



Published in final edited form as:

*Clin Neuropsychol.* 2010 January ; 24(1): 70–79. doi:10.1080/13854040903002232.

## COMPARISON OF THE TRADITIONAL RECALL-BASED VERSUS A NEW LIST-BASED METHOD FOR COMPUTING SEMANTIC CLUSTERING ON THE CALIFORNIA VERBAL LEARNING TEST: EVIDENCE FROM ALZHEIMER'S DISEASE

Dean C. Delis<sup>1,2</sup>, Eric M. Fine<sup>1,2</sup>, John L. Stricker<sup>2</sup>, Wes S. Houston<sup>3</sup>, Spencer R. Wetter<sup>4</sup>, Krystal Cobell<sup>1</sup>, Mark W. Jacobson<sup>1,2</sup>, David P. Salmon<sup>5</sup>, and Mark W. Bondi<sup>1,2</sup>

<sup>1</sup> Veterans Affairs San Diego Healthcare System, CA

<sup>2</sup> School of Medicine, Department of Psychiatry, University of California San Diego, CA

<sup>3</sup> Cincinnati Veterans Affairs Medical Center, Cincinnati, OH

<sup>4</sup> Rady Children's Hospital, San Diego, CA

<sup>5</sup> School of Medicine, Department of Neurosciences, University of California San Diego, CA, USA

### Abstract

For over 50 years, cognitive psychologists and neuropsychologists have relied almost exclusively on a method for computing semantic clustering on list-learning tasks (recall-based formula) that was derived from an outdated assumption about how learning occurs. A new procedure for computing semantic clustering (list-based formula) was developed for the CVLT-II to correct the shortcomings of the traditional method. In the present study we compared the clinical utility of the traditional recall-based method versus the new list-based method using results from the original CVLT administered to 87 patients with Alzheimer's disease and 86 matched normal control participants. Logistic regression and score distribution analyses indicated that the new list-based method enhances the detection of differences in semantic-clustering ability between the groups.

### Keywords

Learning; Memory; Neuropsychology; CVLT

### INTRODUCTION

Individuals vary considerably in the strategies they use to learn lists of words. Examples of common learning strategies include those based on (a) the categorical membership of the words (i.e., semantic clustering); (b) the position of the words on the list (i.e., primacy–recency effect); (c) the order in which the words are presented (i.e., serial clustering); and (d) idiosyncratic strategies such as recalling pairs of words consecutively based on their functional or phonemic properties (i.e., subjective clustering). A number of studies have demonstrated the clinical utility of incorporating formal measures of these different learning strategies into memory instruments (Baldo, Delis, Kramer, & Shimamura, 2002; Bayley et

al., 2000; Delis et al., 1991; Greene, Baddeley, & Hodges, 1996; Hermann et al., 1996; Ribeiro, Guerreiro, & De Mendonca, 2007; Woods, Rippeth, & Conover, 2005).

Past research in cognitive psychology and neuropsychology has found that, in group analyses, semantic clustering is the most effective strategy for learning categorizable lists of words. This strategy serves as a kind of “mental filing system” in which the individual words are organized into a smaller number of semantic units or “chunks” for more efficient encoding into and retrieval from long-term memory (Bousfield, 1953; Craik, 1981; Hunt & Love, 1972). The vast majority of cognitive and neuropsychological studies of semantic clustering have used a formula for computing this strategy that was developed over 50 years ago by the cognitive psychologist Bousfield (1953). Interestingly, this formula is based on an assumption about how learning occurs that runs counter to many modern theories of learning and memory. Nevertheless, cognitive psychologists and neuropsychologists have continued to use this formula for computing semantic clustering simply because it has been the method of choice in past studies; however, the assumption underlying this formula has rarely been discussed, questioned or even known. Consistent with the widespread use of Bousfield’s formula in cognitive psychology in the early 1980s, this formula was used to compute semantic clustering on the original version of the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987; Massman, Delis, Butters, Levin, & Salmon, 1990).

During the development of the second edition of the CVLT, our research group closely examined the original assumption underlying Bousfield’s formula for computing semantic clustering and discovered that it ran counter to current views of learning and memory (Delis, Kramer, Kaplan, & Ober, 2000; Stricker, Brown, Wixted, Baldo, & Delis, 2002). Specifically, the traditional formula assumes that examinees initiate semantic organization of a word list *after* they have retrieved as many words as they can from memory. As a result, the traditional formula’s correction for chance-expected semantic clustering is based, in large part, on the number of categories represented in the *examinee’s recalled words*, rather than on the number of categories represented in the original target list (for this reason, we call the traditional method “recalled-based” semantic clustering).

The problem with the traditional recalled-based method is illustrated in the following example. Consider the examinee who, on a particular recall trial of the original CVLT, recalls only “Grapes, tangerines, plums, and apricots.” This recall reflects good semantic clustering, since all of the words retrieved are categorically organized. However, application of the traditional recall-based method results in a semantic-clustering score that is only at chance level. Although the observed semantic-clustering score (the number of consecutively recalled word pairs from the same category) in this example is high relative to the total number of words recalled (i.e., 3), the chance-expected semantic-clustering score, as quantified by the traditional recall-based formula, is also high (i.e., 3). The reason for this high chance-expected score is that, since all four words recalled came from only one category, then, assuming that organization occurs *after* retrieval, it would be impossible not to cluster the words semantically, even by chance. That is, since “fruits” is the only category represented in the examinee’s recall, then it is impossible not to cluster all the fruit items together. Thus, the *chance*-expected cluster score for this example is the same as the *observed* score. In other words, according to the traditional recalled-based formula, it is impossible to cluster beyond a chance level when words are recalled only from one category.

In contrast to the assumption underlying the recall-based clustering formula, modern theories of learning and memory typically view semantic clustering as occurring *during* the learning process, not after target words have already been retrieved from memory (Butters &

Cermak, 1980; Squire, 1987). Semantic organization is thought to be an active, dynamic process that examinees engage in even as the target words are being presented to them. Such re-organization of words during list learning facilitates both encoding and retrieval processes of memory. In order to derive a semantic-clustering formula that more accurately reflects modern views of learning and memory, a new formula for computing semantic clustering was developed for the second edition of the CVLT (CVLT-II; Delis et al., 2000; Stricker, Brown, Wixted, Baldo, & Delis, 2002). This method, explained in detail in Stricker et al. (2002), assumes that semantic organization occurs during the learning process, not afterwards. By assuming that organization occurs during the presentation of the word list, this method uses the number of categories represented on the *target list* as a correction factor, not the number of categories represented in the examinee's recall. For this reason, we call the new method the *list-based* semantic clustering formula.

As noted above, an examinee who recalls only "Grapes, tangerines, plums, and apricots" does not receive any credit for semantic clustering above a chance level when using the traditional recall-based formula, because the formula assumes that the examinee has mental access to only one category (i.e., fruits). However, recall of these responses would receive a relatively good chance-corrected clustering score when using the new list-based formula, because this formula assumes that the examinee has mental access to all four possible categories represented on the target list during the recall process. The fact that the examinee was able to recall four items consecutively from a single category (i.e., fruits) in the face of having access to a total of four possible categories clearly indicates that the examinee is semantically clustering above a chance level (see Delis et al., 2000; Stricker et al., 2002).

Although the list-based formula intuitively appears more accurate in characterizing the role that semantic clustering plays during encoding and retrieval, the question arises as to whether this formula, as a clinical measure, is more sensitive than the traditional recall-based formula for detecting deficits in semantic clustering in neurological populations. In the present study we compared the two methods for computing semantic clustering—the traditional recall-based formula versus the new list-based formula—in a large sample of patients with Alzheimer's disease (AD) and matched normal control (NC) participants. We hypothesized that the list-based formula would be superior to the recalled-based formula in quantifying semantic-clustering deficits in the AD participants relative to the NC participants.

## METHOD

### Participants

The 173 participants in this study were selected from a larger subject pool enrolled in the Alzheimer's Disease Research Center (ADRC) at the University of California, San Diego, School of Medicine. Participants were selected without regard to gender, ethnicity, or race. Written informed consent was obtained from all participants (or their carers) after the protocol of the study had been fully explained. The diagnosis of "possible" or "probable" AD was made by two senior staff neurologists according to the criteria developed by the NINCDS-ADRDA (McKhann et al., 1984). Extensive medical, laboratory, and neuropsychological testing was performed to rule out other possible causes of dementia. The sample of AD patients ( $N=87$ ) had DRS scores ranging from 97–134. Scores from the CVLT, including the semantic clustering measure, were not used by the neurologists in making the AD diagnoses. The normal control (NC) group ( $N=86$ ) was chosen based on a neurologist's diagnosis of normal functioning and a DRS score of 135 or greater (Lucas et al., 1998; Mattis, 1988; Monsch et al., 1995). Individuals comprising the NC group had been followed at the ADRC with annual neuropsychological and neurological evaluations, and were deemed "normal" for at least 3 subsequent years. The NC participants were matched

individually with the AD patients based on age and education (within 3 years of each). No significant differences were found between the NC and AD groups in terms of age, education, or gender (see Table 1).

### Measures and procedures

As a part of the annual visit at the ADRC, the NC and AD participants were administered the original CVLT by a trained psychometrist. The CVLT 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 trial, followed by short- and long-delay recall and recognition testing of List A. The test results were scored using the CVLT computer-assisted scoring program (Fridlund & Delis, 1987). This scoring program provides the traditional recall-based measure of semantic clustering, originally developed by Bousfield (1953), in which the observed semantic-cluster score is divided by the chance-expected semantic-cluster score. The *observed* semantic-cluster score is computed by adding the number of times a correct target word is recalled immediately following another correct target word from the same category. The *chance-expected* semantic-cluster score (Bousfield, 1953; Delis et al., 1987) is computed using the following formula:

$$\text{Expected Sem Cl}_i = \sum_{n=1}^4 \frac{[Tn_i(Tn_i - 1)]}{MX_i}$$

where:  $n$ =category type (four categories per list);  $i$ =a given trial;  $Tn_i$ =number of correct words recalled from category type  $n$  on trial  $i$ ;  $MX_i$ =total number of words recalled on Trial  $i$ , including intrusions and repetitions.

For all participants, we then computed the new *list-based* measure of semantic clustering developed by Stricker et al. (2002) in which the observed semantic-cluster score is subtracted from the chance-expected score. The observed semantic-cluster score is computed in the same way for the list-based measure as for the traditional recall-based measure. The following formula is used to compute the list-based chance-expected score (see Stricker et al., 2002, for further details about the formula):

$$\text{Expected Sem Cl}_i = \frac{[(r - 1)(m - 1)]}{N_{L_i} - 1}$$

where:  $r$ =number of correct words on trial  $i$ ;  $i$ =a given trial;  $m$ =number of members of each semantic category on the original list;  $N_{L_i}$ =total number of words on the original list.

For both the recall-based and list-based measures, the total chance-corrected semantic clustering score was computed across Trials 1–5.

## RESULTS

### Comparison of recall-based and list-based clustering measures

The results of the study indicate that the NC group exhibited significantly higher semantic-clustering scores than the AD group when computing this strategy using either the traditional recall-based measure,  $t(171)=10.0, p<.001$ ; Mean NC=2.31,  $SD=0.89$  vs Mean AD=1.1,  $SD=0.74$ , or the new list-based measure,  $t(171)=10.7, p<.001$ ; Mean NC=2.23,  $SD=1.97$  vs Mean AD=-0.07,  $SD=0.38$ . Logistic regression analyses were performed to evaluate the utility of the list-based and recall-based formulas in classifying participants as

either AD or normal control. The results from these analyses indicate that while both semantic clustering formulas significantly predicted group membership—List-based measure:  $\chi^2=114.506$  (1, 172),  $p<.0001$ ; Recall-based measure:  $\chi^2=76.326$  (1, 172),  $p<.0001$ —the new list-based formula accounted for more of the variance in diagnostic status (approximately 64%) than the traditional recall-based formula (approximately 47%). In addition, the overall classification rate improved with the list-based formula, with 89.7% of the AD patients successfully predicted with the list-based formula and only 80.5% of the AD patients successfully predicted with the recall-based formula (see Tables 2 and 3 for classification rates).

Further analyses were undertaken to better characterize differences in the score distributions for the traditional recall-based and the new list-based formulas. Specifically, Levene's heterogeneity of variance test was utilized to determine whether the variance of the score distributions for each group were more divergent for the new list-based semantic-clustering formula relative to the traditional recall-based formula. The results of Levene's test indicate that the variance of the score distributions significantly differed between groups for both semantic-clustering formula, but to a greater extent for the new list-based formula ( $F=158.76$ ,  $p<.0001$ ) relative to the traditional recall-based formula ( $F=6.41$ ,  $p<.012$ ). These differences in the distributions of scores for the two semantic-clustering formulas are clearly illustrated in the scatterplots shown in Figures 1a and 1b in which each semantic-clustering formula is represented on the y-axis and Mattis DRS score is represented on the x-axis. First, examination of the score distributions of the normal controls for the two semantic-clustering formulas clearly illustrates that the scores for the new list-based method produced a larger range of scores than that of the traditional recall-based formula, which is desirable in a normative sample. Second, as can be seen in Figures 1a and 1b, the score distribution for the AD patients was more restricted for the list-based than the recall-based formula, which would be predicted in this clinical population.

These scatterplots also reveal a limitation with the recall-based method: a number of individual AD patients displayed comparable levels of semantic-clustering performance compared to NC participants when using the traditional recall-based formula (see the overlap in scatterplots of scores in Figure 1a). In contrast, the scatterplots illustrate that, when applying the new list-based method, the majority of the AD patients clustered around a chance level of semantic-clustering performance and most of the NC participants achieved higher scores when using this method (see Figure 1b). Box plot representations of these data also reveal the greater separation of the participant groups when using the new list-based method relative to the traditional recall-based formula (see Figure 2). Taken together, Figures 1 and 2 clearly reveal a greater separation in semantic-clustering performance between normal controls and the AD patients for the list-based relative to the recall-based methods. Results from the logistic regression and score distribution analyses indicate that the list-based measure is superior to the recall-based formula for detecting deficits in semantic clustering in AD patients.

Finally, we calculated the sensitivity, specificity, and the predictive values associated with different cut-off scores for the list-based semantic clustering measure (see Table 4). For example, based on the output from the logistic regression model, a list-based clustering score of  $-0.22$  has a probability of .85, indicating that AD is likely present, and the positive predictive value for this score is 92.1%. In contrast, a list-based clustering score of  $1.14$  has a probability of .15, indicating that AD is likely not present, and the negative predictive value for this score is 98.1%. Given that predictive values are prevalence dependent and that the base rate of individuals with AD in our sample (i.e., 50%) is higher than in the general population, the predicted probability values should be interpreted with caution.

Nevertheless, the cut-off scores for the list-based measure may provide clinicians with additional information for diagnosing AD.

## DISCUSSION

The measurement of semantic clustering is an important part of the neuropsychological evaluation of list-learning abilities, because it reflects the degree to which individuals are able to use higher-level organizational strategies when attempting to learn verbal material. In this study we compared two methods for computing chance-corrected semantic clustering—the traditional recall-based formula (Bousfield, 1953) and a new list-based formula (Delis et al., 2000; Stricker et al., 2002)—in terms of their sensitivity to clustering deficits in AD patients relative to NC participants. It was hypothesized that the list-based method would be superior to the recall-based method because the list-based formula is more consistent with modern theories regarding when semantic organization occurs during the learning process. The results of the study revealed that, relative to the NC group, the AD patients were significantly impaired in semantic clustering when either the list-based or recall-based formulas were used. However, consistent with our hypothesis, the new list-based method was superior to the traditional recall-based method in terms of yielding (a) better overall prediction of diagnostic status; (b) greater percentage of variance explained; and (c) greater differentiation in the distribution of scores between the groups for the list-based method due to an increase in score variance in the normal control group along with a reduction in variance for the AD patients. Taken together, these findings indicate that the list-based method enhances the dissociation in semantic-clustering ability between the NC and AD groups. Additional studies are needed to determine the utility of the list-based method for (a) differentiating individuals with distinct dementia etiologies; and (b) identifying the earliest cognitive changes associated with AD.

In summary, the current findings indicate that, in both research and clinical practice, the assessment of chance-corrected semantic clustering on word-list memory tasks is better served by employing measures that are derived from list-based rather than recall-based formulas (see also Delis et al., 2000; Stricker et al., 2002). The results also suggest that deficient organization of target information may play an important role in the memory dysfunction of AD patients.

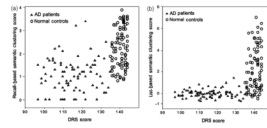
## Acknowledgments

We wish to thank the Alzheimer's Disease Research Center (ADRC) at the University of California, San Diego, School of Medicine for providing the CVLT data used in this study. Dr. Delis is a co-author of the CVLT and receives royalties from this test.

## References

- Baldo JV, Delis D, Kramer J, Shimamura AP. Memory performance on California Verbal Learning Test-II: Findings from patients with focal frontal lesions. *Journal of the International Neuropsychological Society*. 2002; 8:539–546. [PubMed: 12030307]
- Bayley PJ, Salmon DP, Bondi MW, Bui BK, Olichney J, Delis DC, et al. Comparison of the serial position effect in very mild Alzheimer's disease, mild Alzheimer's disease, and amnesia with electroconvulsive therapy. *Journal of the International Neuropsychological Society*. 2000; 6:290–298. [PubMed: 10824501]
- Bousfield WA. The occurrence of clustering in recall of randomly arranged associates. *The Journal of General Psychology*. 1953; 49:229–240.
- Butters, N.; Cermak, LS. *Alcoholic Korsakoff's syndrome: An information-processing approach to amnesia*. New York: Academic Press; 1980.

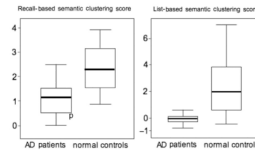
- Craik, FIM. Encoding and retrieval effects in human memory: A partial review. In: Long, JB., editor. Attention and performance. Vol. 9. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1981. p. 383-402.
- Delis, DC.; Kramer, JH.; Kaplan, E.; Ober, BA. California Verbal Learning Test. San Antonio, TX: The Psychological Corporation; 1987.
- Delis, DC.; Kramer, J.; Kaplan, E.; Ober, BA. California Verbal Learning Test. 2. San Antonio, TX: The Psychological Corporation; 2000.
- Fridlund, AJ.; Delis, DC. CVLT Research edition administration and scoring software [computer software]. New York: The Psychological Corporation; 1987.
- Greene JDW, Baddeley AD, Hodges JR. Analysis of the episodic memory deficit in early Alzheimer's disease: Evidence from the doors and people test. *Neuropsychologia*. 1996; 34:537–551. [PubMed: 8736567]
- Hermann BP, Seidenberg M, Wyler A, Davies K, Christeson J, Moran M, et al. The effects of human hippocampal resection on the serial position curve. *Cortex*. 1996; 32:323–334. [PubMed: 8800618]
- Hunt, EB.; Love, T. How good can memory be?. In: Melton, A.; Martin, E., editors. Coding process in human memory. Washington: V. H. Winston & Sons; 1972.
- Lucas JA, Ivnik RJ, Smith GE, Bohac DL, Tangalos EG, Kokmen E, et al. Normative data for the Mattis Dementia Rating Scale. *Journal of Clinical and Experimental Neuropsychology*. 1998; 20:536–547. [PubMed: 9892057]
- Massman PJ, Delis DC, Butters N, Levin BE, Salmon DP. Are all subcortical dementias alike? Verbal learning and memory in Parkinson's and Huntington's disease patients. *Journal of Experimental and Clinical Neuropsychology*. 1990; 12(5):729–744.
- Mattis, S. Dementia Rating Scale: Professional manual. Odessa, FL: Psychological Assessment Resources, Inc; 1988.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984; 34:939–944. [PubMed: 6610841]
- Monsch AU, Bondi MW, Salmon DP, Butters N, Thal LJ, Hansen LA, et al. Clinical validity of the Mattis Dementia Rating Scale in detecting dementia of the Alzheimer type. A double cross-validation and application to a community-dwelling sample. *Archives of Neurology*. 1995; 52(9)
- Ribeiro F, Guerreiro M, De Mendonca A. Verbal learning and memory deficits in mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*. 2007; 29(2):187–197. [PubMed: 17365254]
- Squire, LR. Memory and brain. Oxford, UK: Oxford University Press, Oxford; 1987.
- Stricker JL, Brown GG, Wixted J, Baldo JV, Delis DC. New semantic and serial clustering indices for the California Verbal Learning Test – second edition: Background, rationale, and formulae. *Journal of the International Neuropsychological Society*. 2002; 8:425–435. [PubMed: 11939700]
- Woods SP, Rippeth JD, Conover E. Deficient strategic control of verbal encoding and retrieval in individuals with methamphetamine dependence. *Neuropsychology*. 2005; 19:35–43. [PubMed: 15656761]



**Figure 1.**

(a) Scatterplots of the traditional recall-based semantic clustering scores relative to DRS scores in normal control (NC) participants and Alzheimer's disease (AD) participants. Note the significant overlap in semantic-clustering scores of the NC and AD participants. (b) Scatterplots of the new list-based semantic clustering scores relative to DRS scores in normal control (NC) participants and Alzheimer's disease (AD) participants. Note the greater separation of scores between the two participant groups.





**Figure 2.** Box plot representations of the test results reveal the greater separation of the NC and AD groups when using the list-based method relative to the recall-based formula.

**Table 1**

Demographic and CVLT-II results for the AD patients and normal controls

	NC	AD
<i>N</i>	86	87
Age	70.8 (7.6)	71.03 (7.6)
Age range	52–88	50–87
Education	15.5 (2.8)	14.8 (2.7)
Education range	8–20	6–20
Gender % female	55.8	49.4
DRS	140.5 (2.5)	116.0 (10.3)*
DRS range	135–144	97–136
Total words recalled for trials 1–5 from the CVLT	51 (9.5)	20 (6.7)**

DRS=Dementia Rating Scale, standard deviation in parentheses.

\*  $p < .001$ ;

\*\*  $p < .0001$ .

**Table 2**

Classification table for the new list-based method

Observed diagnosis	Controls	AD	Predicted diagnosis % correct
Controls	67	19	77.9
AD	9	78	89.7
Overall percentage correct		83.8	

**Table 3**

Classification table for the traditional recall-based method

Observed diagnosis	Controls	AD	Predicted diagnosis % correct
Controls	61	25	70.9
AD	17	70	80.5
Overall percentage correct		75.7	

**Table 4**

Sensitivity, specificity, positive and negative likelihood ratios, positive (PPV) and negative predictive values (NPV) for various list-based clustering cut-off scores

List-based cut-off score	Sensitivity %	Specificity %	Overall classification %	Positive likelihood ratio	Negative likelihood ratio	PPV %	NPV %
-0.22	40.2	96.5	68.2	11.5	.61	92.1	61.5
≥0.03	66.6	87.2	76.9	5.2	.38	84.0	72.1
≥0.14	78.2	84.9	81.5	5.2	.26	83.9	79.3
≥0.31	85.1	81.4	83.2	4.6	.18	82.2	84.3
≥0.47	89.7	77.9	83.8	4.1	.13	80.0	88.2
≥0.63	96.5	72.1	84.4	3.4	.04	77.7	95.3
≥1.14	98.8	61.6	80.3	2.6	.002	72.2	98.1