ELSEVIER

Archives of Clinical Neuropsychology 21 (2006) 413–420

Archives of CLINICAL NEUROPSYCHOLOGY

The California Verbal Learning Test – second edition: Test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms

Steven Paul Woods [a,*], Dean C. Delis [b], J. Cobb Scott [c], Joel H. Kramer [d], James A. Holdnack [e]

[a] *Department of Psychiatry (0847), University of California at San Diego, 150 W. Washington Street, 2nd Floor, San Diego, CA 92103-2005, USA*

[b] *Department of Psychiatry (0847), University of California at San Diego, and Psychology Service (116B), VA San Diego Healthcare System, La Jolla, CA, USA*

[c] *Joint Doctoral Program in Clinical Psychology, San Diego State University and University of California at San Diego, San Diego, CA, USA*

[d] *Department of Neurology, University of California at San Francisco, San Francisco, CA, USA*

[e] *Harcourt Assessment, Inc., San Antonio, TX, USA*

Accepted 15 June 2006

Abstract

The California Verbal Learning Test – second edition (CVLT-II) is one of the most widely used neuropsychological tests in North America. The present study evaluated the 1-month test-retest reliability and practice effects associated with the standard and alternate forms of the CVLT-II in a sample of 195 healthy adults. Eighty participants underwent repeat assessment using the standard form of the CVLT-II on both occasions, whereas the remaining 115 individuals received the standard form at baseline and the alternate form at follow-up. Consistent with prior research, results revealed generally large test-retest correlation coefficients for the primary CVLT-II measures in both the standard/standard (range = 0.80–0.84) and standard/alternate (range = 0.61–0.73) cohorts. Despite exhibiting slightly lower test-retest reliability coefficients, participants in the alternate form group displayed notably smaller practice effects (Cohen's *d* range = −0.01 to 0.18) on the primary indices relative to individuals who received the standard form on both occasions (Cohen's *d* range = 0.27–0.61). Reliable change indices were also generated and applied to primary CVLT-II variables to determine the base rates of significant improvements (range = 2–10%), declines (range = 0–7%), and stability (range = 85–97%) in performance over time. Overall, findings from this study support the test-retest reliability of the standard and alternate forms of the CVLT-II in healthy adults and may enhance the usefulness of this test in longitudinal neuropsychological evaluations.

Keywords: Episodic memory; Verbal learning; Test reliability; Practice

The California Verbal Learning Test (CVLT and CVLT-II; Delis, Kramer, Kaplan, & Ober, 1987; Delis, Kramer, Kaplan, & Ober, 2000) is among the five most common assessment instruments used by clinical neuropsychologists in North America (Rabin, Barr, & Burton, 2005). The construct validity of the CVLT as a measure of episodic verbal learning and memory has garnered considerable support in the neuropsychological literature (e.g., Alexander, Stuss,

* Corresponding author. Tel.: +1 619 543 5004; fax: +1 619 543 1235.
E-mail address: spwoods@ucsd.edu (S.P. Woods).

0887-6177/$ – see front matter © 2006 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.acn.2006.06.002

& Fansabedian, 2003; Baldo, Delis, Kramer, & Shimamura, 2002; Crosson, Novack, Trenerry, & Craig, 1988; Kibby, Schmitter-Edgecombe, & Long, 1998). Prior studies also support the test-retest reliability of the original CVLT (e.g., Paolo, Tröster, & Ryan, 1997), with the traditional primary variables (e.g., total trials 1–5 and long-delay free recall) demonstrating particularly robust temporal stability in healthy adults (Delis et al., 1991).

In the only study published to date on the test-retest reliability of second edition of the CVLT, Benedict (2005) reported data on 34 participants with multiple sclerosis who were randomly assigned to receive either: (1) the CVLT-II standard form at baseline and the alternate form at 1-week follow-up; or (2) the standard form at both baseline and 1-week follow-up. Although test-retest reliability coefficients were broadly comparable for the standard ($M\ r = 0.62$, range = 0.50–0.72) and alternate form ($M\ r = 0.75$; range = 0.54–0.89) groups, participants who received the alternate form at retest exhibited notably smaller practice effects across the CVLT-II summary measures (alternate form $M\ d = 0.0$, range = −0.1 to 0.1; standard form $M\ d = 0.76$, range = 0.5–1.0). Findings were interpreted to suggest that use of the CVLT-II alternate form might diminish the confounding effects of practice across repeated administrations without adversely affecting reliability. However, no peer-reviewed studies have been published on the test-retest reliability of the CVLT-II in a nonclinical group, which may yield different results as compared to a disease sample (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003) and provide more broadly applicable psychometric data for clinical and research application. The present study therefore aimed to examine the test-retest reliability and practice effects of the standard and alternate forms of the CVLT-II in healthy adults, as well as to generate reliable change indices (RCIs; Jacobson & Truax, 1991) to provide statistical guidelines for detecting significant changes in individual CVLT-II profiles.

1. Method

Participants were 195 healthy adults who underwent repeat testing with the CVLT-II over at least a 1-week test-retest interval. The average test-retest interval was 29 days (S.D. = 13, range = 9–74). All potential study participants were screened for histories of medical, neurological, or psychiatric conditions known to adversely affect neurocognitive functions (see the CVLT-II technical manual for further details). Eighty participants underwent repeat assessment using the standard form of the CVLT-II on both occasions (standard/standard), whereas the other 115 individuals received the standard form at baseline and the alternate form (standard/alternate) at follow-up. The demographic characteristics of the study groups and their test-retest interval data are displayed in Table 1.

Raw scores were used for all analyses. Wilcoxon Signed Rank tests and Spearman's rho (ρ) correlation coefficients (or their parametric counterparts as determined by results from the Shapiro-Wilk W test of normality) were conducted to assess practice effects and test-retest reliability, respectively. The critical alpha level was set at 0.001 for these analyses

Table 1
Demographic composition of the test-retest study samples

Variable	Standard/standard forms ($N = 80$)			Standard/alternate forms ($N = 115$)		
	Mean	S.D.	Range	Mean	S.D.	Range
Test-retest interval (days)	25.8	13.1	9–74	30.9	13.1	10–74
Age (years)	49.5	22.7	16–88	47.7	22.0	16–88
Education (%)						
≤8 years	7.5%			5.2%		
9–11 years	8.8%			15.7%		
12 years	37.5%			37.4%		
13–15 years	33.8%			21.7%		
≥16 years	12.5%			20.0%		
Sex (%)						
Female	51.3%			53.9%		
Male	48.8%			46.1%		
Ethnicity (%)						
Caucasian	77.5%			83.5%		
Hispanic	10.0%			7.8%		
African-American	10.0%			8.7%		
Other	2.5%			0.0%		