

Marquette University
e-Publications@Marquette

Psychology Faculty Research and Publications

Psychology, Department of

5-1-2015

The Interchangeability of CVLT-II and WMS-IV Verbal Paired Associates Scores: A Slightly Different Story

Indrani K. Thiruselvam
Marquette University

Elisabeth M. Vogt
Marquette University

James B. Hoelzle
Marquette University, james.hoelzle@marquette.edu

Accepted version. *Archives of Clinical Psychology*, Vol. 30, No. 3 (May 2015): 248-255. DOI. © 2015
The Author. Published by Oxford University Press. Used with permission.

The Interchangeability of CVLT-II and WMS-IV Verbal Paired Associates Scores: A Slightly Different Story

Indrani Thiruselvam

*Department of Psychology, Marquette University,
Milwaukee, WI*

Elisabeth M. Vogt

*Department of Psychology, Marquette University,
Milwaukee, WI*

James B. Hoelzle

*Department of Psychology, Marquette University,
Milwaukee, WI*

Abstract

We investigated the similarity of the Wechsler Memory Scale-Fourth Edition (WMS-IV) Auditory Memory Index (AMI) scores when California Verbal Learning Test-Second Edition (CVLT-II) scores are substituted for WMS-IV Verbal Paired Associates (VPA) subtest scores. College students ($n = 103$) were administered select WMS-IV subtests and the CVLT-II in a randomized

Archives of Clinical Neuropsychology, Vol 30, No. 3 (May 2015): pg. 248-255. [DOI](#). This article is © Oxford University Press and permission has been granted for this version to appear in e-Publications@Marquette. Oxford University Press does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Oxford University Press.

order. Immediate and delayed VPA scaled scores were significantly greater than VPA substitute scaled scores derived from CVLT-II performance. At the Index level, AMI scores were significantly lower when CVLT-II scores were used in place of VPA scores. It is important that clinicians recognize the accepted substitution of CVLT-II scores can result in WMS-IV scores that are inconsistent with those derived from standard administration. Psychometric issues that plausibly contribute to these differences and clinical implications are discussed.

Keywords: Assessment, Learning and Memory, Test construction, Wechsler Memory Scale-Fourth Edition, California Verbal Learning Test-Second Edition, Verbal Paired Associates

Topic: psychometrics, mental recall, verbal learning, memory, valproic acid, college student

Introduction

Clinical neuropsychologists routinely evaluate and quantify memory functioning during clinical examinations. It is an essential cognitive construct to consider during the differential diagnosis process. For example, patients with Alzheimer's disease demonstrate more impaired episodic memory whereas patients with vascular dementia demonstrate more impaired semantic memory (Graham, Emery, & Hodges, 2004). The construct is also essential to consider when developing treatment plans. For example, verbal memory functioning is a strong predictor of post-surgical outcome for individuals with epilepsy (Breier et al., 1996; Helmstaedter & Elger, 1996).

A host of stand-alone memory tests and batteries have been developed to assist clinicians in quantifying auditory, visual, immediate, delayed, cued, free recall, and recognition memory (e.g., Wilson, 2002). Survey findings suggest that the Wechsler Memory Scale (WMS) is one of the most frequently utilized measure to evaluate memory functioning (Rabin, Barr, & Burton, 2005). The WMS battery has undergone a number of revisions with each new edition. Despite its wide use, some researchers question if changes have meaningfully improved the clinical utility of the measure (Loring & Bauer, 2010). While there is evidence that the most recent test edition has improved psychometric properties, the relative value of this is

unknown. Hoelzle, Nelson, and Smith (2011) found that the dimensional structure underlying the Wechsler Memory Scale-Fourth Edition (WMS-IV; Wechsler, 2009) was more differentiated than the Wechsler Memory Scale-Third Edition (WMS-III; Wechsler, 1997), but it is unknown how this difference may affect clinical decision making. Nevertheless, literature is emerging that supports the construct validity of the WMS-IV in individuals with traumatic brain injury (Carlozzi, Grech, & Tulskey, 2013) and amnesic mild cognitive impairment (Pike et al., 2013).

The WMS-IV attempts to quantify five different types of memory functioning. This study focuses on auditory memory, which is primarily reflected in the Auditory Memory Index (AMI) score, and is evaluated with Logical Memory (LM) and Verbal Paired Associates (VPA) immediate and delayed subtests. LM entails the immediate and delayed recall of two short stories. VPA involves four learning trials of 14 word pairs, and the subsequent immediate and delayed recall of these word pairs. A unique feature of the WMS-IV, relative to earlier versions of the WMS, is the option of replacing VPA scores with scores obtained from the California Verbal Learning Test—Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000). The CVLT-II is a commonly administered word-learning task (Rabin et al., 2005) in which an examinee is provided a list of 16 words and asked to recall as many words as possible across a number of immediate and delayed trials.

The WMS-IV Technical and Interpretive Manual (Wechsler, 2009, p. 166) acknowledges that the CVLT-II is inherently different from the VPA subtest, and has different normative bases and score metrics. Only moderate correlations are observed between the two tests. Specifically, the correlation between VPA I scaled scores and CVLT-II Trials 1–5 Free-Recall *T* scores is 0.54 and the correlation between VPA II scaled scores and CVLT-II Long-Delay Free-Recall *z* scores is 0.51 in a large normative sample (Wechsler, 2009). Miller and colleagues (2012) speculate that the moderate correlation between VPA and CVLT-II is explained by task discrepancies, such as the explicit associative learning and cued-recall format of the VPA as opposed to the implicit structure and generally free-recall format of the CVLT-II. VPA also allows for potentially richer learning opportunities as test takers are given immediate feedback after each

cue, whereas the CVLT-II does not allow for any performance feedback. Further, there are meaningful differences in the range of possible CVLT-II and VPA scores (i.e., floor and ceiling effects) that may also impact the relationship between test scores. For example, the WMS manual (Wechsler, 2009) sets the maximum possible VPA II scaled score at 13 (z-score of 1.0) for a 20-year-old, whereas the CVLT-II Long-Delay Free-Recall trial maximum z-score is 1.5. Psychometrically, variables with restricted ranges of scores have attenuated associations with other variables.

Despite potentially meaningful differences between tasks, the WMS manual (Wechsler, 2009) provides a method by which scores from the CVLT-II can be converted into scaled scores and substituted for VPA scores. Specifically, VPA I scaled score substitutes are derived from the CVLT-II Trials 1–5 Free-Recall *T* score and VPA II scaled score substitutes are derived from the CVLT-II Long-Delay Free-Recall z score. The rationale underlying these substitutions relate to the conceptual similarities between the VPA and CVLT-II in terms of verbal content, response processes, task demands, and semantic association. The manual reports that the WMS-IV Index Scores derived when using the CVLT-II substitution are “very similar” (p. 167) to those obtained using the standard VPA scores.

Only one published study to-date has investigated the degree to which WMS-IV VPA and substituted VPA scores are interchangeable. Miller and colleagues (2012) utilized archival data from a diverse clinical sample and reported that when the CVLT-II is substituted for VPA scores, index scores were significantly lower for Auditory Memory, but not Delayed Memory or Immediate Memory. They also found that substituted VPA scores were significantly lower than VPA scaled scores for the delayed recall condition, but not for the immediate recall condition.

Miller and colleagues (2012) clearly demonstrate discordance between VPA and substituted VPA scores derived from CVLT-II performance. Despite the moderate correlations between tasks, scores derived from VPA and CVLT-II can result in different performance categorization. This is not surprising and has been observed with other neuropsychological measures that evaluate similar constructs. For example, Stallings, Boake, and Sherer (1995) demonstrated that

despite strong correlations between the CVLT (Delis, Kramer, Kaplan, & Ober, 1987) and the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964), a conceptually similar list learning task, different classification rates emerge. CVLT standard scores obtained from head injured patients were significantly lower than RAVLT standard scores, which presents an interpretive challenge in identifying neurocognitive issues. Even more relevant, Pike and colleagues (2013) found that CVLT-II delayed recall performance was more accurate than VPA delayed recall performance at distinguishing healthy older adults from patients with amnesic mild cognitive impairment. However, their study did not explicitly address whether there was a meaningful difference between VPA delayed recall performance and a substituted VPA delayed recall performance derived from the CVLT-II. Clearly, it would be problematic if substituting CVLT-II scores produced inconsistent results with standard WMS-IV administration.

Given the discordance found in recent studies among clinical populations, this study aims to investigate the concordance of VPA and CVLT-II scores, and the degree to which these scores are interchangeable in deriving the WMS-IV AMI score, among a relatively healthy sample of high functioning young adults. Young, healthy adults often participate in research (e.g., see, Booksh, Pella, Singh, & Gouvier, 2010; Sher, Martin, Wood, & Rutledge, 1997; Suhr & Boyer, 1999) and undergo evaluations in academic or vocational contexts (e.g., see, Prevatt, Welles, Li, & Proctor, 2010). It is expected that these healthy individuals will achieve average or above-average WMS-IV and CVLT-II scores which permits a unique investigation of the CVLT-II substitution. Given the differences between the CVLT-II and VPA subtests in terms of ceiling limits, it is possible that the nature and degree of concordance between the tasks might differ in this sample compared with the clinical sample reported by Miller and colleagues (2012). The present study seeks to inform clinicians and researchers of psychometric implications of CVLT-II substitution for VPA in a young cognitively intact sample.

Method

Participants

A total of 103 students were recruited from a Midwestern university. Four participants were excluded due to missing data and six were excluded due to questionably valid performance as evaluated by the Victoria Symptom Validity Test (Slick, Hoop, & Strauss, 1995; scoring <21 on the difficult condition; Grote et al., 2000). Analyses were therefore conducted on data from the remaining 93 participants. Mean age was 19.16 ($SD = 1.10$) and mean self-reported GPA was 3.31 ($SD = 0.40$). General intelligence was estimated to be in the high average range (Wechsler Test of Adult Reading [Wechsler, 2001] mean standard score = 114.88; $SD = 7.82$). The majority of participants were Caucasian (80.1%; 5.4% African American, 3.2% Hispanic, 1.1% Asian, and 9.7% other or not indicated) and female (71.0%). Thirteen participants indicated on a screening questionnaire that they had a history of a learning disorder ($n = 5$), attention-deficit/hyperactivity disorder ($n = 4$), other neurological disorder ($n = 3$), and/or other psychiatric disorder ($n = 3$). Despite this history, these individuals were considered relatively high functioning. These 13 participants did not report a significantly lower GPA (mean GPA = 3.11, $SD = 0.38$) nor obtain significantly lower WTAR scaled scores (mean standard score = 111.08, $SD = 8.42$), and were therefore included in all analyses.

Primary Neuropsychological Measures

Wechsler Memory Scale, Fourth Edition (Wechsler, 2009) Logical Memory I & II

These subtests assess free-recall memory of two short stories presented verbally. The examinee is asked to recall story details immediately and after a 20- to 30-min delay. Test-retest reliability over a mean period of 23 days varied from $r = .67$ (LM II) to $r = .72$ (LM I). A yes/no recognition test for each story is given after the delayed recall trial.

Wechsler Memory Scale, Fourth Edition (Wechsler, 2009) Verbal Paired Associates I & II

These subtests assess memory for associated word pairs. A list of 14 word pairs is read to the examinee. The examinee is then asked to provide the associated word when given the first word of the pair. This task is repeated across four trials and feedback is given regarding performance on each item. After a 20- to 30-min delay, the examinee is asked to recall the paired word without performance feedback. Test-retest reliability for both VPA I and II over a mean period of 23 days was $r = .76$. A yes/no recognition test of word pairs and a free-recall test of words from the word pairs are administered after the delayed memory trial.

California Verbal Learning Test, Second Edition (Delis et al., 2000)

This verbal memory test evaluates recall and recognition of a word list across immediate and delayed trials. The primary word list (List A) consists of 16 words and is presented in five free-recall trials. A second word list (List B) also consists of 16 words and is used as an interference trial. Following this interference trial, short-delay free-recall and cued-recall trials are administered for List A. After a 20-min delay, long-delay free-recall, long-delay cued-recall, yes/no recognition, and forced-choice recognition trials are administered for List A. Test-retest reliability over a mean period of 21 days was high for Trials 1–5 Free-Recall Total ($r = .82$) and Long-Delay Free Recall ($r = .88$).

Procedure

Following institutional review board approval, participants were recruited from an undergraduate psychology research pool and provided course credit in exchange for participation. Data from this study were collected as part of a larger study investigating psychometric properties of various neuropsychological tests. Order of the memory tests was counterbalanced, so that participants were either administered first, CVLT-II, followed by LM and VPA or second, LM and VPA followed by CVLT-II. Consistent with standardized

administration procedures in the WMS-IV manual, the order of LM and VPA was not counterbalanced. During the 20- to 30-min delay of each verbal memory test, a primarily non-verbal test/group of tests was administered for appropriate interpolated activity. The WTAR was administered upon completion of all other tests.

Analyses

Analyses were conducted using Statistical Package for the Social Sciences version 21 for Windows (IBM, 2012). Alpha levels of $p < .05$ were considered significant. Pearson product-moment correlation statistics were calculated between CVLT-II standard scores and VPA immediate and delayed recall scaled scores to determine the relationship between these variables. Fisher's r -to- z transformation was used to compare correlations from this study with those obtained by Miller and colleagues (2012) and Wechsler (2009). This procedure converts the sampling distribution of Pearson's r (not normally distributed) to the normally distributed z variable to enable comparisons between two different samples (Fisher, 1915; Kenny, 1987). For each substitution, paired samples t test were conducted between original VPA scaled scores and substitute scaled scores from the CVLT-II. Cohen's d was used to evaluate the magnitude of mean differences obtained from these t tests. Guidelines by Cohen (1977) indicate that $d = 0.2$ is a small effect size, $d = 0.5$ is a moderate effect size, and $d = 0.8$ is a large effect size. Exploratory *post hoc* analyses were also conducted to determine if test administration order impacted performances on WMS-IV VPA and CVLT-II.

Results

The pattern of correlations between the CVLT-II and WMS-IV VPA subtests was somewhat different from prior investigations. CVLT-II Trials 1–5 Total T score was not significantly associated with VPA I scaled score, $r = .17$, $p = .10$. The magnitude of association is meaningfully less than those obtained by Miller and colleagues (2012; $r = .49$; Fisher's r -to- z transformation $z = 2.97$, $p < .01$) and Wechsler (2009; $r = .54$; Fisher's r -to- z transformation $z = 3.68$, $p < .01$). The correlation between CVLT-II Long-Delay Free-Recall z score and VPA II scaled score was moderate and significant, $r = .33$, $p < .01$. This

correlation is similar to results obtained by Miller and colleagues (2012; $r = .45$; Fisher's r -to- z transformation $z = 1.19$, $p = .12$), but significantly different from correlations obtained by Wechsler (2009; $r = .51$; Fisher's r -to- z transformation $z = 1.91$, $p = .03$).

Mean level performances and differences in AMI scores are, respectively, presented in Table 1. Consistent with expectations given the nature of the sample, a great majority of the participants performed in the average or greater range on LM (percentage of students in the average or above-average range: LM I = 87.10%; LM II = 88.17%), VPA (VPA I = 95.70%; VPA II = 97.85%), and the CVLT-II (Trials 1–5 Free Recall = 89.25%; Long-Delay Free Recall = 83.87%). Paired samples t -test revealed that utilization of CVLT-II Trials 1–5 Free-Recall T scores led to significantly *lower* VPA I substitute scores, $t(92) = 2.99$, $p < .01$, $d = 0.31$. This finding is inconsistent with those obtained by Miller and colleagues (2012), who reported that VPA I substitute scores were lower, but not significantly different from VPA I scores. Utilization of CVLT-II Long-Delay Free-Recall z scores led to significantly *lower* VPA II substitute scores, $t(92) = 3.90$, $p < .01$, $d = 0.40$. Similar delayed recall results were obtained by Miller and colleagues (2012). Utilization of CVLT-II scores as substitutes for VPA I and II scores led to significantly *lower* AMI scores, $t(92) = 3.68$, $p < .01$, $d = 0.38$, which is also consistent with Miller and colleagues's findings. Fig. 1 displays the strong relationship ($r = .77$) between AMI scores derived by summing either first, LM and VPA or second, LM and CVLT-II performances.

Table 1. Mean auditory memory scores

Auditory Memory Test/Index	Mean	SD	Range	Skewness	Kurtosis
VPA I Scaled Score	11.86	2.29	5–17	–0.33	0.26
VPA II Scaled Score	11.33	1.57	3–13	–2.14	8.02
LM I Scaled Score	10.78	2.53	3–16	–0.39	0.18
LM II Scaled Score	10.53	2.56	5–16	–0.04	–0.12
CVLT Trials 1–5 T Score	55.26	9.18	29–78	–0.28	0.01
VPA I Substitute Scaled Score	10.92	2.36	4–17	–0.33	0.26
CVLT Long-Delay Free-Recall z score	0.31	0.92	–2.50–1.50	–0.62	–0.25
VPA II Substitute Scaled Score	10.51	2.00	3–13	–0.71	2.37
AMI: LM and VPA	106.53	10.30	64–130	–0.71	2.37
AMI: LM and CVLT	103.83	10.51	70–123	–0.56	0.70

Notes: VPA = Verbal Paired Associates; I = Immediate Recall; II = Delayed Recall; LM = Logical Memory; CVLT = California Verbal Learning Test-Second Edition; AMI = Auditory Memory Index.

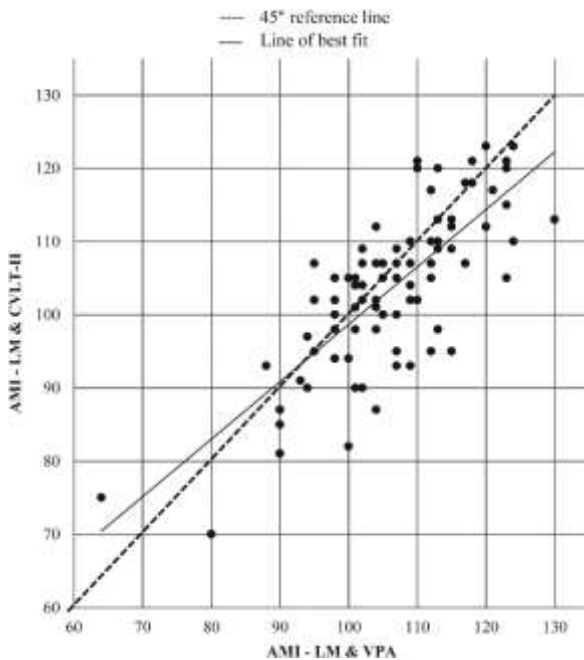


Fig. 1. Distribution of Auditory Memory Index scores derived from Logical Memory and Verbal Paired Associates, and California Verbal Learning Test-Second Edition.

Exploratory *post hoc* analyses were conducted to investigate whether test order might impact VPA and CVLT-II performances (*Post hoc* analyses were based on 79 participants [85% of study sample], for whom test order were recorded.). The order of test administration impacted learning and recalling word lists. When the WMS-IV was administered prior to the CVLT-II, the average CVLT Trials 1–5 Free-Recall *T* score was significantly greater than if the CVLT-II had been administered first, $t(77) = 4.47, p < .01, d = 1.09$. The same pattern emerged for CVLT-II Delayed Free-Recall trial *z* scores, $t(77) = 3.98, p < .01, d = 0.89$. Similar findings were obtained with the VPA substitute scaled scores (VPA I Substitute Scaled Score, $t(77) = 4.76, p < .01, d = 1.07$; VPA II Substitute Scaled Score, $t(77) = 3.34, p < .01, d = 0.75$). However, order of test administration did not significantly impact standard and alternatively generated AMI scores (Standard AMI, $t(77) = -0.81, p = .42, d = -0.18$; Alternative AMI, $t(77) = 1.73, p = .09, d = 0.39$).

Discussion

This study investigated the interchangeability of VPA and CVLT-II scores in deriving the AMI of the WMS-IV. The correlation between CVLT-II Trials 1–5 total *T* score and VPA I scaled score was small, insignificant, and inconsistent with prior investigations (Miller et al., 2012; Wechsler, 2009). The correlation between the CVLT-II Long-Delay Free-Recall *z* score and VPA II scaled score was moderate, significant, and consistent with the correlation obtained by Miller and colleagues (2012), but not Wechsler (2009). In addition, VPA I substitute scores, VPA II substitute scores, and AMI scores derived with CVLT-II scores were significantly lower than corresponding scores derived using VPA scores. Our results pertaining to VPA II substitute scores and AMI scores were consistent with the findings of Miller and colleagues (2012). However, our finding pertaining to VPA I substitute scores represents a unique result that raises further questions regarding the legitimacy of the CVLT-II substitution when deriving WMS-IV scores.

Discrepancy in findings across studies is likely related to unique sample characteristics. The current sample includes relatively high functioning college students, which resulted in a unique distribution of performances. The distribution of scores obtained from Miller and colleague's (2012) clinical sample was more normally distributed (skewness = 0.15 and -0.10 and kurtosis = -0.02 and -0.50 for VPA I and VPA II, respectively) than in the present study. Memory scores were primarily on the higher end of the scales (skewness = -0.33 and -2.14 and kurtosis = 0.26 and 8.02 for VPA I and VPA II, respectively), thus contributing to a more restricted range and weaker associations between tests in the present study. It is debatable whether the current data should have been transformed to more closely approximate the normal distribution prior to conducting analyses (Howell, 2010). This was not done in order to maintain score metrics that are easily interpreted by clinicians and researchers.

To further explore issues pertaining to potentially restricted ranges of VPA scores, we investigated what percentage of participants obtained perfect scores on various trials. Seventy of the 93 participants (75.3%) obtained the maximum raw VPA score of 14 by

the last (i.e., fourth) VPA learning trial, whereas only 27 of the 93 participants (29.0%) obtained the maximum raw CVLT-II score of 16 by the last (i.e., fifth) CVLT-II learning trial. Similarly, 52 participants (55.9%) obtained a perfect VPA II raw score of 14, whereas only 14 participants (15.1%) obtained a perfect CVLT-II Long-Delay free-recall raw score of 16. This pattern mirrors what is observed in the maximum standardized scores possible for VPA and CVLT-II delayed recall performance. Among younger examinees (16–19 years old), VPA II only allows a maximum scaled score of 12 (equivalent z-score of 0.67; 75th percentile) whereas the CVLT-II allows a maximum z-score of 1.5 (equivalent standard score of 14.5; 94th percentile). Collectively, these differences suggest that the upper limit (ceiling) of the VPA subtest is meaningfully lower, and more frequently attained, than that of the CVLT-II. This discrepancy matters psychometrically because the possible range of scores observed is restricted. Moreover, this upper limit compression impacts the clinical utility of the CVLT-II substitution to detect either a decline or an improvement in memory functioning in young adults.

Miller and colleagues (2012) proposed one theoretical explanation for the discrepancy between VPA and CVLT-II performances. Associations formed during the learning trials of VPA are more robust against mnemonic decay compared with the semantic categorization of CVLT-II test items. In addition, VPA test stimuli are presented twice as often as CVLT-II test items. The examiner provides feedback after each response during VPA but not during the CVLT-II. Given that the CVLT-II and VPA subtests are meaningfully different and do not similarly quantify memory functioning, it is not surprising that VPA substitute scores derived from CVLT-II performance do not consistently match VPA scaled scores. Subsequently, differences at the subtest level result in differences at the index level (AMI).

It is important to consider whether observed differences are interpretively meaningful. In other words, are score differences significant enough that a clinician would likely change their interpretation of testing data? It is possible that this may occur as AMI score differences ranged from 0 to 20 (Mean AMI difference = 2.70; $SD = 7.07$), depending on whether VPA or CVLT-II scores were used to derive the AMI score (see Fig. 1). Almost 8% (compared with 6.1% in Miller et al.'s [2012] sample) of participants had AMI scores that were

>15 points (1 *SD*) lower when substituting the CVLT-II for VPA performance. Unlike Miller and colleagues's findings, none of the participants in this study had AMI scores that were >1 *SD* higher when using the CVLT-II to derive the AMI, than when using the VPA (see Table 2). It is also important to recognize that 95th percentile confidence intervals expand as AMI scores become more extreme and regress towards a score of 100. In other words, it is possible that discrepancies further away from a score of 100 are smaller than they visually appear, though it is currently unclear if the same confidence intervals should be applied to alternatively derived AMI scores.

Table 2. Accuracy of substituted WMS-IV subtest scores as a function of *SD*

	AMI^a (%)	VPA I^b (%)	VPA II^b (%)
(Substituted > Original) > 1 <i>SD</i>	0	0	0
(Substituted > Original) ≤ 1 <i>SD</i>	31.18	32.26	18.28
Substituted = Original	10.75	17.20	37.63
(Substituted < Original) ≤ 1 <i>SD</i>	50.54	50.54	44.09
(Substituted < Original) > 1 <i>SD</i>	7.53	0	0

Notes: For example, 31% of AMI scores were between 1 and 15 points higher when generated using CVLT-II rather than VPA.

AMI = Auditory Memory Index; **VPA I** = Verbal Paired Associates, Immediate Recall; **VPA II** = Verbal Paired Associates, Delayed Recall.

^a*SD* of score metric = 15.

^b*SD* of score metric = 3.

As previously described, exploratory *post hoc* analyses revealed that the order of test administration clearly impacted learning and recalling word lists. When the WMS-IV was administered prior to the CVLT-II, CVLT-II performances increased by nearly an *SD*. While noteworthy, the significance of this finding is somewhat unclear because the order effect only resulted in different VPA equivalent scores and did not contribute to a difference between standard and alternatively generated AMI scores. Future research is encouraged to more systematically explore whether this test order effect is uniquely associated with this specific sample of research participants. It is plausible that these bright participants developed effective memory strategies and confidence during the WMS-IV that meaningfully improved their CVLT-II performance. One might hypothesize that an impaired patient would benefit less from exposure to memory tasks than a healthy, young adult. Regardless, clinicians who routinely substitute CVLT-II performances when generating WMS-IV Index scores should recognize the potential meaningful impact of test order.

Unfortunately, it is not clear whether the WMS-IV or CVLT-II was administered first when collecting normative data.

While it is clear that substituting CVLT-II performances for VPA performances results in discrepant scores, it is plausible that each is a valid approximation of verbal memory functioning. Factor analytic research is recommended to further explore whether the CVLT-II and VPA subtests are related to the same theoretical construct, verbal memory functioning. For example, VPA I and VPA II have been included in several factor analytic studies of the WMS-IV (e.g., see Hoelzle et al., 2011; Holdnack, Zhou, Larrabee, Millis, & Salthouse, 2011; Wechsler, 2009). It would be worthwhile to evaluate the congruence of dimensions, loading strength, and amount of variance explained with VPA and VPA equivalent scaled scores. Alternatively, Donders (2008) has identified a multidimensional structure underlying the CVLT-II that consists of Attention Span, Learning Efficiency, Delayed Memory, and Inaccurate Memory. Novel VPA scores could be generated (e.g., Intrusions, Learning Efficiency) and the fit between a similar factor structure (in terms of dimensionality, loading strength, common and unique variances) and VPA performance could be quantified through confirmatory factor analytic methods. Additionally, novel empirical investigations are encouraged to explore whether the standard or alternatively generated AMI score is more predictive of verbal memory functioning, or another relevant outcome variable. The relatively brief assessment battery administered in this research significantly impacts the degree to which additional analyses could be conducted to explore these key issues.

Also due to the limited nature of our assessment battery, we were unable to determine how substitution using the CVLT-II scores affected changes in the WMS-IV Immediate Memory Index (IMI) and Delayed Memory Index (DMI) scores in this sample. One might anticipate that substituting CVLT-II for VPA performances would result in smaller changes for IMI and DMI scores, compared with AMI scores, since the substitution results in a relatively smaller percentage of change (one of four contributing subtest scores is changed in IMI and DMI, whereas two of four contributing subtest scores are changed in AMI). Nevertheless, given these observed differences in scores, we suggest that clinicians exercise caution in deriving the AMI using CVLT-II scores, due to the high likelihood of generating discrepant scores.

If it is necessary to quantify verbal memory functioning, VPA subtests may not be sufficiently challenging for higher functioning young adults. VPA administration, relative to CVLT-II administration, results in a more restricted range of scores, which could ultimately lead to differences in test sensitivity and specificity. Assuming VPA and the CVLT-II evaluate the same construct, our findings suggest that the higher ceiling of the CVLT-II is more sensitive to differences in memory performance among those with relatively strong memory functioning. As an additional advantage, the standard error of measurement (SEM) associated with CVLT-II scores is likely smaller than the SEM associated with VPA scores, given that the CVLT-II has greater test-retest reliability coefficients than the VPA. In other words, there are several meaningful reasons to believe that the CVLT-II would be a more precise instrument to use during research and clinical activities. This belief is consistent with previously documented findings among clinical samples that suggest the CVLT-II is a particularly effective instrument. Specifically, the CVLT-II has been found to be more sensitive than the VPA subtest to memory deficits observed in a sample of patients who have amnesic mild cognitive impairment (Pike et al., 2013). Similarly, the original CVLT was found to be more sensitive than the Hopkins Verbal Learning Test (Brandt, 1991) due to the higher ceiling of the CVLT arising from more items on the word list (Lacritz, Cullum, Weiner, & Rosenberg, 2001).

The quest to more precisely quantify memory functioning continues to challenge neuropsychologists. Future research could also investigate other methods of assessing memory, apart from total correct scores, as is the method used in VPA and CVLT-II. It may be helpful to emphasize and take into consideration learning curves instead of absolute correct or incorrect numbers (Helmstaedter, Wietzke, & Lutz, 2009). In addition, utilizing an item response theory approach, and assigning different scoring weights to individual items based on item difficulty level could increase the precision at which different levels of memory functioning are distinguished (e.g., see, Buschke et al., 2006; Gavett & Horwitz, 2012). Such procedures would likely overcome many difficulties associated with floor and ceiling effects commonly observed on memory tests such as the WMS-IV and CVLT-II. This would ultimately lead to more accurate assessment, which would be a positive development in an era of medicine that

strives for cost-effective and empirically supported assessment and intervention.

References

- Booksh, R. L., Pella, R. D., Singh, A. N., Gouvier, W. D., (2010). Ability of college students to simulate ADHD on objective measures of attention. *Journal of Attention Disorders*, 13(4), 325–338.
- Brandt, J., (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The Clinical Neuropsychologist*, 5, 125–142.
- Breier, J. I., Plenger, P. M., Wheless, J. W., Thomas, A. B., Brookshire, B. L., Curtis, V. L., et al. (1996). Memory tests distinguish between patients with focal temporal and extratemporal lobe epilepsy. *Epilepsia*, 37(2), 165–170.
- Buschke, H., Sliwinski, M. J., Kuslansky, G., Katz, M., Verghese, J., Lipton, R. B., (2006). Retention weighted recall improves discrimination of Alzheimer's disease. *Journal of the International Neuropsychological Society*, 12, 436–440.
- Carlozzi, N. E., Grech, J., Tulskey, D. S., (2013). Memory functioning in individuals with traumatic brain injury: An examination of the Wechsler Memory Scale-Fourth Edition (WMS-IV). *Journal of Clinical and Experimental Neuropsychology*, 35(9), 906–914.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised edition). New York: Academic Press.
- Delis, D. C., Kramer, J. H., Kaplan, E., Ober, B. A., (1987). *The California Verbal Learning Test*. New York: Psychological Corporation.
- Delis, D. C., Kramer, J. H., Kaplan, E., Ober, B. A., (2000). *California Verbal Learning Test-Second Edition*. San Antonio, TX: Psychological Corporation.
- Donders, J., (2008). A confirmatory factor analysis of the California Verbal Learning Test-Second Edition (CVLT-II) in the standardization sample. *Assessment*, 15, 123–131.
- Fisher, R. A., (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507–521.
- Gavett, B. E., Horwitz, J. E., (2012). Immediate list recall as a measure of short-term episodic memory: Insights from the serial position effect and item response theory. *Archives of Clinical Neuropsychology*, 27, 125–135.
- Graham, N. L., Emery, T., Hodges, J. R., (2004). Distinctive cognitive profiles in Alzheimer's disease and subcortical vascular dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(1), 61–71.

- Grote, C. L., Kooker, E. K., Garron, D. C., Nyenhuis, D. L., Smith, C. A., Mattingly, M. L., (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology*, 22(6), 709–719.
- Helmstaedter, C., Elger, C. E., (1996). Cognitive consequences of two-thirds anterior temporal lobectomy on verbal memory in 144 patients: A three-month follow-up study. *Epilepsia*, 37(2), 171–180.
- Helmstaedter, C., Wietzke, J., Lutz, M. T., (2009). Unique and shared validity of the “Wechsler Logical Memory Test,” the “California Verbal Learning Test,” and the “Verbal Learning and Memory Test” in patients with epilepsy. *Epilepsy Research*, 87(2), 203–212.
- Hoelzle, J. B., Nelson, N. W., Smith, C. A., (2011). Comparison of Wechsler Memory Scale-Fourth Edition (WMS-IV) and Third Edition (WMS-III) dimensional structures: Improved ability to evaluate auditory and visual constructs. *Journal of Clinical and Experimental Neuropsychology*, 33(3), 283–291.
- Holdnack, J. A., Zhou, X., Larrabee, G. J., Millis, S. R., Salthouse, T. A., (2011). Confirmatory factor analysis of the WAIS-IV/WMS-IV. *Assessment*, 18(2), 178–191.
- Howell, D. C., (2010). *Statistical methods for psychology*. (7th ed.). Belmont, CA: Wadsworth Cengage.
- IBM Corp. Released 2012. IBM SPSS statistics for Windows, version 21.0. Armonk, NY: Author.
- Kenny, D. A., (1987). *Statistics for the social and behavioral sciences*. Canada: Little Brown and Company Limited.
- Lacritz, L. H., Cullum, C. M., Weiner, M. F., Rosenberg, R. N., (2001). Comparison of the Hopkins Verbal Learning Test-Revised to the California Verbal Learning Test in Alzheimer's disease. *Applied Neuropsychology*, 8(3), 180–184.
- Loring, D. W., Bauer, R. M., (2010). Testing the limits: Cautions and concerns regarding the new Wechsler IQ and memory scales. *Neurology*, 74(8), 685–690.
- Miller, J. B., Axelrod, B. N., Rapport, L. J., Hanks, R. A., Bashem, J. R., Schutte, C., (2012). Substitution of California Verbal Learning Test, for Verbal Paired Associates on the Wechsler Memory Scale. *The Clinical Neuropsychologist*, 26(4), 599–608.
- Pike, K. E., Kinsella, G. J., Ong, B., Mullaly, E., Rand, E., Storey, E., et al. (2013). Is the WMS-IV Verbal Paired Associates as effective as other memory tasks in discriminating amnesic mild cognitive impairment from normal aging? *The Clinical Neuropsychologist*, 27(6), 908–923.

- Prevatt, F., Welles, T., Li, H, Proctor, B., (2010). The contribution of memory and anxiety to the math performance of college students with learning disabilities. *Learning Disabilities Research & Practice*, 25, 39–47.
- Rabin, L. A., Barr, W. B., Burton, L. A., (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, 20, 33–65.
- Rey, A., (1964). *L'examen Clinique en psychologie*. Paris: Presses Universitaires de France.
- Sher, K. J., Martin, E. D., Wood, P. K., Rutledge, P. C., (1997). Alcohol use disorders and neuropsychological functioning in first-year undergraduates. *Experimental and Clinical Psychopharmacology*, 5(3), 304.
- Slick, D. J., Hoop, G., Strauss, E., (1995). *The Victoria Symptom Validity Test*. Odessa, FL: Psychological Assessment Resources.
- Stallings, G., Boake, C., Sherer, M., (1995). Comparison of the California Verbal Learning Test and the Rey Auditory Verbal Learning Test in head-injured patients. *Journal of Clinical and Experimental Neuropsychology*, 17(5), 706–712.
- Suhr, J. A., Boyer, D., (1999). Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. *Journal of Clinical and Experimental Neuropsychology*, 21(5), 701–708.
- Wechsler, D., (1997). *Wechsler Memory Scale-Third Edition (WMS-III) administration and scoring manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D., (2001). *WTAR: Wechsler Test of Adult Reading*. San Antonio, TX: Pearson.
- Wechsler, D., (2009). *Wechsler Memory Scale-Fourth Edition (WMS-IV): Technical and interpretive manual*. Bloomington, MN: Pearson.
- Wilson, B. A., (2002). Assessment of memory disorders. In Baddeley, A. D., Kopelman, M. D., Wilson, B. A., (Eds.), *The handbook of memory disorders* (2nd ed., pp. 617–636). West Sussex: John Wiley & Sons.