[2]
JOANNE HAMILTON,[2] ASSAWIN GONGVATANA,[2,3] JOEL H. KRAMER,[4]
MARK W. BONDI,[1,2] JODY COREY-BLOOM,[2] AND DAVID P. SALMON[2]

[1]Veterans Affairs San Diego Healthcare System, San Diego, California
[2]University of California, San Diego School of Medicine, San Diego, California
[3]San Diego State University, San Diego, California
[4]University of California, San Francisco Medical Center, San Francisco, California

Abstract

Memory tests that are in a recall format have almost universally measured accuracy in terms of the number of target items reported by the examinee. However, this traditional scoring method can, in certain cases, result in artificially inflated memory accuracy scores. That is, just as a "yes" response bias and high false-positive rate on recognition testing can artificially inflate a patient's hit rate, so, too, a liberal response bias and high intrusion rate on recall testing can artificially inflate a patient's level of target recall. Recognition tests correct for this problem by using a discriminability measure that provides a single score of hit rate relative to false-positive rate; however, recall tests rarely provide a single score of recall accuracy that corrects for intrusion rate. In the present study, we examined the utility of a new recall discriminability measure that analyzes target recall relative to intrusion rate. Patients with Alzheimer's disease (AD) or Huntington's disease (HD) were administered the CVLT–II, which provides both the traditional measure of target recall and a new measure of recall discriminability. The results indicate that the new recall discriminability measure was superior to the traditional level of target recall measure in distinguishing the recall performance of AD and HD patients. Implications of these results for clinical practice and theories of memory disorder in dementia are discussed. (*JINS*, 2005, *11*, 708–715.)

Keywords: Memory, Assessment, Recognition discriminability, Alzheimer's disease, Huntington's disease

INTRODUCTION

Improvements in neuropsychological assessment techniques can sometimes occur with the development of new methods for scoring responses even on existing clinical instruments. For example, prior to the 1980s, clinical memory tests that included yes/no recognition memory conditions (e.g., the recognition trial of the Rey Auditory Verbal Learning Test) provided scores and norms only for the total number of target items correctly endorsed (i.e., "hits"). As is now well known, the problem with this scoring method was that many patients with severe memory disorders (e.g., Alzheimer's disease, AD, or alcoholic Korsakoff syn-

drome) typically exhibit strong "yes" response biases on recognition testing, yielding both high hit *and* false-positive rates (i.e., endorsement of non-target distractor items; Delis et al., 1987, 1991; Deweer et al., 1994; Glosser et al., 1998). By scoring only the number of hits, the old scoring method often provided misleading information by (1) failing to detect the severe recognition memory impairments of these patients; and (2) awarding them with above-average recognition scores. For this reason, the original CVLT, which was developed in the early 1980s, incorporated a measure from cognitive psychology called recognition discriminability, which is a *single* score that reflects the ability of the examinee to identify target items and reject distractor items. In numerous studies, this measure has proven useful for distinguishing between patients with different types of memory disorders (see review by Delis et al., 2000). For example, it is now well known that AD patients typically obtain severely

impaired scores on measures of recognition discriminability, whereas patients with predominantly subcortical dysfunction (e.g., early Huntington's disease; HD) often exhibit disproportionate improvement on measures of recognition discriminability relative to free recall (Butters et al., 1985, 1995; Delis et al., 1991). In addition, a series of studies by Massman and colleagues showed that the inclusion of the CVLT recognition discriminability index in a discriminant function analysis resulted in a correct classification rate of 90% between patients with AD and those with HD, and 100% between patients with AD and those with depression (Massman et al., 1993). As a result of these and other studies, it is now common practice for all clinical yes/no recognition memory tests to include some type of recognition discriminability index (Spreen & Strauss, 1998; Wechsler, 1987).

In our work with memory-impaired patients, we have identified another longstanding scoring method that is almost universally used today but that can provide misleading information for certain patients. This method is the primary score used on almost all *recall* memory tests in which only the total number of target items recalled is computed. The problem with this generally universal recall scoring method is analogous to that found on past yes/no recognition tests that provided scores only for the number of hits without also factoring in the false-positive rate. That is, current *recall* scoring methods consider only the number of target items recalled *without* also factoring in the intrusion rate (i.e., extra-list errors). The misleading information that can occur here is again often seen in patients with severe memory disorders, such as those with AD or alcoholic Korsakoff syndrome. These patients often have confabulatory tendencies and tend to generate high intrusion rates on recall trials, particularly if the test includes cued-recall trials (Fuld et al., 1982; Delis et al., 1991). In addition, patients who generate intrusions typically report items that are similar to the target items (Cermak & Stiassny, 1982; Delis et al., 1991). For example, the intrusions generated by patients on word-list memory tasks are often members of the categories represented on the target list (Cermak & Stiassny, 1982; Delis et al., 2000). Similarly, the intrusions reported by patients on design-memory tests are frequently prototypical designs (e.g., square; triangle) that, in part or whole, are also often found on the target designs (Jacobs et al., 1990). If patients who are confabulating generate a high enough number of intrusions, some of their responses will likely be correct by chance. It is not uncommon for patients with high intrusion rates to achieve relatively high standardized scores in terms of their level of target recall, even though, in reality, they may be generating these responses not from explicit memory, but from confabulatory tendencies. For example, we recently tested an elderly patient whose level of correct recall on the CVLT–II fell within the average range on all of the immediate and delayed recall trials, but his overall intrusion rate was also elevated by 5 standard deviations (see Table 1). It is likely that this patient's level of target recall was artificially inflated, at least in part, by his severely elevated intrusion rate.

Table 1. CVLT–2 standardized scores of an 85-year-old man with complaints of declining memory and diagnosis of Cognitive Disorder NOS

CVLT–II variable	Traditional Correct Recall Score	New Recall Discriminability Score
Trials 1–5 Total	T score = 67 (96th %ile rank)	Z score = −3.0 (1st %ile rank)
Short Delay Free Recall	Z score = 0.0 (50th %ile rank)	Z score = −1.5 (7th %ile rank)
Short Delay Cued Recall	Z score = +0.5 (69th %ile rank)	Z score = −1.5 (7th %ile rank)
Long Delay Free Recall	Z score = +0.5 (69th %ile rank)	Z score = −1.5 (7th %ile rank)
Long Delay Cued Recall	Z score = +0.5 (69th %ile rank)	Z score = −1.0 (16th %ile rank)

In an attempt to overcome this problem and to identify patients whose level of target recall may be artificially inflated by a high intrusion rate, Delis et al. (2000) developed a new measure for the CVLT–II called *Recall Discriminability*. Analogous to *recognition* discriminability, this new recall measure provides a *single* score that factors in both target recall *and* intrusion rate. It was thought that by having a single score that analyzes target recall relative to intrusions, a more accurate measure of overall recall accuracy might be obtained.

In past experimental studies (e.g., Shear et al., 1992), some researchers have attempted to analyze target recall relative to incorrect responses by using a ratio or percentage measurement, such as:

$$\frac{\text{\# Target Items Recalled}}{\text{\# Target Items Recalled} + \text{\# Intrusions Reported}} \times 100$$

However, this type of ratio method presents at least two measurement problems. First, it fails to award additional credit for having a high target-recall score relative to a low intrusion rate. For example, 16 target responses and zero intrusions yield the same high ratio score as 1 target response and zero intrusions (i.e., 100% accuracy score for both). Second, the ratio method disproportionately penalizes examinees for having a low intrusion rate if their overall level of target recall also is low. For instance, 1 intrusion with 2 target items recalled yields an accuracy score of only 67%, whereas 1 intrusion with 12 target items recalled yields an accuracy score of 92%.

In order to overcome these psychometric problems, Delis et al. (2000) developed a new method for scoring recall discriminability for the CVLT–II that was adapted from the most commonly used method to compute *recognition* discriminability in cognitive psychology, namely, the d' measure (Macmillan & Creelman, 1991). The recognition d'

measure is based on the hit rate (number of hits/total number of targets) and the false-positive rate (number of false positives/total number of distractors). The raw d' score is analogous to a contrast z score in that it reflects the absolute difference in standard deviation units between the examinee's hit rate (signal) and false-positive rate (noise; Macmillan & Creelman, 1991). For example, if an examinee's hit rate is 84% of the possible targets (approximately 1 *SD* above the expected mean) and his or her false-positive rate is 16% of the possible distractors (approximately 1 *SD* below the expected mean), then this examinee's raw d' score would be about $+2.0$. In this case, the examinee is endorsing hits and rejecting distractors significantly above a chance level. In the reverse scenario, where the hit rate is 16% and the false-positive rate is 84%, the raw d' score would be around -2.0. In this case, the examinee is rejecting targets and endorsing distractors significantly below a chance level. If an examinee's hit rate and false-positive rate are both at 50% accuracy, then d' is zero.

The d' measure that is used to compute recognition discriminability is derived from four values: number of hits, number of possible hits (targets), number of false-positives, and number of possible false-positives (distractors). These four values are easily derived from yes/no recognition tasks, since the examinee is administered all of the possible targets and possible false-positive items. The problem on recall tasks is that, although we know the number of possible targets, we do not know the universe of possible intrusion errors. However, Delis et al. (2000) adapted the recognition d' formula to recall testing by making the assumption that, in general, the number of possible intrusions is the same as the number of possible correct responses. This assumption was made because patients rarely generate more than 16 intrusions on a single recall trial (Delis et al., 2000). Thus, on the CVLT–II, a Recall Discriminability index is computed for a particular recall trial by using the following four values: number of target words correctly reported on that trial; number of possible target words (always 16); number of intrusions reported on that trial; and our assumed number of possible intrusions (16 for most cases). However, if an examinee happens to report more than 16 intrusions on a particular trial (which is a rare occurrence), then the number of reported intrusions is also considered the number of possible intrusions.

The advantages of this new Recall Discriminability index over a ratio or proportion score are that it (1) rewards examinees for reporting higher numbers of target items in conjunction with lower numbers of intrusions, and (2) uses a d' formula that is similar to the yes/no recognition discriminability index employed on the CVLT–II, which affords direct comparisons of recall and recognition raw scores. Despite these apparent advantages, however, there have been no published studies to date that have examined the clinical utility of this new Recall Discriminability measure.

In the present study, the CVLT–II was administered to patients with either AD or HD and their recall performances were compared using the old and new recall scoring methods. These two patient groups were selected for comparison because of the relatively distinct neurocognitive mechanisms that are thought to underlie their memory disorders. That is, patients with AD are thought to have a "cortical dementia," with severely impaired recall, high intrusion rates, and severely impaired recognition memory (Butters et al., 1995; Delis et al., 1991; Deweer et al., 1994; Glosser et al., 1998). In contrast, HD patients are thought to have a "subcortical dementia," with severely impaired recall, lower intrusion rates, and disproportionately better performance on recognition testing relative to free recall (Butters et al., 1985; Kramer et al., 1988; Snodgrass & Corwin, 1988; Delis et al., 1991). We hypothesized that the AD and HD patients would exhibit comparable levels of recall when performance was analyzed using the traditional method of scoring only the number of target words generated. However, when using the new Recall Discriminability index, we hypothesized that the HD patients would exhibit significantly better recall performance than the AD patients, because the HD patients' intrusion rates should be lower than those of the AD patients (Delis et al., 1991), thereby yielding higher recall discriminability scores for the HD patients.

METHODS

Research Participants

Thirty-three patients participated in the study: 16 patients diagnosed with probable Alzheimer's disease (AD) and 17 patients diagnosed with definite Huntington's disease (HD). The AD patients were recruited from the Alzheimer's Disease Research Center at the University of California, San Diego, School of Medicine. Two senior staff neurologists made the probable AD diagnosis according to the criteria developed by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (McKhann et al., 1984). The 17 HD patients were participants of the Huntington's Disease Clinical Research Program at the University of California, San Diego, School of Medicine. All HD patients were diagnosed with definite HD by a senior staff neurologist on the basis of unequivocal motor signs (i.e., chorea) and a positive family history for HD. In some cases, genetic confirmation of expanded CAG repeats was also available. All participants gave informed written consent allowing their data to be used for the study.

Measures and Analyses

The CVLT–II was administered to all participants by trained psychometrists using the standardized procedures (Delis et al., 2000). The CVLT–II involves the oral presentation of a 16-word list (List A) over five immediate-recall trials. An interference list (List B) is then presented for one immediate-recall trial, followed by short- and long-delay free- and cued-recall and recognition testing of List A. During the

long-delay interval (approximately 20 min), nonverbal testing is administered to the subjects. The paper-and-pencil protocols were scored using the CVLT–II scoring software (Delis & Fridlund, 2000). The measures of interest in the present study were the CVLT–II recall trials, including Trials 1–5 Total Immediate Recall, Short Delay Free Recall, Short Delay Cued Recall, Long Delay Free Recall, and Long Delay Cued Recall. For each recall trial, two types of raw scores were derived: (1) the traditional score of the number of target words reported; and (2) our new measure of Recall Discriminability, which is a single score that reflects level of target recall relative to intrusion rate. These raw scores were then transformed to z scores corrected for age and gender based on the CVLT–II national normative study of 1,087 adults matched to the demographics characteristic of the U.S. population (Delis et al., 2000).

Participants were also administered the full version of the Dementia Rating Scale (DRS; Mattis, 1988), which provides a screening test of global cognitive functioning. The DRS provides a brief assessment of several cognitive domains, including attention, memory, language, and visuospatial abilities.

RESULTS

Demographics

Table 2 summarizes demographic variables and DRS scores for the two groups. No significant differences were found between groups for education level and DRS scores (ps > .80 and .30, respectively). The AD group was significantly older that the HD group ($p < .001$), an expected finding given that AD typically affects individuals later in life than HD. For this reason, CVLT–II recall performances in the two groups were compared using the standardized scores that correct for age and gender.

Traditional Correct Recall Measure

The patients' CVLT–II recall performance was first analyzed using the traditional method of scoring only the number of target words recalled. As shown in Figure 1a, the AD and HD groups exhibited comparable levels of severe impairment across the recall trials using this traditional scoring method. Independent samples t tests revealed no significant differences between AD and HD groups on any of the recall

Table 2. Demographics and DRS scores for HD and AD groups

Variable	HD	AD	*p*-value
N	17	16	
Age [years; M (SD)]	47.1 (15.4)	73.6 (12.6)	<.001
Education [years; M (SD)]	13.7 (1.9)	14.5 (2.8)	>.30
DRS [M (SD)]	120.4 (10.5)	120.8 (10.0)	>.90

DRS = Dementia Rating Scale.

trials, including Trials 1–5 Total Immediate Recall [$t(31) = .96, p > .30$]; Short Delay Free Recall [$t(31) = .37, p > .70$]; Short Delay Cued Recall [$t(31) = .36, p > .70$]; Long Delay Free Recall [$t[31] = .35, p > .70$]; and Long Delay Cued Recall [$t(31) = .31, p > .70$].

New Recall Discriminability Measure

The most important finding in the present study concerns the analysis of Recall Discriminability in the two patient groups. As can be seen in Figure 1b, the HD patients obtained higher mean Recall Discriminability scores than the AD patients across all of the recall trials. Independent samples t tests revealed that the HD patients achieved significantly higher Recall Discriminability scores than the AD patients on Short Delay Free Recall [$t(31) = 2.07, p < .05$]; Short Delay Cued Recall [$t(31) = 2.06, p < .05$]; and Long Delay Cued Recall [$t(31) = 2.59, p < .01$]. In addition, the HD patients exhibited a trend to obtain higher Recall Discriminability scores than the AD patients on Long Delay Free Recall [$t(31) = 1.89, p < .06$]. The two patient groups did not differ significantly on the List A Trials 1–5 Total Recall Discriminability measure [$t(31) = .95, p > .30$].
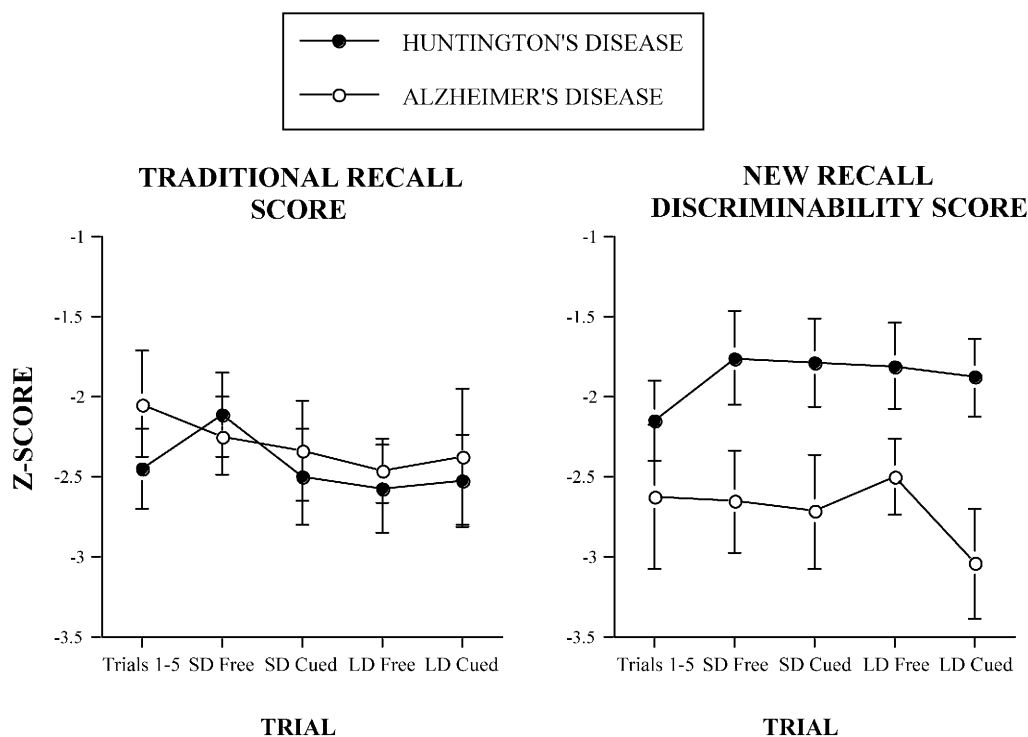
DISCUSSION

It is now widely recognized in neuropsychology that, on yes/no recognition memory testing, an analysis of only the hit rate (i.e., the number of target items endorsed) can provide inaccurate and sometimes highly misleading information. For example, AD patients often obtain normal or even above-average hit rates, but this finding is typically an artifact of their "yes" response bias and high false-positive rate (Delis et al., 1991; Deweer et al., 1994; Glosser et al., 1998). For this reason, almost all recognition memory tests published today employ some type of recognition discriminability index, which is a single score that measures the number of hits relative to the number of false-positive errors.

What is not typically realized in neuropsychology is that a similar problem can occur for *recall* memory testing as well. Almost all existing recall memory tests analyze recall performance in terms of *only* the number of target items reported. However, some patients may recall target items, not because they are accurately retrieving this information from explicit memory, but because they are generating high numbers of intrusion errors and they happen to report some target responses by chance. For example, Table 1 (see above) shows the CVLT–II results of an 85-year-old man who was given a diagnosis of Cognitive Disorder Not Otherwise Specified. This individual's level of target recall was average to above-average across all immediate and delayed recall trials (e.g., long-delay free recall was at the 69th percentile rank). However, the patient also generated 45 intrusion errors across all recall trials (z score = +5.0, *abnormal*), almost all of which were members of the categories found on the target lists. With this high rate of categorically related intrusions, it is likely that the patient generated some target items

Fig. 1. A comparison of CVLT–II recall performance in AD patients and HD patients when using the traditional recall measure (Fig. 1a) and the new Recall Discriminability measure (Fig. 1b).

because of his confabulatory tendencies rather than because of accurate recall. We observed a similar example in an AD patient who drew one to two squares on each response page of the delayed recall trial of the WMS–III Visual Reproduction subtest. This patient received partial credit for Cards C, D, and E that, when totaled, resulted in a correct-recall score that fell in the average range. It may have been the case, however, that this patient was, at least in part, confabulating a prototypical design (square) on each response page rather than remembering components of the target designs from explicit memory (see also Jacobs et al., 1990).

In order to address this shortcoming in current memory assessment practice, Delis et al. (2000) developed a new measure for the CVLT–II that was designed to provide a *single* score of recall performance that factors in both level of target recall *and* intrusion rate. Called "Recall Discriminability," this index was modeled after the standard *d′* measure that is often used to quantify recognition discriminability. Whereas recognition discriminability analyzes hit rate relative to false-positive rate, the new recall discriminability index analyzes target-recall rate relative to intrusion rate.

The present study examined the utility of this measure by comparing the recall performances of AD and HD patients on the CVLT–II using the old and new recall scoring methods. Based on past research, it was hypothesized that the two patient groups would fail to differ when their recall performance was analyzed using the traditional method of scoring only the number of target words reported. How-

ever, we hypothesized that the HD patients would exhibit significantly better recall performance than the AD patients when the new Recall Discriminability index was used, because the HD patients' intrusion rates should be lower than those of the AD patients (Delis et al., 1991). The results of the present study generally bore out these predictions. We found that the AD and HD patients failed to differ significantly on any of the immediate- and delayed-recall trials when their performances were analyzed using the traditional measure of target recall only (see also Butters et al., 1985; Delis et al., 1991; Kramer et al., 1988; Snodgrass & Corwin, 1988). In contrast, when recall performance was measured with the new Recall Discriminability index, the HD patients performed significantly better than the AD patients on the short-delay free recall trial, short-delayed cued recall trial, and long-delay cued recall trial. In addition, the HD patients showed a trend for better Recall Discriminability than the AD patients on the long-delay free recall trial ($p < .06$). Although the mean Recall Discriminability index for the five learning trials was higher for the HD than AD patients, this finding did not reach statistical significance.

The present findings have implications for the practice of memory assessment in general and for characterizations of memory deficits in dementia in particular. Just as a "yes" response bias and high false-positive rate can artificially inflate a patient's hit rate on yes/no recognition testing, so, too, a liberal response bias and high intrusion rate can artificially inflate a patient's level of target generation on recall

testing. In both cases, at least some of the target items endorsed or reported may, in fact, stem from confabulatory tendencies rather than accurate explicit memory skills. It follows then that, just as it is important to use some kind of discriminability measure on *recognition* testing, so, too, it is important to employ some type of discriminability measure on *recall* testing.

With regards to characterizations of memory disorders in dementia, past studies comparing the memory profiles of AD and HD patients have often concluded that the two patient groups exhibit comparable levels of severely impaired recall performance, but that the HD patients displayed better recognition performance (Delis et al., 1991). The present findings suggest that this characterization may not be entirely accurate. That is, given that AD patients typically generate significantly more intrusions than HD patients on memory testing, AD patients are also more likely to report target responses due to their confabulatory tendencies. As revealed in the present study, an analysis of Recall Discriminability, which factors in both target recall and intrusion rate, suggests that AD and HD patients do *not* have comparable levels of recall performance. Rather, HD patients appear to be superior to AD patients in terms of their delayed recall performances (see Figure 1). A potentially important implication of this finding is that HD patients' memory impairment may not be predominantly at the retrieval level, as previously thought. That is, if these patients are performing better on delayed free recall as reflected in their recall discriminability scores, then they may not be showing additional improvement on recognition testing relative to free recall. In a *post-hoc* analysis, we found this pattern of results to be the case for the present sample of HD patients (i.e., they were not performing significantly better on recognition discriminability relative to recall discriminability on the long delay free recall trial). These preliminary findings invite the hypothesis that HD patients have a mild to moderate encoding/storage deficit rather than a primarily retrieval impairment as traditionally thought (see Butters et al., 1995). While the purpose of the present study was not to investigate neurocognitive mechanisms underlying the memory impairment of AD and HD patients, these preliminary findings suggest that the recall discriminability index may prove helpful in increasing our understanding of the mechanisms underlying memory dysfunction.

In examining the usefulness of the Recall Discriminability measure in our clinical practice since the publication of the CVLT–II, we have found that, in patients who do not generate high intrusion rates, the standardized scores on this new measure tend to correspond to their standardized scores on the traditional measure of level of target recall. However, in patients with elevated intrusion rates, a relatively common occurrence in brain-damaged populations, the Recall Discriminability measure can be superior to the traditional target recall measure in characterizing the nature of the patients' memory impairments. In some cases, such as the patient whose CVLT–II results are shown in Table 1, the Recall Discriminability scores can be markedly discrep-

ant from the traditional target recall scores and particularly useful for documenting the memory problems that these patients may be experiencing in their lives.

As was done for the CVLT–II, recall discriminability indices could be readily developed for other memory tests as well. The question arises as to whether or not a recall discriminability measure would be useful for other memory instruments. It may be the case that this measure would have greater utility for memory tests that tend to elicit intrusion errors. For example, word-list tests that use categorized lists, particularly with category-cued recall trials (e.g., CVLT–II), may pull for more intrusion errors that uncategorized word-list tests without category-cued recall trials (e.g., Rey Auditory Verbal Learning Test, RAVLT; Delis et al., 2000; Rey, 1964); thus, a recall discriminability index may be particularly important for an instrument like the CVLT–II. However, both AD patients and frontal-lobe dementia patients have been found to exhibit significantly elevated intrusion errors on the RAVLT (Rouleau et al., 2001). In addition, individual patients with extensive confabulatory tendencies will often generate elevated intrusions responses on any type of memory test administered to them, regardless of the structure (e.g., categorized or uncategorized) or modality (e.g., verbal or nonverbal) of the target material (Barrett et al., 2000; Butters et al., 1995; Dalla Barba, 1993; DeLuca, 1993; Fischer et al., 1995; Sandson & Albert, 1987; Schnider et al., 1996). For these reason, it may be important for all recall memory tests to employ some type of recall discriminability index that can be used, at the very least, as an optional measure to interpret for those individual patients who report high intrusion rates.

A potential limitation in the present study concerns the formula that was developed to reflect recall discriminability. Just as there are different methods for computing recognition discriminability (e.g., d'; nonparametric methods; see Corwin, 1994; Delis et al., 2000), so too there are potentially different methods for computing recall discriminability. As discussed in the Introduction, some researchers have tried ratio methods in attempting to analyze correct target recall relative to intrusion errors; however, these ratio methods typically have significant psychometric shortcomings (e.g., 16 targets recalled with zero intrusions yields the same high ratio score as one target word recalled with zero intrusions; see above). We elected to adapt the d' recognition discriminability formula to recall performance, because (1) it rewards examinees for reporting higher numbers of target words in conjunction with lower numbers of intrusion errors; (2) it is based on a discriminability formula that is the mostly commonly used method for yes/no recognition tests in cognitive-science research (Macmillan & Creelman, 1991); (3) it is well suited for analysis of individual cases (Macmillan & Creelman, 1991); (4) it can be used regardless of whether the recognition test has equal or unequal target and distractor items (because d' examines hit and false-positive *rates* rather than absolute numbers); and (5) it affords a more direct method for comparing recall *versus* recognition performance (because false-positive errors are corrected for

recognition discriminability and intrusion errors are corrected for recall discriminability). However, a potential shortcoming in adapting d' to recall tests is that, while d' is readily computed on yes/no recognition tests because the number of possible false-positive errors is known, the universe of possible intrusion errors is not known on recall tests. For this reason, Delis et al. (2000) developed a conditional formula for computing intrusion rate. That is, since the number of possible correct responses on the CVLT–II is always 16 per recall trial, and since patients rarely generate more than 16 intrusion responses on any one trial, then the assumption was made that, in general, the number of possible intrusions on any one trial is the same as the number of possible correct responses (i.e., 16). However, the Delis et al. (2000) formula has some flexibility in that, if an examinee happens to report more than 16 intrusions on a particular trial (which is a rare occurrence), then the number of reported intrusions is also considered the number of possible intrusions for that trial when computing the d' formula. In those rare cases where a patient does report more than 16 intrusion errors on a single trial, the number of possible intrusions becomes larger than the number of possible correct responses. However, this occasional incident of unequal possible target and intrusion responses was one of the rationales for modeling the recall discriminability formula after d': it is well suited for equal or unequal possible targets and intrusions. Nevertheless, the recall discriminability index developed by Delis et al. (2000) is only one of a number of possible ways of computing this type of measure, and future research should strive to develop new formulas and test the relative merits of different methods for computing this type of measure. In addition, the present study represented a first attempt at examining the utility of this measure in two clinical samples; more studies are needed with larger and different patient populations to further explore the possible clinical and scientific benefits of this type of index.

In summary, the present findings suggest that the Recall Discriminability index may be useful in improving our diagnostic accuracy of memory disorders across dementia populations and in helping to elucidate the neurocognitive mechanisms that may underlie those disorders.

ACKNOWLEDGMENTS

REFERENCES

Barrett, A.M., Crucian, G.P., Schwartz, R.L., & Heilman, K.M. (2000). Testing memory for self-generated items in dementia: Method makes a difference. *Neurology*, *54*, 1258–1264.

Butters, N., Delis, D.C., & Lucas, J. (1995). Clinical assessment of memory disorders in amnesia and dementia. *Annual Review of Psychology*, *46*, 493–523.

Butters, N., Wolfe, J., Martone, M., Granholm, E., & Cermak., L.S. (1985). Memory disorders associated with Huntington's disease: Verbal recall, verbal recognition, and procedural memory. *Neuropsychologia*, *23*, 729–743.

Cermak, L.S. & Stiassny, D. (1982). Recall failure following successful generation and recognition of responses by alcoholic Korsakoff patients. *Brain and Cognition, 1*, 165–176.

Corwin, J. (1994). On measuring discrimination and response bias: Unequal numbers of targets and and distractors and two classes of distractors. *Neuropsychology*, *8*, 110–117.

Dalla Barba, G. (1993). Different patterns of confabulation. *Cortex*, *29*, 567–581.

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

Delis, D.C., Massman, P.J., Butters, N., Salmon, D.P., Cermak, L.S., & 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: A Journal of Consulting and Clinical Psychology*, *3*, 19–26.

Delis, D.C. & Fridlund, A.J. (2000). *CVLT-II Comprehensive Scoring System and Computerized Report.* San Antonio, TX: The Psychological Corporation.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (2000). *California Verbal Learning Test–II, Second Edition.* San Antonio, TX: The Psychological Corporation.

DeLuca, J. (1993). Predicting neurobehavioral patterns following anterior communicating artery aneurysm. *Cortex*, *29*, 639– 647.

Deweer, B., Ergis, A.M., Fossati, P., Pillon, B., Boller, F., Agid, Y., & Dubois, B. (1994). Explicit memory, procedural learning and lexical priming in Alzheimer's disease. *Cortex*, *30*, 113–126.

Fischer, R.S., Alexander, M.P., D'Esposito, M.D., & Otto, R. (1995). Neuropsychological and neuroanatomical correlates of confabulation. *Journal of Clinical and Experimental Neuropsychology*, *17*, 20–28.

Fuld, P.A., Katzman, R., Davies, P., & Terry, R.D. (1982). Intrusions as a sign of Alzheimer dementia: Chemical and pathological verification. *Annals of Neurology*, *11*, 155–159.

Glosser, G., Deutsch, G.K., Cole, L.C., Corwin, J., & Saykin, A.J. (1998). Differential lateralization of memory discrimination and response bias in temporal lobe epilepsy patients. *Journal of the International Neuropsychological Society*, *4*, 502–511.

Jacobs, D., Salmon, D.P., Tröster, A.I., & Butters, N. (1990). Intrusion errors in the figural memory of patients with Alzheimer's and Huntington's disease. *Archives of Clinical Neuropsychology*, *5*, 49–57.

Kramer, J.H., Delis, D.C., Blusewicz, M.J., Brandt, J., Ober, B.A., & Strauss, M. (1988). Verbal memory errors in Alzheimer's and Huntington's dementias. *Developmental Neuropsychology*, *4*, 1–5.

Macmillan, N.A. & Creelman, D.C. (1991). *Detection Theory: A User's Guide*. New York: Cambridge University Press.

Massman, P.J., Delis, D.C., Butters, N., Dupont, R.M., & Gillin, J.C. (1993). The subcortical dysfunction hypothesis of memory deficits in depression: Neuropsychological validation in a subgroup of patients. *Journal of Clinical and Experimental Neuropsychology*, *14*, 687–706.

Mattis, S. (1988). *Dementia Rating Scale: Professional Manual.* Odessa, FL: Psychological Assessment Resources Inc.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlin, E.M. (1984). Clinical diagnosis of Alzheimer's

disease: Report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

Rey, A. (1964). *L'examen clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.

Rouleau, I., Imbault, H., Laframboise, M., & Bédard, M.A. (2001). Pattern of intrusions in verbal recall: Comparison of Alzheimer's disease, Parkinson's disease, and frontal lobe dementia. *Brain and Cognition*, *46*, 244–249.

Sandson, J. & Albert, M.L. (1987). Perseveration in behavioral neurology. *Neurology*, *37*, 1736–1741.

Schnider, A., von Daniken, C., & Gutbrod, K. (1996). The mech- anisms of spontaneous and provoked confabulations. *Brain*, *119*, 1365–1375.

Shear, P.K., Tallal, P., & Delis, D.C. (1992). Verbal learning and memory in language impaired children. *Neuropsychologia*, *30*, 451–458.

Snodgrass, J.G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology*, *117*, 34–50.

Spreen, O. & Strauss, E. (1998). *A Compendium of Neuropsychological Tests*. New York: Oxford University Press.

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