Use of the California Verbal Learning Test to Detect Proactive Interference in the Traumatically Brain Injured

Bobbi Numan, Jerry J. Sweet, and Charan Ranganath
Northwestern University, Evanston Hospital

Recent studies using the California Verbal Learning Test (CVLT) to investigate the learning and memory capacities of traumatically-brain injured (TBI) individuals have suggested that this population does not show the expected buildup of proactive interference (PI). The purpose of this study was to investigate whether PI could be detected on the CVLT, in a TBI sample, if PI were calculated using alternative methods. CVLT data from 25 TBI individuals with varying degrees of brain injury and 21 healthy controls were compared. Results from the various analyses suggested that TBI individuals show buildup and release from PI when learning and attempting to recall competing forms of information if appropriate methods of analysis are used. Although the CVLT differs considerably from traditional PI paradigms (e.g., Wickens, 1970), our results suggest it can be used to detect PI in TBI individuals. © 2000 John Wiley & Sons, Inc. J Clin Psychol 56: 553–562, 2000.

Keywords: proactive interference; CVLT; traumatic brain injury

The California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987) is a clinical measure of verbal learning and memory that appears relevant to assessing the effects of numerous neurological conditions, including traumatic brain injury (TBI). The developers of this test sought to incorporate principles from cognitive science into the CVLT so that those involved in clinical assessment and clinical research could make practical use of processes and functions identified by those in the fields of cognitive psychology and cognitive neuropsychology (Delis, Freeland, Kramer, & Kaplan, 1988; Wickens, 1970). Correspondence concerning this article should be addressed to Jerry J. Sweet, Department of Psychiatry, Evanston Hospital, 2650 Ridge Ave., Evanston, IL, 60201.
Delis, Kramer, Fridlund, & Kaplan, 1990). Traditional tests of verbal memory generally yield only a single, overall measure of performance. Delis and his colleagues point out that a single score does not allow determination of the cognitive strategies used by the test-taker. Furthermore, in most cases, a single score will not allow differentiation of cognitive deficits of various patient populations because it is possible for different groups to achieve the same overall score for very different reasons.

Briefly, the CVLT consists of two word lists. The first (List A) is made up of 16 shopping items; four items each from four different categories (fruits, spices, tools, and clothing). List A is presented five times with immediate, free recall after each presentation. After the five learning trials, an interference list (List B) is presented for one immediate, free recall trial. List B consists of 16 new shopping items, eight from two of the List A categories (fruits, spices) and eight from two new categories (fish, utensils). Free recall and category-cued recall of List A is then tested. Following a 20-minute interval consisting of nonverbal testing, free recall, cued recall, and recognition memory of List A are tested.

The various components of the CVLT allow for a more complete evaluation of learning and memory. Along with an assessment of general memory functioning, the CVLT also yields measures of the rate of learning, recall consistency, and the use of various learning strategies (serial order clustering vs. semantic clustering). The recognition test provides measures of discriminability and response biases.

As mentioned above, the presence of List B allows one to measure the effects of interference on learning and memory. While the CVLT is able to provide indices of both proactive and retroactive interference, only proactive interference (PI) will be discussed here. As defined by Postman (1971), PI refers to the “detrimental effects of prior learning on the retention of subsequently learned materials” (p. 1029). With respect to the CVLT, this means that the recall of List B will be impaired because of the prior learning of List A.

The CVLT does not measure the buildup and release from PI in the same manner as some classic experimental paradigms (e.g., Postman, 1971; Wickens, 1970). For example, Wickens (1970) describes a series of experiments that are modifications of the Peterson & Peterson (1959) technique. In these experiments, series of word triads are presented for recall, with the retention interval held constant. For the first four trials, all the words belong to the same class or category. On the fifth trial a shift occurs, and the words are taken from a different category. The typical finding is that recall decreases across the first four trials (buildup of PI) and then increases with the category shift on trial five (release from PI). The amount of release depends upon the type of shift that has occurred. When the shift is one of taxonomic class, the release is quite marked (Loess, 1967; Wickens, 1970).

Despite methodological differences, a construct validation study of interference effects on the CVLT conducted by Kramer and Delis (1991) showed that the CVLT is sensitive to the effects of PI in a population of normal adults. As described earlier, half the items from List B come from the same categories as those on List A (shared category items), while the remainder are from new categories (nonshared category items). Since the buildup of PI is greater for semantically related words (Loess, 1967; Wickens, 1970), one would predict that recall of shared category items on List B would be less than that of nonshared category items. Kramer and Delis (1991) compared the number of shared and nonshared category items recalled from Trial 1 of List A to the number of shared and nonshared category items recalled from List B and found that on List B, recall of shared category items was decreased (reflecting the buildup of PI), whereas recall of nonshared category items was enhanced (reflecting release from PI). Kramer and Delis (1991) sug-
gested that the CVLT be used to study various patient populations who may be abnormally vulnerable to the effects of interference.

The CVLT has been used to explore the verbal-learning and memory capabilities of patients with Alzheimer’s Disease (Delis, Masserman, Butters, Salmon, Cermak, & Kramer, 1991); Korsakoff’s Syndrome (Delis et al., 1991); Parkinson’s Disease (Masserman, Delis, Butters, Levin, & Salmon, 1990); Huntington’s Disease (Delis et al., 1991; Masserman et al., 1990); schizophrenia (Karaken, Moberg, & Gur, 1996); bipolar disorder (Dupont et al., 1990); epilepsy (Hermann, Wyler, Steenman, & Richey, 1988); and alcoholism (Kramer, Blusewicz, & Preston, 1989). Two studies have focused on CVLT performance in the traumatically brain injured (Crosson, Novack, Trennery, & Craig, 1988; Vanderploeg & Eichler, 1990).

In the study by Vanderploeg and Eichler (1990), PI was measured by comparing recall of List B to recall of Trial 1 of List A. Under normal circumstances of learning, a buildup of PI is expected to occur as an individual is exposed to similar lists of stimuli, due to the competing nature of the information being learned. The buildup of PI is represented in a decreasing efficiency of learning performance across trials. Because Lists A and B of the CVLT are composed of similar stimuli and recall of List B was better than recall of List A in TBI patients, Vanderploeg and Eichler concluded that TBI patients do not show normal buildup of PI. It should be noted that although this method is the most commonly used by clinicians for assessing vulnerability to PI, it may not be the most sensitive measure available. However, even using a more sensitive measure, Crosson et al. (1988) also found that relative to normal controls, their TBI group did not show the expected buildup of PI.

In Crosson et al.’s (1988) study, the authors reasoned that of the items recalled from List B, 50% (by chance) should be shared category items, while the remainder should be nonshared category items. When they examined the percentages of shared category items recalled from List B, they found no differences between their TBI group and normal controls. However, when they examined each group individually, they found that the percentage of shared category items recalled from List B was significantly less than chance in the normal control group, but did not differ from chance expectations in the TBI group. These authors concluded that the TBI group failed to show a buildup of PI, even though the fact that the controls did not recall a significantly smaller percentage of shared category items than the TBI group does not appear to provide conclusive evidence that the TBI group failed to show PI.

Despite their methodological differences, the fact that two different groups of researchers (Crosson et al., 1988; Vanderploeg & Eichler, 1990) failed to show a buildup of PI in a TBI population suggests that this group merits more intensive study regarding the capacity of the CVLT to assess PI. These findings are unexpected in that another study, using a measure that differed from the CVLT, revealed normal buildup of PI in TBI individuals. Goldstein, Levin, and Boake (1989) compared TBI individuals to normal controls on the release from the PI paradigm developed by Wickens (1970). While controls exhibited higher overall recall of words, there were no differences between the groups with respect to buildup or release from PI. This raises the possibility that the CVLT simply may not be sensitive enough to detect the buildup and release from PI in TBI populations.

Another possible reason that Vanderploeg and Eichler (1990) and Crosson et al. (1988) failed to detect PI in TBI patients may lie in the analytical methods used for measuring PI. Both studies used measures of PI that differed from the method used by Kramer and Delis (1991) in their CVLT study on normal adults. Kramer and Delis compared the numbers of shared and nonshared category items recalled from List B to those
recalled from Trial 1 of List A. (They also found similar results using a weighted average of shared and nonshared category items recalled from Trial 1 of List A in order to account for the effects of the following four learning trials.) In their sample of normal adults, they found no differences in recall of shared and nonshared category items from List A, Trial 1. On List B, however, they found that recall of shared category items had significantly decreased (reflecting the buildup of PI), and that recall of nonshared category items was enhanced (reflecting release from PI). Furthermore, it appears that the methods used by Kramer and Delis can also be effective in clinical samples, since Karaken et al. (1996), using similar methods, found evidence of PI on the CVLT in a sample of schizophrenic patients.

The purpose of this study was to resolve the question of whether TBI individuals show buildup and release from PI on the CVLT. Because it appears that methodological variations in measuring PI are important, PI was measured in several different ways and these measurements were compared. Using the CVLT allows one to compare overall levels of recall between lists, and to compare recall of shared vs. nonshared category items on each list. We expected that PI could be detected in TBI individuals if the methods of Kramer and Delis (1991) were used.

Method

Participants

The TBI sample consisted of 14 males and 11 females, ranging in age from 18 to 49 years ($M = 31.12; SD = 10.76$) and having from 10 to 19 years of education ($M = 14.44; SD = 2.06$), who were referred to the Neuropsychology Service of The Evanston Hospital (Evanston, IL) for clinical assessment. Participants were tested at various intervals following their injuries, with this time period ranging from two weeks to four years, eight months ($M = 10.65$ months; $SD = 13.30$).

In order to be included in the sample, there had to have been a clear and documented history of TBI, such as a recorded period of time spent in a coma, or evidence of brain injury on computerized tomography (CT) or magnetic resonance imaging (MRI) scans (available on 12 of the patients). Severity of brain injury was assessed by noting the duration of post-traumatic amnesia (PTA) either from the participant, from family members, or from medical records. Length of PTA ranged from 0 to 42 days ($M = 7.88; SD = 11.80$). Following Jennett’s (1983) guidelines relating length of PTA and severity of injury, seven of the 25 TBI participants would be described as having a very mild injury, one a mild injury, one a moderate injury, five a severe injury, eight a very severe injury, and three an extremely severe injury. Given that documented TBI does not rule out later malingering on neuropsychological evaluation (Nies & Sweet, 1994), all participants were required to pass generally accepted motivational measures in order to be included in the present study. These measures included specific motivational measures, such as the Multi-Digit Memory Test and the Rey Memory for 15 items test. Participants also were required to demonstrate an absence of indicators of insufficient effort on actual tests of ability (e.g., CVLT, Booklet Category Test; see Sweet, 1999 for review).

The healthy control group consisted of 21 Northwestern University undergraduates (18 males, 3 females), who were enrolled in an Introduction to Psychology class and received course credit for their participation. The age range of this group was 17 to 21 years ($M = 19.00, SD = 1.14$), and years of education ranged from 12 to 15 years ($M = 13.00, SD = 1.05$). Participants in the control group denied any history of brain injury or neurological illness.
Although an age difference between the groups is apparent, the normative data found within the CVLT manual (Delis et al., 1987) indicates that the mean ages of both groups receive the same amount of correction when transformed into age-corrected standard scores. Thus, significant effects of age on CVLT performance were not expected (see further mention of this issue below).

Procedure

The CVLTs were administered individually by graduate-level students or doctoral-level psychologists trained in neuropsychological assessment following a standardized protocol. TBI participants completed the CVLT individually as part of a larger neuropsychological battery. Controls completed only a subset of measures necessary to fill the period between short- and long-delay recall. For both groups, only nonverbal measures were administered during the time delay portion of the CVLT. CVLT protocols were scored using the CVLT computer-scoring program (Fridlund & Delis, 1987), and additional measures of PI from the CVLT were scored by hand.

Results

Overall CVLT Performance

Before assessing the effects of interference on memory in this population, other data from the CVLT were explored in order to characterize the learning and memory abilities of TBI patients in general. The CVLT computer-scoring program (Fridlund & Delis, 1987) computed a raw score for each learning variable as well as a standard score comparing each individual to an age- and education-matched normative group. Table 1 presents the mean raw scores and the mean standard scores for the TBI patients and healthy participants. In order to compare learning profiles of healthy and TBI participants, we conducted $t$ tests on age-corrected standard scores for each variable. Our data indicate that across the five

<table>
<thead>
<tr>
<th>Variable</th>
<th>TBI Patients ($n = 25$)</th>
<th>Controls ($n = 21$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw Score</td>
<td>Standard Score</td>
</tr>
<tr>
<td>Total Recall A1–5</td>
<td>49.48 (11.65)</td>
<td>34.92* (15.70)</td>
</tr>
<tr>
<td>Learning Slope</td>
<td>1.32 (.63)</td>
<td>.00 (1.16)</td>
</tr>
<tr>
<td>Serial Clustering</td>
<td>3.22 (2.40)</td>
<td>.88 (1.64)</td>
</tr>
<tr>
<td>Semantic Clustering</td>
<td>1.72 (.72)</td>
<td>−.96 (.98)</td>
</tr>
<tr>
<td>Primacy Effect</td>
<td>28.60% (8.94)</td>
<td>.00 (1.68)</td>
</tr>
<tr>
<td>Middle Effect</td>
<td>40.16% (9.60)</td>
<td>−.84 (1.80)</td>
</tr>
<tr>
<td>Recency Effect</td>
<td>31.28% (10.57)</td>
<td>.72 (1.77)</td>
</tr>
<tr>
<td>Recall Consistency</td>
<td>84.72% (9.38)</td>
<td>−.24 (1.13)</td>
</tr>
</tbody>
</table>

*All scores are reported as age- and education-matched $z$-scores with the exception of Total Recall A1–5, which is reported as a $t$ score.
learning trials on List A, the TBI sample showed a nonsignificant trend to recall fewer words than controls, \( t(44) = 1.73, p = .09 \). No significant differences were observed in rate of learning, or learning slope, \( t(44) = 1.41, p = .165 \).

Serial clustering refers to the tendency to learn the list in the same order in which it was presented, and semantic clustering refers to the ability to organize recall by semantic category. Since List A is organized so that no two items from the same category are presented consecutively, any semantic clustering observed in recall indicates that the participant is imposing organization on the list, which is a more effective learning strategy. TBI patients did not differ significantly from healthy controls on serial clustering, \( t(44) = 1.12, p = .27 \), but showed a trend towards reduced semantic clustering relative to controls, \( t(44) = 1.92, p = .06 \).

The CVLT computer scoring program also generates data reflecting whether participants tended to recall words primarily from the beginning, middle, or end of the list. Since the CVLT consists of 16 items, thereby exceeding the capacity of short-term memory, recalling items from the beginning or middle of the list (i.e., primacy and middle effects in Table 1) suggests that these items were stored in and are being recalled from long-term memory. Recall of items from the end of the list (recency effect) tends to reflect recall from short-term memory (i.e., working memory). Although the two groups did not differ significantly on recall of early list items, \( t(44) < 1 \), there was a marginal trend for TBI patients to recall fewer middle items than controls, \( t(44) = 1.71, p = .09 \). TBI participants showed an enhanced recency effect relative to controls, \( t(44) = 2.19, p = .03 \). Finally, recall consistency reflects whether the participant is able to recall the same items from trial to trial. Our two groups did not differ significantly on this variable, \( t(44) < 1 \).

**Proactive Interference**

There are no age corrections available for CVLT measures of PI. Further, it should be noted that the mean ages of both groups fell within the same age-norm correction range for traditional CVLT measures, such as Total Recall A1–5. Clinically, the most commonly utilized measure of PI on the CVLT is the difference between recall of Trial 1 of List A to recall of List B. An analysis of variance (ANOVA) was conducted with Group (TBI vs. control) and List (A vs. B) as factors. This analysis revealed no significant effect for Group, \( F(1,44) < 1 \); List, \( F(1,44) = 1.18, p = .283 \); or the List \( \times \) Group interaction, \( F(1,44) = 1.56, p = .218 \). This analysis therefore revealed no evidence of proactive interference in either patients or controls. In an attempt to replicate the findings of Crosson et al. (1988), the percentages of shared category items recalled on List A, Trial 1 and List B were compared. Participants recalled significantly smaller percentages of shared category words on List B than on List A, Trial 1, \( F(1,44) = 22.48, p < .001 \). No main effect of Group was observed, \( F(1,44) < 1 \); however, a trend toward a Group \( \times \) List interaction was evident, \( F(1,44) = 3.06, p = .087 \). Inspection of the means (see Table 2) indicated that both groups showed a reduction in the percentage of shared category items recalled on List B relative to List A; however, the magnitude of the decrease was slightly larger in controls.

Another technique to assess PI was used by Kramer & Delis (1991) in which the numbers of shared and nonshared category items recalled from Trial 1 of List A and List B were compared, and their methods are replicated here. In the first analysis, an ANOVA was conducted on raw numbers of shared and nonshared category word recall on List A, Trial 1 and List B with Group as a between-subjects variable. Overall, participants recalled
more nonshared than shared category words, $F(1,44) = 15.33, p < .001$. However, this main effect was qualified by a significant List × Category interaction, $F(1,44) = 29.16, p < .001$. Follow-up planned comparisons indicate that mean recall of shared and nonshared category words did not differ at List A, Trial 1, $t(45) < 1$. During List B recall, however, participants recalled significantly fewer shared than nonshared category words, $t(45) = -5.518, p < .001$. There was no significant List × Category × Group interaction, $F(1,44) = 1.32, p = .257$, suggesting that both TBI patients and controls showed similar patterns of performance. No other effects approached significance (all $F$s < 1). Means for recall of shared and nonshared words on the two lists are presented for each group in Figure 1.

Finally, another ANOVA was computed using a weighted average of shared and nonshared category items recalled across all five trials of List A. This weighted average was computed to control for the fact that “the relative proportion of shared and nonshared

---

Table 2
Mean Number of Shared and Nonshared Category Words on List A, Trial 1, Weighted Average from List A, Trials 1–5, and List B

<table>
<thead>
<tr>
<th></th>
<th>TBI Patients ($n = 25$)</th>
<th>Controls ($n = 21$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shared</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>List A, Trial 1</td>
<td>3.28 (1.40)</td>
<td>3.28 (1.02)</td>
</tr>
<tr>
<td>List A, Weighted Average*</td>
<td>3.18 (1.14)</td>
<td>2.39 (1.21)</td>
</tr>
<tr>
<td>List B</td>
<td>2.60 (1.26)</td>
<td>3.96 (1.54)</td>
</tr>
</tbody>
</table>

*The weighted average is computed by multiplying the percentage of items recalled that are from shared or nonshared categories across the five trials from List A by the total number of items recalled on List A, Trial 1.

---

Figure 1. Mean numbers of shared and nonshared category items recalled from List A and List B of the CVLT by subject group.
category items recalled across all five learning trials affects the potential for interference during recall of List B” (Kramer & Delis, 1991, p. 300). The weighted average consists of the percentage of items recalled across the five trials of List A that are shared (or non-shared) category items, multiplied by the number of items recalled from Trial 1 of List A (Kramer & Delis, 1991). In this analysis, main effects were observed for both list, $F(1,44) = 7.86, p = .007$ and category, $F(1,44) = 13.12, p = .001$, indicating an overall tendency for participants to recall more words on List B than the weighted average for List A, and for higher recall of nonshared than shared category words. These main effects, however, were qualified by a significant List $\times$ Category interaction, $F(1,44) = 38.59, p < .001$.

The weighted averages of shared and nonshared items recalled on List A, Trial 1 did not differ significantly, $t(44) = 1$ (see Table 2 for means), indicating that this interaction reflects the effect of proactive interference during List B recall. A near-significant trend was observed for TBI participants to recall fewer words than controls, $F(1,44) = 3.12, p = .084$. Furthermore, there was a near-significant trend for a Category $\times$ Group interaction, $F(1,44) = 3.780, p = 0.058$. Inspection of group means suggests that this interaction reflects a tendency for TBI participants to recall fewer shared than nonshared category words across the five learning trials, whereas there was little difference in recall of words from the categories by controls.

**Discussion**

Our results demonstrate that the CVLT is sufficiently sensitive to detect the effect of proactive interference in heterogeneous traumatically brain-injured participants. Despite differences in age, education, and gender composition of the TBI and control groups, both groups showed qualitatively similar PI effects. Since it might be argued that demographic differences contribute to between-group variability in CVLT performance, the similarity in PI effects between the two groups could be viewed either as demonstrating a lack of support for this methodological concern or a robustness of the PI finding in normals and TBI patients. Present data do not allow such a distinction to be made; these data also do not rule out the possibility that larger TBI samples stratified by severity might show a relationship to PI.

An important finding from this study was that PI could be detected in both subject groups, provided that the proper measures were utilized. That is, as long as the dependent measure of PI took into account the relative recall of shared and nonshared category items from Lists A and B, it was sensitive to PI. Unfortunately, the most common clinical measure of proactive interference, comparison of total recall on List A, Trial 1 versus List B, was not sensitive to PI in our data as well as in a previous study of TBI patients (Vanderploeg & Eichler, 1990). This technique may not be a sensitive measure of PI for two reasons. First, by the time participants are presented with List B, they have already experienced five learning trials with List A. Thus, unimpaired recall on List B may be due to the practice of learning a 16-item list five times prior, and this may overshadow any interference effects that may be present. In addition, this technique does not differentiate shared from nonshared category items. Since the potential for PI is greater with semantically related material, this method may not be sensitive enough to detect PI that is generated among the shared category items.

Detecting PI in a TBI population is not without precedent. Goldstein et al. (1989) made use of a more traditional research paradigm and clearly demonstrated buildup and release from PI in their TBI sample. Their results, like ours, suggest that despite impairments in learning and memory, TBI individuals are sensitive to the effects of interference during verbal learning.
Nonetheless, it is still unclear that what is being called PI on the CVLT is conceptually the same as the PI that is measured with the more classical experimental paradigms. The fact that List A is presented for five learning trials before the interference list is presented makes this paradigm very different from the classical buildup and release from PI study. In the classical study, there are typically four interference trials in which there are three different items from the same category, followed by a category shift on the fifth trial in which three items from a new category are presented (Wickens, 1970). While Kramer and Delis’s (1991) weighted average is intended to help control for the five learning trials, it is nevertheless difficult to determine what the difference in recall between Lists A and B represents. It may represent PI, but it may also reflect something different. Across repeated trials of the same task, a weighted average reflects a strengthening of learning that is not possible in the single-exposure design of the traditional PI study. Further research comparing PI on the CVLT to PI obtained using more standard research designs may resolve the question of whether the nature of PI on the CVLT is different from the PI measured with more traditional methods. Perhaps by the time the CVLT is revised the instrument will have been investigated sufficiently with regard to PI and allow the authors to recommend a specific approach for clinicians to use. At present, there are no clear recommendations suggested by the research to date.

References


