Pearson

Substitution of CVLT–3 Scores for WMS–IV Verbal Paired Associates Scores

James A. Holdnack, Lisa Whipple Drozdick, and Troy G. Courville

Abstract

Score substitution usually is associated with measures of cognitive ability and reserved for unintended administration problems. Score substitution enables clinicians to obtain alternate index scores when standardized test administration has been violated. The WMS–IV introduced a score substitution model that enables clinicians to intentionally use scores from the CVLT–II instead of Verbal Paired Associates in WMS–IV index scores. This study updates the substitution method, using the CVLT–3. Methods: Participants were 380 adults, ages 16–90, who participated in the CVLT–II/WMS–IV linking study. The CVLT–3 learning, short-, and long-delay recall trials contain the same words and administration as the CVLT–II. All CVLT–II scores were updated with the current norms derived for the CVLT–3. Results: CVLT–3 scores moderately correlated with WMS–IV VPA and LM subtests, and are consistent with the correlation observed between VPA and LM. Scores from CVLT–3 were scaled to WMS–IV norms using equipercentile equating. Discussion: CVLT–3 substitution for VPA does not presume construct equivalence; the equating reflects a rescaling of CVLT–3 to WMS–IV norms. This substitution model is designed to improve clinicians' workflow and allows for intentional, a priori, decisions to be made regarding the best memory model for a specific diagnostic hypothesis. Rationale and caveats for use are discussed.

Introduction

The substitution of one score for another in the computation of composite or index scores is a common practice typically associated with measures of general cognitive ability (e.g., Wechsler, 1997; Wechsler, 2009). On measures of general cognitive functioning, a supplementary score typically can be used in an index only when a violation of the standard administration for the core subtest in that domain has occurred. Developing alternate ability scores using substituted subtests is ***discouraged.*** Substitution of scores is less common in neuropsychological measures and serves a different purpose than the traditional "spoiled" subtest model utilized in a general cognitive ability test. This new use of score substitution was first introduced in the Wechsler Memory Scale, fourth edition (WMS–IV; Wechsler, 2009), to create an alternative model for auditory memory assessment by allowing the substitution of measures of verbal learning.

This score substitution enables the examiner to select the subtests most relevant to a particular client when creating the Verbal Memory Indexes. For example, in 2004, McDowell, Bayless, Moser, Meyers, and Paulsen demonstrated that the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987) word list was more sensitive to memory impairment than the WMS–III word list. Because of potential differential sensitivity among verbal memory measures, WMS–IV developers derived CVLT–II equivalent scaled scores to allow substitution of CVLT–II scores for Verbal Paired Associates scores in deriving index standard scores.

The use of the California Verbal Learning Test, second edition (CVLT–II; Delis et al., 1999) scores as a substitute for the WMS–IV Verbal Paired Associates (VPA) subtest scores involves the ***intentional*** decision of the clinician to measure components of verbal memory better evaluated with the CVLT–II list learning test than VPA. Two studies have evaluated the substitution of CVLT–II for VPA in clinical (Miller et al., 2012) and high-ability (Thiruselvam, Vogt, & Hoelzle, 2015) populations. Both studies report that the substitution yields somewhat different index scores in these populations. In the clinical sample, the substitution of CVLT–II/VPA equated scores yielded lower Auditory Memory Index scores (AMI), compared to the standard derivation of that index in a mixed clinical population (Miller et al., 2012). The actual AMI values in the study were 86.8 (+ 16.1) and 85.7 (+ 15.3) for the standard and substituted index scores, respectively (Miller et al., 2012). The large sample size (n = 261) resulted in statistical significance, but the overall performance in both groups was remarkably similar at the group level. In the high-ability college student study, the authors note a low correlation between CVLT–II and VPA scores, which is attributable to the known ceiling effects with VPA in younger age groups (Thiruselvam et al., 2015). The AMI standard and equated scores from that study were 106.5 (+ 10.3) and 103.8 (+ 10.5), respectively (Thiruselvam et al., 2015). The difference at the group level is larger than reported in the clinical sample, yet the overall difference is relatively small, 2.7 standard score points or < .3 standard deviations.

Both studies further analyzed the data to determine the rate of large differences (e.g., + 1 SD) in examinee performance when substitution is applied. In the clinical sample, 8.4% of examinees' AMI scores differed by one or more standard deviations with a bias toward substituted scores being lower than the standard score (6.1% versus 2.3%: Miller et al., 2012). In the high-ability group, the results were similar with 7.5% of examinees' scores differing by method, with all cases having lower scores when substitution is used (Thiruselvam et al., 2015). The CVLT–II scores tend to be slightly lower in both clinical and young high-ability individuals. Neither result was unexpected due to the differences between VPA and CVLT–II; the CVLT–II has more items to recall and includes recall interference in delayed trials. The studies reported correlations slightly lower in the clinical sample and substantially lower in the high-ability sample than were found in the original sample (Wechsler, 2009). Both studies generally expressed concern about substitution when the correlations are in the moderate range; however, the correlations reported for VPA with CVLT–II (.5s: Wechsler, 2009) are similar to those reported for allowable substitutions in the WAIS–IV (e.g., Picture Completion, the substitutable PRI subtest, and core subtests within the Perceptual Reasoning Index [Wechsler, 2008]). Both studies expressed concerns with the equivalence of the AMI when substitution is applied.

The current study describes the development of alternate Auditory, Immediate, and Delayed index scores for the WMS–IV, incorporating scores from the CVLT–3. Examiners were able to substitute the CVLT–II Trials 1–5 Total Correct score and Long-Delay Free Recall Correct scores for the Verbal Paired Associates Immediate and Delayed recall scores, respectively. Given that the two tests were standardized on two different samples and at two different times, equipercentile equating (Kolen & Brennan, 2009) methods were used to adjust CVLT–II scores to the WMS–IV normative scores. With the release of the CVLT–3, new score equating is required to adjust the CVLT–3 to the WMS–IV. The intent of the analysis was not to establish equivalence between the scores, but to scale the CVLT–3 data to the WMS–IV normative data. The current study extends the original substitution model to allow CVLT–3 cued recall trials to be substituted for VPA I and II in the various indexes. VPA measures verbal learning through a cued association paradigm that is procedurally closer to the CVLT–3 cued recall trials more than free recall trials.

Method

Participants

Equating Sample

The WMS–IV and CVLT–II were administered to a sample of 380 examinees, ages 16–90, as part of the WMS–IV standardization. The sample was stratified in two broad age groups, adult and older adult, to accommodate slight differences in the Verbal Paired Associates test structure between older (OA: Older Adult Battery) and younger examinees (FA: Full Adult battery). Table 1 presents the demographic characteristics of the samples completing the OA and FA batteries. The normed scores used in this initial study were based on the CVLT–II (Delis et al., 2000) standardization.

Table 1. Demographic Characteristics of CVLT–3 and WMS–IV Equating Sample

	Full Adult	Older Adult
N	190	190
Age		
Mean	40.4	75.7
SD	17.9	7.2
Range	16–69	65–90
Sex		
Female	51.6	60.0
Male	48.4	40.0
Education		
8 or fewer years	5.8	11.6
9–11 years	5.8	15.3
12 years	31.0	34.2
13–15 years	30.0	18.9
16 or more years	27.4	20.0
Race/Ethnicity		
African American	11.6	10.5
Asian	4.2	1.0
Hispanic	13.7	6.3
Other	2.6	1.2
White	67.9	81.0

Measures

CVLT–II/CVLT–3

The CVLT–II/CVLT–3 is a 16-word, list-recall task. The recall items can be organized into four semantically related categories. The examiner reads the word list (words are not presented by category) to the examinee, who must recall as many of the words from the list as possible. There are five learning trials:

- an interference trial
- short-delay free and cued (e.g., examinees' recall is prompted by giving them the semantic category) recall trials
- long-delay free and cued recall trials
- a recognition trial
- a forced-choice recognition trial

The test measures multiple aspects of verbal leaning, including rate of acquisition, strategy, interference, recall monitoring, encoding, retrieval, forgetting, and effort. The CVLT–II/CVLT–3 has a standard, an alternate, and a brief form. All examinees were administered the standard form; however, only the total learning score trials 1–5, short-delay cued recall, long-delay free recall, and long-delay cued recall scores were used in this study.

The CVLT–3 (Delis et al., 2017) was standardized on a large sample of adults (ages 16–90) and stratified by age, education, race/ethnicity and region, based on the 2015 census update. The test content remained unchanged from the CVLT–II for all forms, except the forced-choice recognition trial, in which half of the items were replaced. Score scales were changed from the z- and T-score metric to scaled scores and standard scores. Finally, all measures are referenced to the examinee's age on CVLT–3; CVLT–II scores were adjusted by age and sex. Education- and demographic-adjusted scores are available for the CVLT–3; however, for the purposes of this study, only age-referenced scaled or standard scores are utilized.

WMS–IV

The WMS–IV is a battery of auditory and visual memory tasks that measure story memory, verbal learning, memory for visual-spatial information, and visual working memory. The WMS–IV index structure includes Immediate Memory, Delayed Memory, Auditory Memory, Visual Memory, and Visual Working Memory. For the purposes of this study, only the Immediate, Delayed, and Auditory Indexes were utilized. The WMS–IV, the most recent edition of the test battery, was standardized in 2009 on a large, stratified (2005 census update) sample of adults (ages 16–90).

The WMS–IV contains two auditory memory subtests: Logical Memory (LM) and Verbal-Paired Associates (VPA). Each subtest includes immediate recall, delayed recall, and delayed recognition trials. The WMS–IV auditory subtests vary in content for age groups 16–69 (FA: Full-Adult battery) and 65–90 (OA: Older Adult battery). For Logical Memory, examinees who took the FA battery were administered two stories of about 7–8 sentences each, without repetition. Examinees who completed the OA battery were administered a shorter story

(5 sentences) and one of the FA battery stories that was repeated once. The VPA subtest on both batteries includes four learning trials; however, the FA battery presents 14 word pairs and the OA battery presents 10 word pairs for the examinee to remember. Logical Memory measures the examinee's ability to recall semantically organized content in the form of a story and VPA measures the examinee's ability to learn semantically related and unrelated pairs of words. LM I and II measure free recall ability and VPA I and II measure cued recall ability.

Results

The data analysis was completed using Version 9.4 of the SAS System for Windows 7 (W32_7PRO platform). Table 2 lists the abbreviations used in the results text and tables.

Table 2. WMS–IV and CVLT–3 Subtest Abbreviations

Test/Score	Abbreviation
WMS–IV	
Logical Memory Immediate	LM I
Logical Memory Delayed	LM II
Verbal Paired Associates Immediate	VPA I
Verbal Paired Associates Delayed	VPA II
Designs Immediate	DE I
Designs Delayed	DE II
Visual Reproduction Immediate	VR I
Visual Reproduction Delayed	VR II
Spatial Addition	SA
Symbol Span	SSP
CVLT–3	
Trials 1-5 Correct	CV15
Short Delay Cued Recall Correct	CVSDC
Long Delay Free Recall Correct	CVLDF
Long Delay Cued Recall Correct	CVLDC

WMS–IV and CVLT–3 Correlations

CVLT–3 scores were derived for examinees who participated in the WMS–IV/CVLT–II validity study. The correlation between the original CVLT–II normed scores and new CVLT–3 scores ranged from .91 (short-delay cued recall) to .95 (trials 1–5 total correct). Correlation of CVLT–3 selected scores and the WMS–IV core measures are presented for WMS–IV FA (Table 3), WMS–IV OA (Table 4), and the combined (Table 5) samples. LM I and II correlate in the moderate range (.4s) with both VPA I and II, and the four CVLT–3 measures in the WMS–IV FA battery (ages 16–69). For the WMS–IV OA battery, the correlations between LM I and II and VPA I and II are moderate (.4s) similar to that observed in the WMS–IV FA battery. However, the LM II correlations with the CVLT–3 measures was generally higher (.4s–.6s) than for VPA. The higher correlation between CVLT–3 and LM II is also observed in the total sample. The CVLT–3 cued recall conditions did not have a higher correlation with VPA I and II in comparison to the CVLT–3 free recall scores, in any of the samples.

The correlations among immediate recall mesasures reveal a higher correlation of CVLT–3 Trials 1–5 total correct compared to VPA I with VR I, but not DE I. However, CVLT–3 cued recall correct did not have a higher correlation with the visual immediate measures compared to VPA. Among delayed recall measures, the correlation were similar between CVLT–3 delayed measures and VR II and DE II, to those observed with VPA II.

Table 3. Score Intercorrelations for Full Adult Battery, Ages 16–69

Score	LM I	LM II	VPA I	VPA II	DE I	DE II	VR I	VR II	SA	SSP	CV15	CVSDC	CVLDF	CVLDC
LM I														
LM II	.86													
VPA I	.44	.43												
VPA II	.42	.46	.85											
DE I	.28	.32	.32	.33										
DE II	.26	.29	.35	.35	.69									
VR I	.39	.32	.41	.35	.37	.32								
VR II	.27	.29	.33	.30	.39	.44	.51							
SA	.38	.36	.37	.37	.44	.44	.40	.31						
SSP	.52	.49	.44	.41	.50	.40	.47	.40	.40					
CV15	.49	.51	.53	.53	.33	.31	.48	.34	.42	.48				
CVSDC	.41	.48	.48	.50	.31	.33	.42	.32	.32	.33	.72			
CVLDF	.45	.49	.56	.54	.35	.38	.42	.32	.36	.44	.80	.86		
CVLDC	.46	.48	.50	.49	.34	.32	.44	.28	.33	.39	.75	.88	.87	
Mean	10.2	10.4	10.1	10.0	10.6	10.4	10.4	10.4	10.2	10.2	102.8	10.6	10.4	10.4
SD	3.1	3.1	3.0	3.1	3.0	2.8	2.8	2.7	3.0	3.1	15.3	3.3	3.0	3.2

Note. CVLT–3 Trials 1–5 standard score uses a mean of 100 and a standard deviation of 15. All other score are scaled scores with a mean of 10 and a standard deviation of 3.

Table 4. Score Intercorrelations for Older Adult Battery, Ages 65–90

Score	LM I	LM II	VPA I	VPA II	VR I	VR II	SSP	CV15	CVSDC	CVLDF	CVLDC
LM I											
LM II	.77										
VPA I	.45	.49									
VPA II	.42	.47	.85								
VR I	.29	.24	.40	.36							
VR II	.15	.20	.33	.32	.65						
SSP	.40	.37	.38	.35	.58	.51					
CV15	.55	.54	.57	.58	.34	.26	.43				
CVSDC	.47	.57	.56	.55	.27	.18	.37	.76			
CVLDF	.45	.60	.51	.54	.27	.28	.37	.77	.83		
CVLDC	.46	.60	.51	.52	.26	.20	.36	.73	.90	.89	
Mean	10.0	10.0	10.1	10.0	10.1	10.0	9.9	102.2	10.0	9.6	9.9
SD	3.0	3.1	2.9	3.0	3.2	2.8	3.1	15.8	3.4	3.2	3.4

Note. CVLT–3 Trials 1–5 standard score uses a mean of 100 and a standard deviation of 15. All other score are scaled scores with a mean of 10 and a standard deviation of 3.

Table 5. Score Intercorrelations, All Ages

Score	LM I	LM II	VPA I	VPA II	DE I	DE II	VR I	VR II	SA	SSP	CV15	CVSDC	CVLDF	CVLDC
LM I														
LM II	.81													
VPA I	.45	.46												
VPA II	.42	.47	.85											
DE I	.28	.32	.32	.33										
DE II	.26	.29	.35	.35	.69									
VR I	.34	.28	.40	.36	.37	.32								
VR II	.21	.25	.33	.31	.39	.44	.59							
SA	.38	.36	.37	.37	.44	.44	.40	.31						
SSP	.46	.43	.41	.38	.50	.40	.53	.46	.40					
CV15	.52	.52	.55	.55	.33	.31	.40	.30	.42	.46				
CVSDC	.44	.52	.52	.53	.31	.33	.34	.25	.32	.36	.74			
CVLDF	.45	.55	.53	.53	.35	.38	.35	.30	.36	.40	.78	.85		
CVLDC	.46	.55	.50	.50	.34	.32	.34	.24	.33	.38	.74	.89	.89	
Mean	10.1	10.2	10.1	10.0	10.6	10.4	10.3	10.2	10.2	10.0	102.5	10.3	10.0	10.1
SD	3.0	3.1	2.9	3.0	3.0	2.8	3.0	2.8	3.0	3.1	15.5	3.4	3.1	3.3

Note. CVLT–3 Trials 1–5 standard score uses a mean of 100 and a standard deviation of 15. All other score are scaled scores with a mean of 10 and a standard deviation of 3.

Equating

Prior to equating, the means, standard deviations, and skew for each score was evaluated for examinees who completed the WMS–IV FA and WMS–IV OA editions of VPA I and II and associated CVLT–3 scores. This analysis revealed that the CVLT–3 scores in the sample completing the WMS–IV FA battery showed a greater negative skew and slightly higher overall mean performance (e.g., age referenced scaled scores) compared to score distribution of the sample completing the WMS–IV OA. Initial attempts at equating, using the entire sample (as was done with the CVLT–II scores), revealed data fit issues (e.g., higher or lower differences than expected at the tails of the distribution) in the WMS–IV FA sample. These results, combined with some differences in the correlations among the CVLT–3 and WMS–IV measures between the samples, suggested that the sample (for this edition) should not be combined for equating.

Frequency distributions for VPA I and II and the four CVLT–3 measures (Trials 1–5 total correct, short-delay cued recall, long-delay free recall, and long-delay cued recall) were calculated for each sample independently. The frequency distributions were converted to mid-point percentiles to z-normalized distributions. CVLT–3 scores were aligned to VPA scores in the current sample via equipercentile equating and then linked back to the VPA scaled scores. Table 6 presents the CVLT–3 to VPA scaled score equivalents for immediate and delayed recall measures, by WMS–IV battery.

Table 6. CVLT–3 to VPA Equivalent Scores by Battery

	Full Adult				Older Adult			
VPA equivalent	Trials 1–5 correct	Short-Delay Cued correct	Long-Delay Free correct	Long-Delay Cued correct	Trials 1–5 correct	Short-Delay Cued correct	Long-Delay Free correct	Long-Delay Cued correct
1	< 61	1	1	1	< 67	1–3	1	—
2	61–66	2	2	2	67–68		2	1
3	67–70	3–4	3	3	69–71		—	2
4	71–75	5	4–5	4–5	72–73	4	3	3
5	76–79	6	6	6	74–79	—	4	4–5
6	80–84	—	—	—	80–84	5	5–6	—
7	85–90	7	7	7	85–89	6	—	6
8	91–94	8	8	8	90–93	7–8	7	7–8
9	95–100	9	9	9	94–99	9	8	9
10	101–106	10	10	10	100–102	10	9–10	10
11	107–110	11–12	11–12	11–12	103–109	11	11	11
12	111–116	13	13	13	110–117	12	12	12
13	117–119	14–15	14	14	118–120	13–14	13	13–14
14	120–126	16	—	—	121–127	15	14	15
15	127–128	17	15	15	128–129	16	15	—
16	129–133	—	16	16	130–139	17	—	16–17
17	134–135	—	17	17	140–145	18	16	–
18	136–140	—	18	18	146	–	17–18	18–19
19	> 140	18–19	19	19	> 146	19	19	—

Performance differences were observed when using the CVLT–3 – VPA equated scores in the WMS–IV indexes. Table 7 presents descriptive statistics and base rates comparing standard WMS–IV index scores and WMS–IV indexes using the CVLT–3 score substitution. In the overall sample, 90% of individuals' Auditory Memory Index scores were within 10 points (2/3 of a standard deviation) across age groups when the VPA or CVLT–3 free recall measures were used in the index. For the Immediate and Delayed Memory Indexes, 90% were within 7 points (1/2 a standard deviation) across age groups. Reliability and standard error of measurement for CVLT–3 substituted WMS–IV index scores are provided in Tables 8 and 9.

Table 7. Statistics for Pairwise Comparisons of Standard WMS–IV and CVLT–3 Substituted Index Scores

Reference group	Comparison	Average Difference		Base rate							
				≤ 25%		≤ 10%		≤ 5%		≤ 1%	
		Mean	SD	(+)	(–)	(+)	(–)	(+)	(–)	(+)	(–)
CLVT Free Recall Substitution											
Overall sample	Auditory STD-CVLT3	.03	7.6	5	5	9	10	12	13	17	18
	Immediate STD-CVLT3	.25	5.4	4	4	7	7	10	8	13	11
	Delayed STD-CVLT3	–0.13	5.7	4	4	7	7	9	10	12	14
16–69	Auditory STD-CVLT3	.07	7.7	6	5	9	10	10	12	19	20
	Immediate STD-CVLT3	.29	5.0	4	3	7	6	8	8	12	11
	Delayed STD-CVLT3	–0.1	5.1	4	4	6	6	7	9	12	14
65–90	Auditory STD-CVLT3	–0.02	7.5	5	5	9.5	10	13	13	17	17
	Immediate STD-CVLT3	.22	5.9	4	4	8.5	7	11	9	15	12
	Delayed STD-CVLT3	–0.16	6.2	4	4	8	9	10	11	14	16
CVLT–3 Cued Recall Substitution											
Overall sample	Auditory STD-CVLT3	–0.06	8.2	5	5	10	11	13	13	19	21
	Immediate STD-CVLT3	.12	5.6	4	3	7	7	9	9	13	16
	Delayed STD-CVLT3	–0.04	5.9	4	4	7	8	9	10	14	16
16–69	Auditory STD-CVLT3	–0.23	8.4	5	5	10	11	13	13	21	26
	Immediate STD-CVLT3	0.17	5.1	3	3	7	6	8	8	15	16
	Delayed STD-CVLT3	–0.25	5.6	3	4	6	7	8	10	14	16
65–90	Auditory STD-CVLT3	0.10	8.0	5	6	10	10	12	14	19	21
	Immediate STD-CVLT3	0.07	6.0	4	4	7	7.5	9	10	13	17
	Delayed STD-CVLT3	0.16	6.3	4	4	8	8	10	10	16	15

Table 8. Reliability of WMS–IV Composites Applying VPA CVLT–3 Equivalent Scores

Score	Age group		
	16–69	65–90	Total
CVLT–3 Free Recall scores			
Auditory Memory	.91	.93	.92
Immediate Memory	.93	.92	.93
Delayed Memory	.93	.93	.94
CVLT–3 Cued Recall Scores			
Auditory Memory	.89	.92	.91
Immediate Memory	.90	.95	.91
Delayed Memory	.91	.90	.92

Table 9. Standard Error of Measurement of WMS–IV Composites Applying VPA CVLT–3 Equivalent Scores

	Age group		
Score	**16–69**	**65–90**	**Total**
CVLT–3 Free Recall Scores			
Auditory Memory	4.5	4.0	4.2
Immediate Memory	4.0	4.2	4.0
Delayed Memory	4.0	4.0	3.7
CVLT–3 Cued Recall Scores			
Auditory Memory	5.0	4.2	4.5
Immediate Memory	4.7	3.4	4.5
Delayed Memory	4.5	4.7	4.2

Discussion

This study derived alternate, equated WMS–IV auditory, immediate, and delayed indexes by substituting the CVLT–3 free recall or cued recall scores for WMS–IV VPA. The current results update the published equating tables (Wechsler, 2009), using the most recent CVLT–3 normative data. For this study, a new sample of cases was not collected given that the CVLT–II and CVLT–3 contain the exact same words, rather the existing study was updated by applying the new norms and norms format (e.g., age scaled scores and not age and sex z-scores).

The CVLT–3 free and cued recall measures had moderate correlations with both VPA and LM from the WMS–IV. The correlation between VPA I and II with LM I and II was also in the moderate range and did not show a significantly different relationship from that observed with the CVLT–3 scores. Similarly, the CVLT–3 scores correlated at a low to moderate range with other WMS–IV subtests, consistent with reported correlations for VPA I and II. The similarity of the correlations among the measures indicates that CVLT–3 scores are equally related to the constructs of immediate, delayed, and auditory memory as VPA scores.

Alternate index scores were created through equipercentile equating. The scores were calibrated back to the WMS–IV normative data to be consistent with the other scores appearing in the specific indexes. For the CVLT–3 substitution, the age referenced scaled scores were used in the alternate indexes, whereas in the previous model (e.g., CVLT–II) the age and sex referenced scores were equated to the WMS–IV VPA scaled scores. The scores also were equated by WMS–IV form. In this study, there were small differences in the distribution of CVLT–3 scores in the samples taking the WMS–IV FA and WMS–IV OA versions of VPA. Using the entire sample resulted in larger score differences between the standard index scores and the CVLT–3 index scores, than equating the scores by VPA form.

The moderate correlations between VPA and CVLT–3 scores will result in similar, but not identical, results when the CVLT–3 substituted scores are used. In most cases, CVLT–3 substituted index scores will vary by only 3–5 points higher or lower than the standard VPA based index scores. Larger differences can occur in some cases, even among typically developing and aging individuals.

Rationale for Use

Many CVLT–3 users also administer some or all the subtests from the WMS–IV. The CVLT–3 substitution model was designed to shorten administration time for examiners who wished to have WMS–IV index scores but did not wish to administer VPA in addition to the CVLT–3. Examiner's using the CVLT–3 can compare performance on memory tests compared to overall cognitive functioning or compare auditory and visual memory functions. Index scores are also more reliable than individual scores and cover a broader range of memory functions than individual subtest scores.

Clinicians unfamiliar with the CVLT–3 may wish to evaluate its utility relative to VPA. In particular, the CVLT–3 provides information regarding the impact of executive functioning on verbal learning. Additionally, the clinician may test hypotheses regarding the effects of proactive and retroactive interference and specific types of memory errors (e.g., semantic versus unrelated intrusion errors). The VPA subtests provide information regarding memory errors as related to cued recall (Pearson, 2009) and components of memory, but to a lesser degree.

The CVLT–3 list-learning task differs from VPA in many respects, but the primary difference relates to learning and recall of information. On the CVLT–3, subsets of words are semantically related, so encoding may be facilitated by strategic use of the associations among items. Only a small subset of VPA items is semantically related, such that a specific learning strategy is not reinforced on the test. The second difference is that CVLT–3 is a *free recall* test in which the examinee must say the word from memory in the absence of a specific cue. Even in the cued recall condition, the examinee must have organized the information by category to facilitate recall and recall is still free (e.g., not-linked to a specific associated word). On VPA, the examinee must store information as a specific association to the paired word and the examiner drives the recall by asking about each word. Finally, the CVLT–3 requires the examinee to learn a longer list of words (e.g., 16 versus 14 or 16 versus 10).

Though the tests measure similar components of verbal learning, they are quite different in structure of encoding and recall. Each test is not considered equivalent nor equally sensitive to memory difficulties associated with different clinical populations. The examiner must decide which test provides the best assessment of verbal learning for the examinee. This is facilitated by a review of the literature, which is beyond the scope of this study. The CVLT–3 and Logical Memory subtests have a very long and large research base that exceeds the VPA subtest research base, which may also be a consideration in deciding which measure to use.

Caveats

The current study is based on performance data collected from 2007–2008. Though the CVLT–3 content remains unchanged, it is assumed that recall in the study group would have been replicated. Studies indicate that the CVLT–3 may be more difficult than VPA for some groups, particularly younger high-ability examinees. Ceiling effects on VPA persist and can affect the use of substitution in these samples; however, these results appear more prominent when the cued measures are used, rather than free recall measures (see Table 7). The clinician should be aware that, though 95% of scores are within one standard deviation when comparing VPA and substituted index scores in this sample, percentages may be higher in clinical samples. Also, 25% of scores are outside of a 5-point difference, which could result in a change in classification. Therefore, the clinician should choose one model based on the needs of the examinee and not administer both scales to identify the lowest score. Doing so would capitalize on known effects of administering multiple tests on rates of low scores, possibly leading to misdiagnosis (Brooks et al., 2008, 2009, 2011). The clinician should be aware that the standard and substituted index scores are not equivalent in terms of the constructs measured. The scores are equated to scale the tests to a common performance metric.

References

Brooks, B., Holdnack, J., & Iverson, G. (2011). Advanced clinical interpretation of the WAIS–IV and WMS–IV: Prevalence of low scores varies by level of intelligence and years of education. *Assessment, 18,* 156–167.

Brooks, B. L., Iverson, G. L., Feldman, H. H., & Holdnack, J. A. (2009). Minimizing misdiagnosis: Psychometric guidelines for determining possible and probable memory impairment in older adults. *Dementia and Geriatric Cognitive Disorders, 27,* 439–450.

Brooks, B. L., Iverson, G. L., Holdnack, J. A., & Feldman, H. H. (2008). The Potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale–III in healthy older adults. *Journal of the International Neuropsychological Society, 14,* 463–478.

Kolen, M. J., & Brennan, R. L. (2009). *Test equating, scaling, and linking: Methods and practices.* New York: Springer-Verlag.

Miller, J. B., Axelrod, B. N., Rapport, L. J., Hanks, R. A., Bashem, J. R., & Schutte, C. (2012). Substitution of California Verbal Learning Test, second edition for Verbal Paired Associates on the Wechsler Memory Scale, fourth edition. *The Clinical Neuropsychologist, 26,* 599–608, DOI: 10.1080/13854046.2012.677478

Pearson. (2009). *Advanced Clinical Solutions for use with WAIS–IV and WMS–IV.* San Antonio, TX: Author.

Thiruselvam, I., Vogt, E. M., & Hoelzle, J. B. (2015). The Interchangeability of CVLT–II and WMS–IV Verbal Paired Associates Scores: A slightly different story. *Archives of Clinical Neuropsychology, 30,* 248–255

Wechsler, D. (1997). *Wechsler adult intelligence scale,* third edition. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler adult intelligence scale,* fourth edition. San Antonio, TX: Pearson.

Wechsler, D. (2009). *Wechsler memory scale,* fourth edition. San Antonio, TX: Pearson.

Pearson