Test-Retest Stability of the Continuous Visual Memory Test in Elderly Persons

Anthony M. Paolo and Alexander I. Tröster

University of Kansas Medical Center

Joseph J. Ryan

Dwight D. Eisenhower Department of Veterans Affairs Medical Center

Ninety-two normal elderly persons were administered the Continuous Visual Memory Test (CVMT) on two occasions an average of 1.1 years apart. Means for age and education were 69.02 years (SD = 5.79) and 14.80 years (SD = 2.64), respectively. Stability coefficients were .49 on the Total score, .44 on d-prime, and .48 on Delayed Recognition. D-Prime and Delayed Recognition scores demonstrated significant change on retest. To assist clinicians in the process of detecting meaningful change on retest, standard error of difference values and 90% cut-off scores were provided. © 1998 National Academy of Neuropsychology. Published by Elsevier Science Ltd

The Continuous Visual Memory Test (CVMT; Trahan & Larrabee, 1988) measures visual learning and memory. It was designed to eliminate the motor component associated with drawing tasks and reduce the verbal labeling that may occur on tests that utilize simplistic geometric designs. In addition, the CVMT includes a 30-minute recognition trial and a visual discrimination task that helps distinguish visual memory problems from visual discrimination deficits. The construct validity of the test has been demonstrated in normal (Larrabee, Trahan, & Curtiss, 1992) and clinical (Larrabee & Curtiss, 1995) samples. The diagnostic utility of the test has been demonstrated in persons with unilateral cerebral vascular disease, closed head injury, amnestic syndrome, and dementia of the Alzheimer’s type (Trahan & Larrabee, 1988; Trahan, Larrabee, & Quintana, 1990). Limited information concerning the reliability of the CVMT is available. The test manual (Trahan & Larrabee, 1988) reports split-half reliability coefficients of at least .80 for normal and head-injured persons. Test-retest stability coefficients reported in the manual are .85 for the Total score, .80 for d-Prime, and .76 for Delayed Recognition. Although the magnitude of these coefficients is generally adequate, they were calculated from only 12 neurologically normal persons that were tested 7 days apart. Additional stability coefficients for a sample of 40 normal persons (mean age = 30.53 years) tested 1 week apart were .66 for the Total score, .53 for d-Prime, and .57 for the Delayed recognition (Trahan, Larrabee, Fritzche, &
Curtiss, 1996). Trahan et al. (1996) also reported 1-week stability coefficients of .71 for the Total score, .72 for d-Prime, and .45 for Delayed Recognition in a diagnostically heterogeneous clinical sample with an average age of 42.8 years. Little information concerning the test-retest stability of the CVMT in elderly persons has been reported.

The purpose of this study was to provide test-retest stability coefficients for the CVMT in a group of elderly persons. A second goal of the investigation was to present cut-off scores to assist in the detection of meaningful change in CVMT scores from test to retest.

METHOD

Participants

Participants were 30 male and 62 female volunteers recruited from the community and retirement centers. Means for age, education, and Dementia Rating Scale total score (DRS; Mattis, 1988) at initial assessment were 69.02 years ($SD = 5.79$; range 55–80), 14.80 years ($SD = 2.64$; range 8–20), and 138.49 ($SD = 3.40$), respectively. The majority of participants were White (97%); minority representation was 1% Black and 2% Hispanic. The primary language of all participants was English. The average Beck Depression Inventory (BDI; Beck & Steer, 1993) score at initial assessment was 4.68 ($SD = 3.14$; range 0–11), indicating the absence of significant self-reported mood disturbance in the group. The test-retest interval ranged from 9 to 21 months with an average of 13.5 months ($SD = 2.09$ months).

Procedure

Participants were normal volunteers who served as controls in a longitudinal study of neurodegenerative disease. As part of this investigation, each participant completed the CVMT on two separate occasions. Persons were recruited through print and broadcast media and received $25.00 at each assessment probe. All were interviewed and asked to complete a health questionnaire. Each participant had received a comprehensive neurological examination prior to inclusion in the study. Exclusion criteria included a history of neurologic disorder, psychiatric illness (including significant depression), and illicit drug or alcohol abuse. All participants had adequate vision and hearing. Individuals with minor age-related conditions (e.g., senile diabetes, essential hypertension, mild neurosensory hearing loss, etc.) were not excluded. To ensure that each person displayed normal cognitive functions at both assessment probes, each was required to have an initial DRS total score $> 130$ (Shay et al., 1991) and no evidence of significant decline (i.e., a drop of $\geq 10$ points; Smith et al., 1994) in total DRS score on retest. Additionally, participants had no evidence of dementia according to criteria of the Diagnostic and Statistical Manual of Mental Disorders, third edition, revised (DSM-III-R; American Psychiatric Association, 1987). All participants had normal CVMT Visual Discrimination scores of 6 or more on both assessments.

Participants completed the CVMT according to standard instructions (Trahan & Larrabee, 1988) as part of a larger battery of neuropsychological tests administered by two trained technicians. The CVMT provides six scores, but only the Total score, d-Prime, and Delayed Recognition are recommended for clinical interpretation (Trahan & Larrabee, 1988) and, as such, these scores are the focus of this study. Separate linear regression analyses were conducted to assess the impact of demographic characteristics and retest interval on changes in CVMT performance from test to retest. The Pearson product-moment coefficient was employed for the calculation of test-retest stability coefficients and paired $t$-tests were used for test-retest mean comparisons.
TABLE 1
Means, Standard Deviations, and Stability Coefficients for Continuous Visual Memory Test (CVMT)

<table>
<thead>
<tr>
<th>CVMT Score</th>
<th>Time 1</th>
<th>Time 2</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Total score</td>
<td>73.17</td>
<td>6.62</td>
<td>74.12</td>
</tr>
<tr>
<td>d-Prime</td>
<td>1.77</td>
<td>0.46</td>
<td>2.06*</td>
</tr>
<tr>
<td>Delayed recognition</td>
<td>3.63</td>
<td>1.43</td>
<td>4.11*</td>
</tr>
</tbody>
</table>

Note. M = mean; SD = standard deviation.
*Paired t-test significant at p < .05.

RESULTS

The impact of age, gender, education, and test-retest interval on changes in test scores across the two evaluation periods was evaluated via separate regression analyses. Raw score differences (i.e., the difference between the first and second testing raw scores for each of the three CVMT scores) served as the dependent variables and the demographic and interval characteristics as independent variables. Age, gender, education, and retest interval did not significantly contribute (all ps > .05) to changes in CVMT scores from test to retest.

Table 1 presents means, standard deviations, and stability coefficients for the CVMT scores for both assessment points. Test-retest coefficients were .44 on d-Prime, .48 on Delayed Recognition, and .49 on the Total score. These stability coefficients are well below the .76 to .85 reported in the test manual. Mean comparisons revealed significant retest gains on d-Prime, t(91) = 4.86, p < .05, and Delayed Recognition, t(91) = 3.24, p < .05. Significant differences did not emerge for the Total score. The average gain for d-Prime represents a 16% gain and the average improvement on Delayed Recognition was less than one design (.48).

To assist clinicians in the detection of meaningful change on retest, the standard error of the difference ($SE_d$) was calculated (Jacobson & Truax, 1991; Chelune, Naugle, Lüders, Sedlak, & Awad, 1993) for each score. The $SE_d$ provides information concerning the range of scores expected on retest due to measurement error. That is, it provides a means of detecting a statistically significant test-retest difference score so large that it is unlikely to have occurred solely by chance. The $SE_d$ values and 90% cut-off scores (i.e., 5% for a loss and 5% for a gain) are provided in Table 2.

DISCUSSION

The lack of consistent and meaningful effects of selected demographic characteristics and retest interval on CVMT difference scores suggests that clinicians may ignore this information when interpreting test-retest changes in CVMT results for elderly examinees. However, it must be noted that the present study did not include very short (i.e., < 9 months) or extended (i.e., 2 or more years) retest intervals and, as such, these findings have limited generalizability. Moreover, the educational level of our sample was relatively high and the age range did not include anyone younger than 55 years of age. Therefore, it is possible that education and age would have exerted a more meaningful impact on CVMT scores from test to retest had a more heterogeneous sample been examined. Additional research on the impact of demographic characteristics and retest interval on change scores is warranted. It should also be noted that clinical samples tend to be more heterogeneous in terms of level of test...
performance than cognitively normal samples and thus, test-retest correlations may be higher in clinical samples.

The test-retest stability coefficients for all of the CVMT scores attained by normal elderly were well below the recommended minimum of .80 for tests used in clinical decision-making (Anastasi, 1988; Sattler, 1988). The low stability coefficient for the Delayed Recognition Score may, in part, reflect a restricted range of scores (i.e., 0–7). However, the low coefficients for the other scores suggest that the rank order of examinees from test to retest is quite different. Comparisons of the coefficients in the test manual with the current values revealed substantially lower scores for the present participants. These lower coefficients may reflect differences in the retest interval since as the time between assessments increases, stability coefficients decrease. The interval in the test manual was 7 days, while the current retest interval averaged 1.1 years.

Inspection of test-retest stability coefficients of other memory tests reveals a wide range of values generally consistent with those found for the CVMT. Stability coefficients for the California Verbal Learning Test for a 1-year interval range from .12 to .79 ($M = .53$) for persons averaging 33 years of age (Delis, Kramer, Kaplan, & Ober, 1987) and range from .24 to .76 ($M = .47$) for persons averaging 70 years of age (Paolo, Tröster, & Ryan, 1997). Mitrushina and Satz (1991) reported 1-year stability coefficients for elderly persons (average age = 70 years) ranging from .41 to .68 on the Rey Auditory Verbal Learning Test, .64 to .68 on the Visual Reproduction subtest of the Wechsler Memory Scale, and .56 to .57 on the Rey Complex Figure. Although the CVMT yielded stability coefficients below the suggested minimum for tests used in clinical decision making, the CVMT demonstrates similar stability as other memory tests currently employed in neuropsychological assessment.

Mean comparisons revealed significant improvement from test to retest on d-Prime and Delayed Recognition. These findings, in conjunction with the low stability coefficients, suggest that meaningful practice effects and measurement error need to be considered when interpreting CVMT scores obtained on a second administration. $SEd$ and 90% cut-off values were provided to assist clinicians in the detection of statistically reliable differences when the CVMT is administered on two separate occasions. The $SEd$ provides a way of determining whether a particular test-retest change is beyond that expected from measurement error. It is suggested that any test-retest discrepancy that exceeds the 90% $SEd$ cut-offs be further investigated in order to determine whether it has significant clinical meaning.
REFERENCES


