Demographically-corrected normative data for the HVLT-R, BVMT-R, and Aggregated Learning Ratio values in a sample of older adults

Dustin B. Hammers[1], **Kevin Duff**[1], **Robert J. Spencer**[2,3]

[1]Center for Alzheimer's Care, Imaging, and Research, Department of Neurology, University of Utah, Salt Lake City, UT, USA

[2]Mental Health Service, VA Ann Arbor Healthcare System, Ann Arbor MI, USA

[3]Michigan Medicine, Department of Psychiatry, Neuropsychology Section, Ann Arbor MI, USA

Abstract

Background: The Learning Ratio (LR) is a novel learning slope score that has been developed to reduce the inherent competition between the first trial and subsequent trials in traditional learning slopes. Recent findings suggest that LR is sensitive to AD pathology along the AD continuum – more so than the traditional learning calculations that employ raw changes across trials. However, research is still experimental and not yet directly applicable to clinical settings. Consequently, the objective of the current study was to develop demographically-corrected normative data on these LR learning slopes.

Method: The current study examined the influence of age and education on LR scores for the HVLT-R, BVMT-R, and an Aggregated HVLT-R/BVMT-R in 200 cognitively intact adults aged 65 years and older using linear regression.

Results: Age negatively correlated with all LR metrics, and education positively correlated with most. No sex differences were identified. LR values were predicted from age and education, which can be compared to observed LR values and converted into demographically-corrected *T* scores.

Conclusions: By comparing observed and predicted LR scores calculated from regression-based prediction equations, interpretations are permitted that aid in clinical decision making and treatment planning. Co-norming of the HVLT-R and BVMT-R also allows for comparisons between verbal and visual learning slope scores in individual patients. We hope normative data for LR enhances its utility as a clinical tool for examining learning slopes in older adults administered the HVLT-R and/or BVMT-R.

Keywords

Learning; Memory; Alzheimer's disease; Mild Cognitive Impairment

---

Corresponding author: Dustin B. Hammers, Ph.D., ABPP(CN), Center for Alzheimer's Care, Imaging & Research, University of Utah, 650 Komas Dr., Suite 106-A, Salt Lake City, UT, 84108, dustin.hammers@hsc.utah.edu, Ph. (801) 585-3929, Fax (801) 581-2483.

INTRODUCTION

The assessment of learning and retention of information over multiple trials is standard in clinical neuropsychological evaluations examining older adults (Lezak et al., 2012; Suhr, 2015). The Hopkins Verbal Learning Test – Revised (HVLT-R; Brandt & Benedict, 1997) and the Brief Visuospatial Memory Test – Revised (BVMT-R; Benedict, 1997) are examples of frequently-used measures of verbal and visual learning/memory, respectively. Each measure generates information regarding total recall and delayed recall capacity of individuals being evaluated. In addition to these variables, the steepness of the learning slope or learning curve can also provide information about the potential for an individual to benefit from repeated exposure to information over multiple trials, which is frequently shallow in individuals with memory impairments, or conditions like Alzheimer's disease (AD; Gifford et al., 2015). As such, there is potential for these data to provide valuable diagnostic and rehabilitative information about an individual's clinical picture. Although learning slope data are readily available in many test manuals, the calculation of such learning slopes has historically considered the raw difference between the final and first learning trial (or between the first and best trial; Bender et al., 2020; Benedict, 1997; Bonner-Jackson et al., 2015; Brandt & Benedict, 1997; Wehling et al., 2007), though some variation to this "Final Trial minus First Trial" calculation has been observed (e.g., regression-based learning slope calculations; Gifford et al., 2015).

Recently, there has been an increase in research on an alternate metric for calculating learning slope that controls for performance on initial learning trials, enabling a more accurate examination of the proportion of information learned over successive trials. Spencer and colleagues (Spencer et al., 2020) developed the Learning Ratio (LR) to better account for initial trial learning when calculating learning slopes. LR is calculated as the number of items learned after Trial One divided by the number of items yet to be learned after Trial One (see detailed equation in the Methods). For example, for a Patient that learns 4 items on Trial One of a 12-item word list and 10 items on the Final Trial, the individual would obtain an LR value of 0.75; specifically, the patient learned 6 items over successive trials out of a potential of 8 additional items after Trial One. By dividing by the number of yet-to-be-learned items, LR represents a proportion of information learned over successive trials relative to information available to be learned. In the case of our example, the patient learned 75% of information left to learn after Trial One. This metric therefore incorporates the opportunity for future learning, which will vary depending on Trial One success. It also controls for the competition between Trial One and subsequent trials, because the more information learned at Trial One results in the less information available to learn in successive trials. This appears to be particularly advantageous because previous research has shown that traditional learning slope scores tend to result idiosyncratic findings. Specifically, learning slope data from both the HVLT-R and BVMT-R manuals suggest that older adults display *better* learning capacity than younger individuals, which is counterintuitive given the known effects of age on learning and memory (Salthouse, 2009, 2010). Conversely, applying the LR equation to data from these manuals results in a predictable trend of steady LR decline with advancing age, of which controlling for Trial One competition has been proposed as the mechanism for these differential findings (Spencer et al., 2020).

Several studies have highlighted the benefit of this LR equation over traditional calculations (e.g., "Final Trial minus First Trial"), and have shown overall criterion and convergent validity for the metric. First, Spencer et al. (2020) validated LR scores from the List Learning and Story Memory subtests of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 2012) in 289 older veterans from an outpatient neuropsychology clinic. They demonstrated that the LR equations – for List Learning, Story Memory, and an Aggregated metric of the two – possessed superior correlations with standard measures of memory than traditional learning slopes, and better discriminated between those with and without a neurocognitive diagnosis (Spencer et al., 2020). Second, Hammers and colleagues (Hammers et al., In Press) subsequently validated this LR learning slope calculation for the HVLT-R and the BVMT-R in an independent sample of 56 memory clinic patients undergoing assessment. Like the RBANS LR scores in Spencer's study, lower HVLT-R, BVMT-R, and Aggregated HVLT-R/BVMT-R LR values correlated with both poorer performances on traditional memory measures (HVLT-R Total Recall and Delayed Recall, and BVMT-R Total Recall and Delayed Recall) and smaller total hippocampal volumes; once again, the magnitude of these relationships was greater for LR than traditional learning slope scores. Further results showed that patients with AD possessed smaller LR scores than those with Mild Cognitive Impairment (MCI; Hammers et al., In Press). Third, Hammers and colleagues (Hammers et al., In Press) have shown in a separate sample of participants across the AD continuum that lower LR scores were observed for the MCI and AD groups than the Normal Cognition group for the HVLT-R, BVMT-R, RBANS List Learning, RBANS Story Memory, and associated Aggregate LR values. LR scores were again positively correlated with performances on traditional learning measures (e.g., smaller LR values related to worse Total/Immediate Recall performances from HVLT-R, BVMT-R, and RBANS List Learning and Story Memory), and LR scores displayed excellent receiver operator characteristics when differentiating those with and without cognitive impairment. Further, Hammers et al. (In Press) have observed that LR values discriminated between those with and without cognitive impairment as well as the respective Total Recall values.

Taken together, these recent findings suggest that LR is sensitive along the AD continuum – more so than the traditional raw learning calculation – and that reducing the competition between the first trial and subsequent trials can better depict learning capacity. Although prior research has shown comparable utility for LR relative to immediate recall scores when predicting impairment, LR appears to permit a more nuanced and accurate understanding of trial-by-trial learning capacity than either a total recall score or the currently used raw learning slope calculations (Hammers et al., In Press), which may permit greater personalization of treatment recommendations for some patients. However, the aforementioned findings are still generally experimental and are not directly applicable in clinical settings. To permit greater use clinically, research is needed to develop normative data on these LR learning slopes in cognitively intact or "clean" individuals (Goodwill et al., 2019; Harrington et al., 2017) for commonly used memory measures. Consequently, the purpose of this study is to develop demographically-corrected normative data in a large sample of older adults with intact cognition. Given the relationships previously observed between both the HVLT-R and the BVMT-R with demographic variables of age, education,

and sex (Duff, 2016), it was expected that these variables would be associated with LR metrics from the HVLT-R and BVMT-R, and would be predictive of these LR scores in an older adult sample. Because these two memory measures tend to be given in tandem during cognitive assessments, the current approach would also permit co-norming within the same sample, to better allow for comparisons between verbal and visual learning slope abilities in patients. As a result of developing a normative sample for this learning metric, we hope to enhance its utility as a clinical tool for the examination of learning slopes in older adults administered the HVLT-R and/or the BVMT-R.

METHODS

Sample and Procedure

Cognitively intact community-dwelling older adults were recruited from the community (e.g., senior centers and independent living facilities) from two different samples for the current study. The first sample was comprised of 148 cognitively intact community-dwelling older adults recruited from 2008 to 2013 as a control group for a study of practice effects and MCI (see Duff et al., 2017). The second sample was comprised of 52 cognitively intact community-dwelling older adults recruited as a control group for a study of practice effects and AD biomarkers (2019 to present). The first sample's mean age was 75.6 (SD = 7.1, range = 65 – 96) years old, and the second sample was 72.5 (SD = 4.9, range = 65 – 91) years old. The first sample averaged 15.4 (SD = 2.6, range = 8 – 20+) years of education, and the second sample averaged 16.7 (SD = 2.1, range = 12 – 20) years of education. Both samples were predominantly Caucasian, with the first sample being predominantly female (83.1% female) and the second sample having a slightly higher proportion of females than males (61.5% female). Premorbid intellect at baseline was average to high average according to the Wide Range Achievement Test – Third and Fourth editions (WRAT; Wilkinson, 1993; Wilkinson & Robertson, 2006) Reading subtest for both samples (standard score: M = 107.4, SD = 7.2, range = 81 – 126 for the first sample using the WRAT-3, and M = 110.6, SD = 7.4, range = 88 – 126 for the second sample using the WRAT-4). Self-reported depression was generally low for both samples, including an average of 4.0 (SD = 3.6, range = 0 – 14) according to the 30-item Geriatric Depression Scale (GDS; Yesavage et al., 1982) for the first sample, and an average of 0.9 (SD = 1.0, range = 0 – 5) for the second sample using the 15-item GDS (Sheikh & Yesavage, 1986). Of note, self-reported depression was part of the exclusion criteria for the parent study of the second sample, therefore scores were not observed of GDS 5 in this sample.

For inclusion in the study, all participants from both samples were classified as being cognitively intact, or free of cognitive impairment (e.g., MCI or dementia due to AD). Classification of participants from the first sample has been described previously (Duff et al., 2017). Briefly, all participants in this sample performed within 1.5 SD of the mean for each domain of a baseline cognitive evaluation described below. Classification of participants from the second sample was based on the classification battery developed in the Alzheimer's Disease Neuroimaging Initiative (ADNI2, 2020), which included the Mini Mental State Examination (Folstein et al., 1975), the Clinical Dementia Rating Scale

(Morris, 1993), and the Wechsler Memory Scale-Revised (Wechsler, 1987) Logical Memory II Paragraph A.

The two cognitively intact samples differed on age, $t(130.19) = 3.49$, $p = .001$, $d = 0.56$, education, $t(111.11) = -3.09$, $p = .001$, $d = -0.50$, premorbid intellect, $t(198) = -2.76$, $p = .008$, $d = -0.44$, and sex, $\chi^2 (1) = 9.07$, $p = .003$, $Phi = -0.23$. No differences between samples existed for ethnic distribution, $\chi^2 (1) = 0.01$, $p = .99$, $Phi = -0.02$. Additionally, differences existed between samples for BVMT-R Total Recall, $t(198) = -7.09$, $p < .001$, $d = -0.81$, BVMT-R Delayed Recall, $t(198) = -4.44$, $p < .001$, $d = 0.51$, and BVMT-R LR values, $t(198) = -3.03$, $p = .003$, $d = 0.35$, but not for HVLT-R Total Recall, $t(198) = 1.18$, $p = .24$, $d = 0.14$, HVLT-R Delayed Recall, $t(198) = -0.14$, $p = .89$, $d = 0.02$, or HVLT-R LR values, $t(198) = 0.48$, $p = .63$, $d = 0.06$. Of note, both samples remained within normal limits on average for BVMT-R Total Recall (T of 46 versus 56) and Delayed Recall (T of 50 versus 56). The first sample's BVMT-R LR value was 0.56 and the second sample's LR value was 0.67. Overall, any differences in the samples were generally smaller in magnitude and reflected variation within the distribution of intact individuals, therefore these groups were pooled together to create a cognitively intact combined normative sample with a total sample size of 200 participants. Please see Table 1 for the demographic values for the combined normative sample, which displayed average abilities for immediate and delayed memory skills, visuospatial skills, language, attention, and executive functioning.

General inclusion criteria for the study involved being aged 65 years or older and functionally independent (according to participant and/or knowledgeable informant), along with possessing adequate vision, hearing, and motor abilities to complete the cognitive evaluation. General exclusion criteria included neurological conditions likely to affect cognition, dementia, major psychiatric condition, current severe depression, substance abuse, anti-convulsant or anti-psychotic medications, or residence in a skilled nursing or living facility.

Procedure—All procedures were approved by the local Institutional Review Board before the study commenced. All participants provided informed consent before completing any procedures. The following primary measures were administered:

- HVLT-R (Brandt & Benedict, 1997) is a verbal memory task with 12 words learned over three trials, with the correct words summed for the Total Recall score (range = 0 – 36). The Delayed Recall score is the number of correct words recalled after a 20 – 25-minute delay (range = 0 – 12).

- BVMT-R (Benedict, 1997) is a visual memory task with 6 geometric designs in 6 locations on a card learned over three trials, with correct designs and locations summed for the Total Recall score (range = 0 – 36). The Delayed Recall score is the number of correct designs and locations recalled after a 20 – 25-minute delay (range = 0 – 12).

Of note, for HVLT-R and BVMT-R scores, age-corrected normative comparisons generated T score values ($M = 50$, $SD = 10$) for Total Recall and Delayed Recall derived from norms associated with the respective publisher's test manuals. Learning slope performances were

evaluated by raw data from individual trials of each of the memory measures. For both raw scores and *T Scores*, higher values indicate better performance.

- WRAT Reading subtest – Third and Fourth editions (Wilkinson, 1993; Wilkinson & Robertson, 2006) are used as estimates of premorbid intellect for the first and second samples, respectively. During this task an individual attempts to pronounce irregular words, and the raw score is normalized to standard scores ($M = 100$, $SD = 15$) relative to age-matched peers. Higher values indicate better performance.

- The 30-item GDS (Yesavage et al., 1982) was used to assess self-reported depression for the first sample, and the 15-item GDS (Sheikh & Yesavage, 1986) was used to assess self-reported depression for the second sample. Higher scores indicated more self-reported depression for both measures, with the second sample using a cut-off of 5/15 (or higher) as exclusion for the parent study.

- Additional measures of cognition were administered in the parent studies as follows and are included in the tables to further cognitively describe the sample. The Symbol Digit Modalities Test (Smith, 1973) is a divided attention and psychomotor speed task, with the number of correct responses in 90 seconds being the total score (range = 0 - 110). Trail Making Test, Parts A and B (Reitan, 1992) are tests of visual scanning/processing speed and set shifting/complex mental flexibility, respectively. For each part, the score is the time to complete the task (range = 0 – 180 seconds for Part A, and range = 0 – 300 seconds for Part B). For the Digit Symbol Modality Test and the Trail Making Test, Parts A and B, age- and education-corrected normative comparisons generated *T Scores* ($M = 50$, $SD = 10$) derived from Smith et al. (1973) and Ivnik et al. (1996), respectively. Higher scores indicate better cognition. Finally, the RBANS (Randolph, 2012) is a neuropsychological test battery comprising 12 subtests that are used to calculate Index scores for domains of immediate memory, visuospatial/constructional, attention, language, delayed memory, and global neuropsychological functioning. The index scores utilize age-corrected normative comparisons from the test manual to generate standard scores ($M = 100$, $SD = 15$). Higher scores indicate better cognition.

Calculation of Learning Slopes

For the HVLT-R and BVMT-R, the LR score were computed as a proportion where differences in performance between the Final Trial and Trial One is in the numerator, and the difference between the total points available for a trial and performance on Trial One serves as the denominator. More specifically, for both the HVLT-R and the BVMT-R, the total points available for a trial was 12. The manual for the BVMT-R specifies that learning is calculated as the difference between the first trial and the better of the remaining two trials, but for the purposes of this study, and for consistency with both the HVLT-R and with previous research (Hammers et al., In Press; Spencer et al., 2020), we will use the differences between the first and last trials for each test. The aggregated LR score was computed as the combined difference between Trial One and the Final Trial for both tests, divided by the difference between the combined total points available for a trial for both

tests and the sum of Trial One from both tests. The formulas for LR for the HVLT-R, BVMT-R, and the Aggregated HVLT-R/BVMT-R are as follows:

$$\text{LR for each test} = \frac{(\text{Final Trial} - \text{Trial One})}{(\text{Total Points Available For a Trial} - \text{Trial One})}$$

$$\text{Aggregated LR} = \frac{(\text{Measure 1:Final Trial} - \text{Trial One}) + (\text{Measure 2:Final Trial} - \text{Trial One})}{(\text{Combined Total Points Available for a Trial from Both Tests} - \text{Sum of Trial One from Both Tests})}$$

Data Analysis

For the sample demographics, independent samples t tests for the continuous demographic variables (e.g., age, education, and premorbid intellect) and chi square analyses for the dichotomous demographic variables (e.g., sex and ethnicity) were calculated to assess the appropriateness of combining the two samples into a larger normative sample. For the pooled normative sample, bivariate correlation coefficients were then calculated between the various LR values and the demographic variables of age and education, to better understand their influence on the LR metrics. Independent samples *t* tests were additionally calculated for the categorical demographic variable of sex for HVLT-R LR, BVMT-R LR, and the Aggregated HVLT-R/BVMT-R LR in the pooled sample.

To generate demographically-corrected normative data, linear regression analyses were conducted for the HVLT-R LR, BVMT-R LR, and Aggregated HVLT-R/BVMT-R LR scores (Cherner et al., 2007; Duff, 2016; Norman et al., 2011). Specifically, the individual LR raw scores were the criterion variable, and demographic variables of age and education were the predictor variables. Sex was not included in the model because descriptive analyses did not show an association between sex and LR performance.

Measures of effect size were expressed throughout as Cohen's *d* values for continuous data, and *Phi* coefficients for categorical data. Given the number of comparisons in the current study, a two-tailed alpha level was set at .01 for all statistical analyses.

RESULTS

Demographics and Memory Testing

Table 1 reflects demographic characteristics of the normative sample, along with a characterization of the sample's performance among a variety of neuropsychological measures. Table 2 displays the sample's mean and *SD* values for the HVLT-R, BVMT-R, and the Aggregated HVLT-R/BVMT-R LR metrics, along with individual trial performances for the HVLT-R and BVMT-R. The mean LR value for the 200 participants in the current study was 0.69 (*SD* = 0.3; range −0.20 - 1.00) for HVLT-R, 0.59 (*SD* = 0.2; range 0.00 - 1.00) for BVMT-R, and 0.62 (*SD* = 0.2; range 0.07 - 1.00) for the Aggregated HVLT-R/ BVMT-R LR. This equates to the sample, on average, learning 59% to 69% of the available information after Trial One for these memory measures. Additionally, Table 2 indicates that bivariate correlation coefficients for all three LR metrics were significant with age (*p*s

< .003), and coefficients for BVMT-R LR and Aggregated HVLT-R/BVMT-R LR metrics were significant for education (ps < .004), but not for HVLT-R LR (p = .06). Conversely, sex differences were not observed for HVLT-R LR, $t(198) = -1.27$, $p = .21$, $d = 0.18$, BVMT-R LR, $t(198) = 0.61$, $p = .51$, $d = 0.09$, or the Aggregated HVLT-R/BVMT-R LR, $t(198) = -0.43$, $p = .67$, $d = 0.06$. Similarly, individual trials of HVLT-R and BVMT-R performance were also significantly correlated with age (ps < .002) and mostly with education (ps < .01 for four of six comparisons), and sex differences were not observed for any of the HVLT-R or BVMT-R individual trials (ps = .05 to .77). As a result, demographic variables of age and education were included in the subsequent linear regression analyses.

Linear Regression Analyses

Table 3 displays the results of the linear regression analyses for the HVLT-R LR, BVMT-R LR, and Aggregated HVLT-R/BVMT-R LR as the criterion variables, and age and education as the predictor variables. Briefly, the model containing both demographic variables significantly predicted all three LR metrics (ps < .004).

DISCUSSION

The current study sought to advance the literature on learning assessment in older adults by developing demographically-corrected normative data for the Learning Ratio (LR), which is a novel method of assessing learning slope that controls for initial trial learning. Our results suggest that LR performance was consistently associated with demographic variables of age and education – but not sex – in our sample of cognitively intact older adults (see Table 2). This is mostly consistent with our expectations. Specifically, bivariate correlations indicated that each of the LR calculations – HVLT-R LR, BVMT-R LR, and Aggregated HVLT-R/ BVMT-R LR – were negatively correlated with age, with increased age being associated with worse LR performance. Similarly, education was positively correlated with two of the three LR metrics, such that greater levels of education were associated with greater LR performance. These LR results similarly correspond to the relationships between demographic variables and the individual learning trials for the HVLT-R and the BVMT-R in the current study. They are also consistent with prior research suggesting associations between age/education and performance on immediate/delayed total scores and learning trials for the HVLT-R and BVMT-R (Cherner et al., 2007; Duff, 2016; Kane & Yochim, 2014; Norman et al., 2011; Vanderploeg et al., 2000), and argue for the appropriateness of including demographic corrections in the normative data for the LR metrics for these measures. Regarding the lack of sex differences in our LR equations, a review of the literature suggests that research into sex differences in HVLT-R and BVMT-R performance has been somewhat mixed. Although several studies found sex differences using these measures (Brunet et al., 2020; Munro et al., 2012), several others have not found significant differences (Duff, 2016; Gale et al., 2007; Hester et al., 2004; Kane & Yochim, 2014). Similarly, when examining the manuals for these measures, the BVMT-R manual indicates that sex did not contribute to performance in their normative samples, and the HVLT-R manual indicated that the contribution was so minor that the developers did not include sex in their norms either.

Table 3 shows regression-based prediction equations developed for the HVLT-R LR, BVMT-R LR, and Aggregated HVLT-R/BVMT-R LR. For each LR metric, predicted LR scores could be generated from models containing the demographic variables of age and education. When examining the R^2 values closely, it can be observed that each model accounted for a relatively small proportion of variance (6%, 11%, and 11% for HVLT-R LR, BVMT-R LR, and HVLT-R/BVMT-R, respectively). This is not necessarily surprising given the small bivariate correlations with LR for age and education in Table 2, and the lower magnitude of the beta weights for age and education in the prediction equations in Table 3. These results were similarly unsurprising given that Duff (2016) has previously shown that age accounted for 10-11% of the variance with HVLT-R and BVMT-R scores, and education tended to share only 2% of the variance. Despite the low levels of variance accounted for, inclusion of these easily-obtainable variables permits greater accuracy of the predicted LR performances, subsequently leading to more accurate and patient-specific normative comparisons.

Table 4 provides an example of how to apply these prediction equations to LR performance for an individual, though it is of note that the interested reader can contact the first author to obtain an Excel spreadsheet that will automatically calculate these demographically-corrected values. The excel spreadsheet can additionally be found here: (Hammers, D. (2021, April 3). Hammers HVLT-R LR and BVMT-R LR Normative Calculator. Retrieved from https://osf.io/8ugzv/. In essence, these equations are used to generate predicted LR performances for each measure, which can then be compared to the observed LR performances to assess how much an individual deviates from his/her same-age and -education-matched peers. More specifically, following the calculation of the observed LR value for the individual, the discrepancy between the observed LR value and the predicted LR value is determined (observed LR - predicted LR/Standard Error of the Estimate [SE_{est}]). The resulting calculation yields an age- and education-corrected LR Discrepancy z-score value. For the example of a 65-year-old woman with 12 years of education who obtained an observed HVLT-R LR value of 0.86 (recalling 86% of available information after Trial One), the difference between the predicted and observed LR value was 0.11, and when divided by the 0.28 (SE_{est} from Tables 3 or 4) led to a z value of 0.39. This LR Discrepancy z-score value can then be translated into an age- and education-corrected T score (mean of 50, SD of 10; multiplying by 10, adding 50). As a result, an HVLT-R LR score of 0.86 is equivalent to a T score of 54 for a 65-year-old woman with 12 years of education, which is consistent with a learning slope performance at the upper limit of the average range.

For the reader seeking to calculate normative values for LR scores using Observed LR performance, participant age, and participant education in a single step, these equations are displayed in Table 5.

Conversely, suppose our 65-year-old woman with 12 years of education performed worse than in our previous example. As can be seen in Table 6, when obtaining scores of 7, 7, and 8 on trials 1-3 of the HVLT-R, respectively, this results in an HVLT-R LR value of 0.20 (recalling 20% of available information after Trial One). The difference between her predicted and observed LR value was −0.55, and when divided by the SE_{est}, this led to a z value of −1.97. After converting this LR Discrepancy z-score value to a T score (multiplying

by 10, adding 50), it can be observed that her HVLT-R LR score of 0.20 is equivalent to a T score of 30. This is consistent with a learning slope performance in the impaired range.

While the creation of look-up tables for each age and education group is beyond the scope of this manuscript (and would result in dozens of tables), Table 7 reflects the performance distribution on the LR metrics for a 75-year-old with 16 years of education. As can be observed, for this set of demographic characteristics the mean ($T = 50$) raw LR values would be around 0.70 for HVLT-R, 0.55 for BVMT-R, and 0.60 for Aggregated HVMT-R/BVMT-R, reflecting that the average 75-year-old with 16 years of education learned 70%, 55%, and 60% of available information after Trial One on subsequent trials, respectively. This corresponds to the LR means in Table 2. Weaker performances ($T < 44$) on the LR metrics were observed below LR values of 0.41 to 0.54 (41% to 54% of available information learned), and impaired performances ($T < 31$) were below 0.14 and 0.19 (14% and 19% learned) for the BVMT-R and HVLT-R LR metrics, respectively, and below 0.27 (27% learned) for the Aggregated LR metric. When examining Table 7 more closely, it can be seen that the regression equations tend to result in a ceiling effect for LR performance. In particular for the HVLT-R LR, it is possible to perform very poorly using these norms (learning 0% results in a T score of 24), but performance cannot exceed the high average range (learning score of 100% results in a T score of 61). As such, the metric appears to be slightly more sensitive at identifying individuals with learning problems than those with exceptional learning capacities.

Additional consideration of the two sets of performance data for our 65-year-old case example highlights further clinical benefit of the use of this LR metric. Namely, the existence of normative data for learning slope now permits a more nuanced understanding of learning for some patients. In our first example of a 65-year-old woman with 12 years of education (Table 4), her HVLT-R Total Recall performance was $T = 46$, and her BVMT-R Total Recall performance was $T = 45$ according to the test manuals. This generally equates to learning abilities at the lower limit of the average range. When examining her individual trial performances on the HVLT-R and BVMT-R, however, it can be observed that she tends to possess a rather steep learning curve, such that her initial trial learning is limited but she steadily improves with successive exposures to the material. Her observed LR values of 0.86 and 0.78 suggest that she learned 78% to 86% of available information after Trial One on subsequent trials of the HVLT-R and BVMT-R, respectively, which after applying the demographically-corrected normative comparisons equates to T scores of 54 and 59, respectively. These values are generally consistent with the upper limit of the average and high average ranges, respectively, and suggest a stronger learning capacity for this individual than her Total Recall T scores would imply. Such data reflect a clinical picture of an individual with weaker learning upon initial exposure, but a strong capacity to benefit from repeated exposure. In contrast, in our second example of a 65-year-old woman with 12 years of education (Table 6), her HVLT-R Total Recall performance was $T = 40$ according to the test manual, equating to learning abilities in the low average range. When again examining her individual trial performances on the HVLT-R, however, it can be observed that she tends to possess a rather flat learning curve, such that her initial trial learning is modest but she displayed limited improvement with successive exposures to the material. Her observed LR value of 0.20 suggested that she learned 20% of available information after Trial One on

subsequent trials of the HVLT-R, which after applying the demographically-corrected normative comparisons equates to *T* score of 30. This suggests an impaired learning slope, which is weaker than her Total Recall *T* score would imply. This data reflect a clinical picture of an individual with adequate learning upon initial exposure, but poor capacity to benefit from repeated exposure. Taken together, these norms provide a more detailed clinical picture of learning performance, and can be particularly helpful to aid in treatment recommendations for individual patients.

The current study is not without limitations. First, it is unclear if the current results would be similarly observed in a study incorporating more heterogeneous participants in regards to premorbid functioning, education, ethnicity, or sex. In particular, few participants in the current sample were non-Caucasian, and the sample was predominantly female. Consequently, it is unknown how these normative comparisons perform in populations of other ethnicities, and it is possible that a more evenly distributed sample regarding gender would have led to different results, though it is comforting that no sex-based differences were observed on any LR metrics in the study. While future research should consider normative comparisons in samples that are not primarily well-educated Caucasian females, the current study's proportion of highly educated females appear to reflect long-standing trends in research participation. Specifically, it has been observed that women tend to volunteer more than men across all age ranges (United States Bureau of the Census, 2015), reaching a difference of upwards of 30% (U.S. Bureau of Labor Statistics, 2016), and that individuals with higher education and Caucasians consistently volunteer at greater levels (United States Bureau of the Census, 2015). Second, all individuals in our sample were over the age of 65, therefore it is not recommended that these demographically-corrected normative comparisons be applied to younger patients. Third, these results are specific to the LR metric derived from the HVLT-R and BVMT-R, using the equations from Spencer et al. (2020). Future investigation is encouraged to consider normative data for LR metrics from other memory measures, such as the learning subtests from the RBANS or common verbal memory tasks like the California Verbal Learning Test - II (Delis et al., 2000) or the Rey Auditory Learning Test (Schmidt, 1996). Fourth, it is possible that the use of these normative data may be less advantageous for patients performing very strongly on Trial One of either the HVLT-R or BVMT-R. For such individuals, strong Trial One performance (e.g., learning 10 of out 12 items on Trial One) means that there are few opportunities to learn additional information on subsequent trials (i.e., only 0 – 2 items). Spencer et al. (2020) has previously used the phrase "the law of small numbers" to indicate that this can result in small and potentially unstable denominators in the LR equation. The Aggregated HVLT-R/ BVMT-R LR score was created to partially mitigate this ceiling effect issue, consequently it is recommended that the Aggregated LR measure be used instead of individual HVLT-R LR and BVMT-R LR values for strongly performing participants. Finally, regression equations suggested that age and education accounted for a modest amount of variance in LR performance in the current study. While these demographic variables were selected based on (1) significant relationships with the criterion variables and (2) ease of accessibility, future research should consider other demographic information to potentially improve prediction accuracy of these norms.

Despite these limitations, the current study is the first to calculate demographically-corrected normative comparisons for the LR learning slope metric for the HVLT-R and BVMT-R. As a result, an individual's performance on this metric can now be interpreted to aid in clinical decision making and treatment planning. Co-norming of the HVLT-R and BVMT-R additionally allows for comparisons between verbal and visual learning slope abilities in individual patients, which is particularly relevant given that these two memory measures are frequently administered in tandem during cognitive assessments.

Funding:

References

ADNI2. (2020). Alzheimer's Disease Neuroimaging Initiative: ADNI2 Procedures Manual. Retrieved May 21 from https://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf

Bender AR, Brandmaier AM, Duzel S, Keresztes A, Pasternak O, Lindenberger U, & Kuhn S (2020, 4 14). Hippocampal Subfields and Limbic White Matter Jointly Predict Learning Rate in Older Adults. Cereb Cortex, 30(4), 2465–2477. 10.1093/cercor/bhz252 [PubMed: 31800016]

Benedict R (1997). Brief Visuospatial Memory Test-Revised. Psychological Assessment Resources, Inc.

Bonner-Jackson A, Mahmoud S, Miller J, & Banks SJ (2015, 10 15). Verbal and non-verbal memory and hippocampal volumes in a memory clinic population. Alzheimers Res Ther, 7(1), 61. 10.1186/s13195-015-0147-9 [PubMed: 26468086]

Brandt J, & Benedict R (1997). Hopkins Verbal Learning Test-Revised. Psychological Assessment Resources, Inc.

Brunet HE, Caldwell JZK, Brandt J, & Miller JB (2020). Influence of sex differences in interpreting learning and memory within a clinical sample of older adults. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 27(1), 18–39. 10.1080/13825585.2019.1566433 [PubMed: 30663493]

Cherner M, Suarez P, Lazzaretto D, Fortuny LA, Mindt MR, Dawes S, Marcotte T, Grant I, Heaton R, & group H (2007, 3). Demographically corrected norms for the Brief Visuospatial Memory Test-revised and Hopkins Verbal Learning Test-revised in monolingual Spanish speakers from the U.S.-Mexico border region. Arch Clin Neuropsychol, 22(3), 343–353. 10.1016/j.acn.2007.01.009 [PubMed: 17293078]

Delis D, Kramer J, Kaplan E, & Ober B (2000). California Verbal Learning Test - Second Edition. The Psychological Corporation.

Duff K (2016). Demographically corrected normative data for the Hopkins Verbal Learning Test-Revised and Brief Visuospatial Memory Test-Revised in an elderly sample. Appl Neuropsychol Adult, 23(3), 179–185. 10.1080/23279095.2015.1030019 [PubMed: 26496163]

Duff K, Atkinson TJ, Suhrie KR, Dalley BC, Schaefer SY, & Hammers DB (2017, 5). Short-term practice effects in mild cognitive impairment: Evaluating different methods of change. J Clin Exp Neuropsychol, 39(4), 396–407. 10.1080/13803395.2016.1230596 [PubMed: 27646966]

Folstein MF, Folstein SE, & McHugh PR (1975, 11). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res, 12(3), 189–198. 10.1016/0022-3956(75)90026-6 [PubMed: 1202204]

Gale SD, Baxter L, Connor DJ, Herring A, & Comer J (2007). Sex differences on the Rey Auditory Verbal Learning Test and the Brief Visuospatial Memory Test-Revised in the elderly: normative data in 172 participants. J Clin Exp Neuropsychol, 29(5), 561–567. 10.1080/13803390600864760 [PubMed: 17564921]

Gifford KA, Phillips JS, Samuels LR, Lane EM, Bell SP, Liu D, Hohman TJ, Romano RR 3rd, Fritzsche LR, Lu Z, Jefferson AL, & Alzheimer's Disease Neuroimaging, I. (2015, 7).

Associations between Verbal Learning Slope and Neuroimaging Markers across the Cognitive Aging Spectrum. J Int Neuropsychol Soc, 21(6), 455–467. 10.1017/S1355617715000430 [PubMed: 26219209]

Goodwill AM, Campbell S, Henderson VW, Gorelik A, Dennerstein L, McClung M, & Szoeke C (2019, 5). Robust norms for neuropsychological tests of verbal episodic memory in Australian women. Neuropsychology, 33(4), 581–595. 10.1037/neu0000522 [PubMed: 30829514]

Hammers DB, Gradwohl BD, Kucera A, Abildskov T, Wilde EA, & Spencer RJ (In Press). Preliminary validation of a measure for learning slope for the HVLT-R and BVMT-R in older adults. Cognitive and Behavioral Neurology.

Hammers DB, Suhrie KR, Dixon A, Gradwohl BD, Duff K, & Spencer RJ (In Press). Validation of HVLT-R, BVMT-R, and RBANS learning slope scores along the Alzheimer's continuum. Arch Clin Neuropsychol.

Harrington KD, Lim YY, Ames D, Hassenstab J, Rainey-Smith S, Robertson J, Salvado O, Masters CL, Maruff P, & Group AR (2017, 3 1). Using Robust Normative Data to Investigate the Neuropsychology of Cognitive Aging. Arch Clin Neuropsychol, 32(2), 142–154. 10.1093/arclin/acw106 [PubMed: 27932344]

Hester R, Kinsella GJ, Ong B, & Turner M (2004). Hopkins Verbal Learning Test: Normative data for older Australian adults. Australian Psychologist, 39(3), 251–255.

Ivnik R, Malec J, Smith G, Tangalos E, & Petersen R (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE token, WRAT-R reading, AMNART, STROOP, TMT, and JLO. Clin Neuropsychol, 10(3), 262–278. 10.1080/13854049608406689

Kane KD, & Yochim BP (2014, 11). Construct validity and extended normative data for older adults for the Brief Visuospatial Memory Test, Revised. Am J Alzheimers Dis Other Demen, 29(7), 601–606. 10.1177/1533317514524812 [PubMed: 24604892]

Lezak MD, Howieson DB, Bigler ED, & Tranel D (2012). Neuropsycholgoical Assesssment (5th ed.). Oxford University Press.

Morris JC (1993, 11). The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology, 43(11), 2412–2414. 10.1212/wnl.43.11.2412-a

Munro CA, Winicki JM, Schretlen DJ, Gower EW, Turano KA, Munoz B, Keay L, Bandeen-Roche K, & West SK (2012, 11). Sex differences in cognition in healthy elderly individuals. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 19(6), 759–768. 10.1080/13825585.2012.690366 [PubMed: 22670852]

Norman MA, Moore DJ, Taylor M, Franklin D Jr., Cysique L, Ake C, Lazarretto D, Vaida F, Heaton RK, & Group H (2011, 8). Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. J Clin Exp Neuropsychol, 33(7), 793–804. 10.1080/13803395.2011.559157 [PubMed: 21547817]

Randolph C (2012). Repeatable Battery for the Assessment of Neuropsychological Status. The Psychological Corporation.

Reitan R (1992). Trail Making Test: Manual for administration and scoring. Reitan Neuropsychology Laboratory.

Salthouse TA (2009, 9). Decomposing age correlations on neuropsychological and cognitive variables. J Int Neuropsychol Soc, 15(5), 650–661. http://www.ncbi.nlm.nih.gov/pubmed/19570312 [PubMed: 19570312]

Salthouse TA (2010, 9). Influence of age on practice effects in longitudinal neurocognitive change. Neuropsychology, 24(5), 563–572. 10.1037/a0019026 [PubMed: 20804244]

Schmidt M (1996). The Rey Auditory Verbal Learning Test. Western Psychological Services.

Sheikh JI, & Yesavage J (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontologist, 5, 165–172.

Smith A (1973). Symbol Digit Modalities Test. Western Psychological Services.

Spencer RJ, Gradwohl BD, Williams TF, Kordovski VM, & Hammers DB (2020, 7 11). Developing learning slope scores for the repeatable battery for the assessment of neuropsychological status. Appl Neuropsychol Adult, 1–7. 10.1080/23279095.2020.1791870

Suhr JA (2015). Psychological Assessment: A Problem-Solving Approach. Guilford Press.

U.S. Bureau of Labor Statistics. (2016). Volunteering in the United States, 2015 https://www.bls.gov/news.release/volun.nr0.htm

United States Bureau of the Census. (2015). Current Population Survey, September 2014: Volunteer Supplement Inter-university Consortium for Political and Social Research [distributor]. 10.3886/ICPSR36154.v1

Vanderploeg RD, Schinka JA, Jones T, Small BJ, Graves AB, & Mortimer JA (2000, 8). Elderly norms for the Hopkins Verbal Learning Test-Revised. Clin Neuropsychol, 14(3), 318–324. 10.1076/1385-4046(200008)14:3;1-P;FT318 [PubMed: 11262707]

Wechsler D (1987). Manual for the Weschler Memory Scale - Revised. The Psychological Corporation.

Wehling E, Lundervold AJ, Standnes B, Gjerstad L, & Reinvang I (2007, 10 31). APOE status and its association to learning and memory performance in middle aged and older Norwegians seeking assessment for memory deficits. Behav Brain Funct, 3, 57. 10.1186/1744-9081-3-57 [PubMed: 17974013]

Wilkinson GS (1993). WRAT-3: Wide Range Achievement Test Administration Manual. Wide Range, Inc.

Wilkinson GS, & Robertson GJ (2006). WRAT 4: Wide Range Achievement Test, professional manual. Psychological Assessment Resources, Inc.

Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, & Leirer VO (1982). Development and validation of a geriatric depression screening scale: a preliminary report [Research Support, U.S. Gov't, Non-P.H.S.]. J Psychiatr Res, 17(1), 37–49. http://www.ncbi.nlm.nih.gov/pubmed/7183759 [PubMed: 7183759]

Table 1.

Demographic, cognitive, and behavioral variables for the total sample

Variable	*Mean* (*SD*)	Range
n	200	
Age (years)	74.8 (6.7)	65 - 96
Education (years)	15.8 (2.6)	8 - 20
Sex (% female)	77.5%	
Race (% Caucasian)	99.5%	
WRAT Reading	108.2 (7.4)	81 - 126
RBANS Total Scale	109.2 (12.8)	81 - 146
RBANS Immediate Memory Index	108.1 (14.3)	57 - 152
RBANS Visuospatial/Constructional Index	105.6 (14.2)	64 - 131
RBANS Language Index	104.5 (11.1)	75 - 137
RBANS Attention Index	105.6 (14.6)	72 - 138
RBANS Delayed Memory Index	108.6 (10.3)	75 - 137
HVLT-R Total Recall	56.1 (8.6)	26 - 74
HVLT-R Delayed Recall	53.7 (8.2)	27 - 67
BVMT-R Total Recall	48.5 (10.9)	20 - 75
BVMT-R Delayed Recall	51.9 (9.7)	23 - 75
Trial Making Test, Part A	51.9 (9.6)	20 - 77
Trial Making Test, Part B	50.8 (9.7)	21 - 73
Symbol Digit Modalities Test	52.8 (8.0)	27 - 73

Note: WRAT Reading = Wide Range Achievement Test – Third and Fourth Edition Reading Subtest, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status, HVLT-R = Hopkins Verbal Learning Test – Revised, BVMT-R = Brief Visuospatial Memory Test – Revised. WRAT score and RBANS scores listed as a *Standard Score*, and HVLT-R, BVMT-R, Trail Making Tests, and Digit Symbol Modality Test scores are listed as *T Scores*. All values are *Mean* (*Standard Deviation*) unless listed otherwise.

Table 2.

Learning slope scores for the total sample

Variable	M (SD)	Range	r with age	r with education
n	200			
HVLT-R LR	0.69 (0.3)	–0.20 - 1.00	–.22*	.14
BVMT-R LR	0.59 (0.2)	0.00 - 1.00	–.28*	.22*
Aggregated HVLT-R / BVMT-R LR	0.62 (0.2)	0.07 - 1.00	–.30*	.21*
HVLT-R Trial 1	7.04 (1.8)	2 - 12	–.27*	.22*
HVLT-R Trial 2	9.44 (1.8)	5 - 12	–.33*	.20*
HVLT-R Trial 3	10.3 (1.6)	5 - 12	–.24*	.18*
BVMT-R Trial 1	4.14 (2.3)	0 - 11	–.33*	.16
BVMT-R Trial 2	7.11 (2.4)	1 - 12	–.38*	.17
BVMT-R Trial 3	8.55 (2.3)	1 - 12	–.36*	.23*

Note: HVLT-R = Hopkins Verbal Learning Test – Revised, LR = Learning Ratio, BVMT-R = Brief Visuospatial Memory Test – Revised.

*
p < .01.

Table 3.

Regression equations for demographically corrected scores on the HVLT-R LR, BVMT-R LR, and the Aggregated HVLT-R/BVMT-R LR.

	$F(df)$, p, R^2	Equation	SE_{est}
HVLT-R LR	$F(2, 197) = 5.89$, $p = .003$, $R^2 = .06$	$1.14-(age*0.008) + (education*0.011)$	0.28
BVMT-R LR	$F(2, 197) = 11.88$, $p < .001$, $R^2 = .11$	$0.97-(age*0.009) + (education*0.016)$	0.22
Aggregated HVLT-R / BVMT-R LR	$F(2, 197) = 12.51$, $p < .001$, $R^2 = .11$	$1.03-(age*0.008) + (education*0.012)$	0.19

Note: HVLT-R = Hopkins Verbal Learning Test – Revised, LR = Learning Ratio, BVMT-R = Brief Visuospatial Memory Test – Revised. Age and education are both in years. SE_{est} = Standard Error of the estimate.

Table 4.

Case example of a 65-year-old woman with 12 years of education

	HVLT-R	BVMT-R	Aggregated HVLT-R/ BVMT-R
Trial 1	5	3	8
Trial 2	9	6	15
Trial 3	11	10	21
Total Recall *T* Score	46	45	--
Observed LR Value	0.86	0.78	0.81
Predicted LR Value	0.75	0.58	0.65
Observed - Predicted LR Value	0.11	0.20	0.16
SE_{est}	0.28	0.22	0.19
Age/Education Corrected LR Discrepancy *Z-score* Value	0.39	0.92	0.82
Age/Education Corrected LR *T Score* Value	54	59	58

Note: HVLT-R = Hopkins Verbal Learning Test – Revised, BVMT-R = Brief Visuospatial Memory Test – Revised, LR = Learning Ratio, SE_{est} = Standard Error of the Estimate of the regression equations. Predicted Scaled Score LR Values are derived from the regression formula from Table 3. Age/Education Corrected LR Discrepancy Z-score Value = Observed-Predicted Scaled Score LR Value/ SE_{est}.

Table 5.

Single-step equations to generate demographically corrected scores on the HVLT-R LR, BVMT-R LR, and the Aggregated HVLT-R/BVMT-R LR using Observed LR, participant age, and participant education.

	Equation
HVLT-R LR	$((($Observed LR$ - (1.14 - (0.008*age) + (0.011*$education$))) / 0.28) * 10) + 50$
BVMT-R LR	$((($Observed LR$ - (0.97 - (0.009*age) + (0.016*$education$))) / 0.22) * 10) + 50$
Aggregated HVLT-R/BVMT-R LR	$((($Observed LR$ - (1.03 - (0.008*age) + (0.012*$education$))) / 0.19) * 10) + 50$

Note: HVLT-R = Hopkins Verbal Learning Test – Revised, LR = Learning Ratio, BVMT-R = Brief Visuospatial Memory Test – Revised. Age and education are both in years.

Table 6.

A second case example of a 65-year-old woman with 12 years of education

	HVLT-R	BVMT-R	Aggregated HVLT-R/BVMT-R
Trial 1	7	4	11
Trial 2	7	5	12
Trial 3	8	7	15
Total Recall *T* Score	40	40	--
Observed LR Value	0.20	0.38	0.31
Predicted LR Value	0.75	0.58	0.65
Observed - Predicted LR Value	−0.55	−0.20	−0.34
SE_{est}	0.28	0.22	0.19
Age/Education Corrected LR Discrepancy *Z-score* Value	−1.97	−0.92	−1.81
Age/Education Corrected LR *T Score* Value	30	41	32

Note: HVLT-R = Hopkins Verbal Learning Test – Revised, BVMT-R = Brief Visuospatial Memory Test – Revised, LR = Learning Ratio, SE_{est} = Standard Error of the Estimate of the regression equations. Predicted Scaled Score LR Values are derived from the regression formula from Table 3. Age/Education Corrected LR Discrepancy Z-score Value = Observed-Predicted Scaled Score LR Value/ SE_{est}.

Table 7.

Sample demographically corrected LR value T Score performance distribution for a 75-year-old woman with 16 years of education

HVLT-R LR		BVMT-R LR		Aggregated HVLT-R/BVMT-R LR	
LR Value	*T Score*	LR Value	*T Score*	LR Value	*T Score*
0.00	24	0.00	25	0.00	17
0.10	28	0.10	30	0.10	23
0.20	32	0.20	34	0.20	28
0.30	35	0.30	39	0.30	33
0.40	39	0.40	43	0.40	38
0.50	42	0.50	48	0.50	44
0.60	46	0.60	52	0.60	49
0.70	49	0.70	57	0.70	54
0.80	53	0.80	61	0.80	59
0.90	57	0.90	66	0.90	65
1.00	61	1.00	70	1.00	70

Note: HVLT-R = Hopkins Verbal Learning Test – Revised, LR = Learning Ratio, BVMT-R = Brief Visuospatial Memory Test – Revised. LR Values reflect raw LR scores.