Modeling Learning and Memory Using Verbal Learning Tests: Results From ACTIVE

Alden L. Gross,[1] George W. Rebok,[2] Jason Brandt,[2,3] DougTommet,[1] Michael Marsiske,[4] and Richard N. Jones[1]

[1]Hebrew SeniorLife, Boston, Massachusetts.
[2]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.
[3]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland.
[4]Department of Clinical and Health Psychology, University of Florida, Gainesville.

Objective. To investigate the influence of memory training on initial recall and learning.

Method. The Advanced Cognitive Training for Independent and Vital Elderly study of community-dwelling adults older than age 65 ($n = 1,401$). We decomposed trial-level recall in the Auditory Verbal Learning Test (AVLT) and Hopkins Verbal Learning Test (HVLT) into initial recall and learning across trials using latent growth models.

Results. Trial-level increases in words recalled in the AVLT and HVLT at each follow-up visit followed an approximately logarithmic shape. Over the 5-year study period, memory training was associated with slower decline in Trial 1 AVLT recall (Cohen's $d = 0.35$, $p = .03$) and steep pre- and posttraining acceleration in learning ($d = 1.56$, $p < .001$). Findings were replicated using the HVLT (decline in initial recall, $d = 0.60$, $p = .01$; pre- and posttraining acceleration in learning, $d = 3.10$, $p < .001$). Because of the immediate training boost, the memory-trained group had a higher level of recall than the control group through the end of the 5-year study period despite faster decline in learning.

Discussion. This study contributes to the understanding of the mechanisms by which training benefits memory and expands current knowledge by reporting long-term changes in initial recall and learning, as measured from growth models and by characterization of the impact of memory training on these components. Results reveal that memory training delays the worsening of memory span and boosts learning.

Key Words: AVLT—Growth modeling—HVLT—Memory training—Older adults.

MEMORY decline among older adults is common and affects the ability to function independently in society (Verhaeghen, Geraerts, & Marcoen, 2000). Pathological memory impairment is indicative of dementia (Albert, 2008). Word list–learning tasks that measure episodic memory in clinical settings have a long and venerable history in psychology (Ebbinghaus, 1895/1964; Underwood, 1963). The Rey Auditory Verbal Learning Test (AVLT; Rey, 1964; Schmidt, 2004) and Hopkins Verbal Learning Test (HVLT; Brandt, 1991) are among the most widely used such tests. Their relatively simple administration and multiple trials yield a number of useful performance measures that assess distinct concepts in modern theories of memory and learning (Baddeley, Lewis, Eldridge, & Thomson, 1984; Butters & Cermak, 1980; Schmidt, 2004; Vakil & Blachstein, 1997). The sum of correctly recalled words across trials is commonly used, as is a learning curve or slope. This learning curve might be an arithmetic difference between last and first trials (Brandt & Benedict, 2001), a linear fitting function across all the trials (Delis, Kramer, Kaplan, & Ober, 2000), or a model-estimated curve (Jones et al., 2005; Royall, Palmer, Chiodo, & Polk, 2003).

In addition to a recall sum score and a learning score, initial recall of words on the first trial of a multitrial learning task provides a measure of memory span or attentional control for verbal recall (Delis et al., 2000; Jones et al., 2005; Lezak, Howieson, & Loring, 2004). Although necessary for learning, attention is distinct from memory (Hayden et al., 2011). The famous memory disorder patient Henry Gustav Molaison, known as H. M., demonstrated normal digit span and immediate memory but shallow learning and profound forgetting (Scoville, 1968). This finding, replicated in later studies and upheld in theories of learning, is now widely accepted (Ranganath & Blumenfeld, 2005; Ryan & Cohen, 2004; Shallice & Warrington, 1970; Speer, Jacoby, & Braver, 2003). Previous research has shown that initial recall and learning on verbal learning tasks are uncorrelated (Nettelbeck, Rabbitt, Wilson, & Batt, 1996) and that recall on the first trial is least correlated with other trials or with learning, reflecting its attentional component (Macartney-Filgate & Vriezen, 1988; Magalhães & Hamdan, 2010; Ryan, Geisser, Randall, & Georgemiller, 1986; Vakil & Blachstein, 1997).

Memory training interventions for older adults teach strategies to help encode and retrieve information

(McDaniel, Einstein, & Jacoby, 2008; Rebok, Carlson, & Langbaum, 2007; Verhaeghen, Marcoen, & Goossens, 1992). The motivation behind memory training is that memory is modifiable among older adults (Rebok et al., 2007). Although memory training interventions often include attentional components as part of training (Buschkuehl et al., 2008; Calero & Navarro, 2007; Caprio-Prevette & Fry, 1996; Scogin & Prohaska, 1992), training related to attention or working memory is often a stand-alone intervention target outside of memory training (Li et al., 2008; Mahncke, Bronstone, & Merzenich, 2006). Results from the Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) trial and from other training studies generally show training effects that are specific to cognitive functions that were trained (Rebok & Balcerak, 1989; Zehnder, Martin, Altgassen, & Clare, 2009). Results from the ACTIVE speed-of-processing intervention group, for example, shows no transfer from attention tasks to memory. Significant memory-training effects across multiple studies for measures of immediate recall, but not for attention or short-term memory, were reported in a recent Cochrane review of cognitive training (Martin, Clare, Altgassen, Cameron, & Zehnder, 2011; Zehnder et al., 2009). From that review, improvement attributable to training was not different from improvement in active control groups (Martin et al., 2011; Zehnder et al., 2009). With respect to memory training, previous studies from ACTIVE have demonstrated short-term improvement (effect size: 0.26 *SD*s; Ball et al., 2002) and long-term maintenance (effect size: 0.23 *SD*s; Willis et al., 2006) of memory training for up to five years on memory ability among older adults. These findings are largely consistent with research from other training studies with longitudinal follow-up (Borella, Carretti, Riboldi, & De Beni, 2010; Hastings & West, 2009; Neely & Bäckman, 1993; Stigsdotter & Bäckman, 1989; Willis & Nesselroade, 1990). Other studies using ACTIVE data have explored the heterogeneity of training effects and their underlying mechanisms. One study (Langbaum, Rebok, Bandeen-Roche, & Carlson, 2009) reported heterogeneous classes of responses to memory training, which were defined by elevated performance following training on particular tests, including the HVLT (Brandt & Benedict, 2001) and the AVLT (Rey, 1964). Subsequent work exploring mechanisms underlying memory improvement found significant effects of memory training on longitudinal changes in strategy use objectively measured from the AVLT and HVLT using strategy clustering scores (Gross & Rebok, 2011). This study reported that older adults use more semantic clustering in the HVLT, a test comprised of semantically related words; and more serial and subjective clustering in the AVLT, after memory training. The effects of training on strategies were robust for up to five years after initial training, and changes in strategy use variables predicted changes in memory as well as everyday function (Gross & Rebok, 2011).

Prior studies that have modeled verbal learning–trial recall among older adults report that the learning rate is best approximated by a logarithmic (Royall et al., 2003) or approximately logarithmic (Jones et al., 2005) curve (Poreh, 2005). This study simultaneously estimated memory span (initial recall) and learning across trials using latent growth–modeling methods. The study expands current knowledge by studying longitudinal changes in these factors, by both considering initial recall and growth in HVLT performance and characterizing the long-term impact of memory training on memory span and learning. Inferences about memory training from longitudinal models of trial-specific recall sums are contrasted with results using sum of recall scores. Based on prior research, we hypothesized that trial learning follows an approximately logarithmic shape and that memory training affects learning but not initial recall. It was further hypothesized that growth models using sum of recall as outcomes detect immediate pre- and posttraining gains during recall but not training-related effects on attention or learning.

METHOD

Participants and Procedures

The ACTIVE study is a large, multicenter, longitudinal, and randomized controlled trial of cognitive training among community-dwelling adults age 65 and older. The study's primary purpose was to determine whether three cognitive training interventions improved proximal cognitive outcomes and more distal aspects of everyday function (Jobe et al., 2001). Participants were recruited from six university-based sites across the United States. Eligible participants were living in a noninstitutional setting at entry. Individuals that were excluded were younger than 65, reported disabilities in basic or instrumental activities of daily living, were on chemotherapy, had been diagnosed with cancer in the previous five years, had been diagnosed with Alzheimer's disease, had a stroke in the previous 12 months, scored lower than 23 on the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975), had any substantial sensory (e.g., vision, hearing) impairment that could interfere with training, or had received cognitive training within the past 2 years. Additional features of the study design and recruitment strategy are described in detail elsewhere (Ball et al., 2002; Jobe et al., 2001; Willis et al., 2006).

In all, 2,802 participants were randomized to training in memory, inductive reasoning, speed of cognitive processing, or to a no-contact control group. This study used data from memory-trained ($n = 703$) and no-contact control ($n = 698$) participants assessed through the fifth annual visit. Reasoning and processing speed groups were excluded from this study because we sought to describe the impact of memory training on memory and learning. Each ACTIVE intervention was administered in 10 small-group training sessions, each lasting 60–75 min, offered over a course of 10 weeks. The first of 10 sessions provided didactic training on how memory works and how to maximize benefits of training. The second through fifth sessions involved specific memory strategy instruction. Memory strategies during these sessions

included organization, association, visualization, and the method of loci. These strategies were included in training to encourage adoption in everyday cognitive tasks. Participants were provided time and exercises to practice each of these strategies during the remaining five training sessions, during which no new strategies were taught (Jobe et al., 2001).

Participants were assessed at baseline with a thorough battery of neuropsychological tests and questionnaires and then followed up immediately after training (10 weeks after the baseline visit) and one, two, three, and five years after training ended.

Measures

The ACTIVE study used modifications of the AVLT and HVLT, which are tests of word-list memory. The AVLT uses a list of 15 unrelated words administered for five immediate-recall trials, followed by an interference trial with a different word list and a short-delay recall with the first list (Rey, 1964; Schmidt, 2004). The HVLT uses a 12-word list consisting of three sets of four semantically related words repeated in three immediate-recall trials (Brandt, 1991; Brandt & Benedict, 2001). In a modification to the tests' standard clinical administration in which respondents report the words aloud, participants were asked to write down as many words as they could after each trial. For this study, sums of words correctly recalled on each immediate-recall trial were used to model initial recall and learning. We also compared the effects of memory training from these models of trial-specific growth with estimates from models using sums of words recalled across all trials.

Demographic variables selected a priori for this study included age, sex, years of education, self-rated health status, and ethnicity. Self-rated health status was measured on a scale of 1–5 (Ware & Sherbourne, 1992). In addition to demographic predictors, we also considered several indicators of strategy use in the HVLT and AVLT objectively measured at the immediate posttraining assessment. In the HVLT, which contains semantically related words, semantic clustering measures the extent to which a participant recalls words in semantic groupings and subjective clustering measures the extent to which a participant recalls words in the same order from trial to trial. In the AVLT, both subjective and serial clustering processes, which measures the degree to which participants recall words in the order originally presented, were calculated. Strategy clustering scores were calculated using chance-adjusted, list-based equations provided by Stricker et al. (2002) and are described in more detail in Gross and Rebok (2011).

Alternate but nonequivalent word lists from the AVLT and HVLT were administered at different study visits to reduce retest effects, complicating within-person comparisons across time. As a result, trial-level and overall recall sums were adjusted using an equipercentile equating procedure to account for differences in word-list difficulty (Gross et al., 2012; Kolen & Brennan, 1995). In equipercentile equating, trial sum recall scores at follow-up visits for each trial were equated using percentiles from the distribution of analogous trial sum scores at the baseline visit (e.g., all Trial 1 recall sum scores at follow-up visits were equated to the distribution of baseline Trial 1 recall scores). Because equipercentile equating assumes that underlying memory ability is equivalent at follow-up waves, and because in fact the samples at follow-up differ from baseline because of normal aging, attrition, and training effects, we adapted an equipercentile equating algorithm to preserve these differences over time in ACTIVE. Details of the specific procedure are provided elsewhere (Gross et al., 2012). We first identified a restricted equating sample of participants in which to calculate appropriate test percentiles and derive an equating algorithm. We then applied the equating algorithm to the full study sample. We removed aging, attrition, and training effects from the equating sample by restricting it to control group participants whose ages were observable at all study visits (70–91 years). We estimated analytical weights using a direct standardization procedure for age to ensure the same age distribution at each study wave and adjusted these weights for the probability of dropouts. In this selected sample free from the effects of attrition, aging, and training, we applied equipercentile equating using a published R program (Albano, 2011). We then applied the equating algorithm to the full study sample including memory-trained participants. This procedure ensures that all forms are on the same metric with a comparable mean and standard deviation, thus enabling valid within-person longitudinal comparisons (Kolen & Brennan, 1995). All measures retain sensitivity to aging and training effects.

Analysis Plan

Multiple-group latent growth models (LGMs) were used to model initial recall and learning in verbal learning–test recall between the first and final AVLT or HVLT trials in ACTIVE's memory-trained and control groups (McArdle & Bell, 2000; McArdle & Hamagami, 1996; Muthén, 1997; Muthén & Curran, 1997). The latent variables for initial recall (Trial 1 score) and the learning curve are person-specific growth factors summarizing observed trial-level word recall.

The final model setup is provided graphically in Figure 1 for the AVLT. The model for the HVLT was identical except that there were only three HVLT trials per administration (T1–T3). On the trial level, the sum of words correctly recalled was modeled with a series of growth parameters, which are vectors representing the initial-recall and learning-curve components of memory performance (Bollen & Curran, 2006; Duncan, Duncan, & Stryker, 2006). These trial-level parameters, in turn, served as indicators for longitudinal growth processes, defined by an intercept (initial or baseline level), immediate training effect, and linear trajectory. Longitudinal growth parameters were regressed on covariates.

Parameters of interest were means, variances, and covariances of the initial-recall and learning-curve factors. Factor
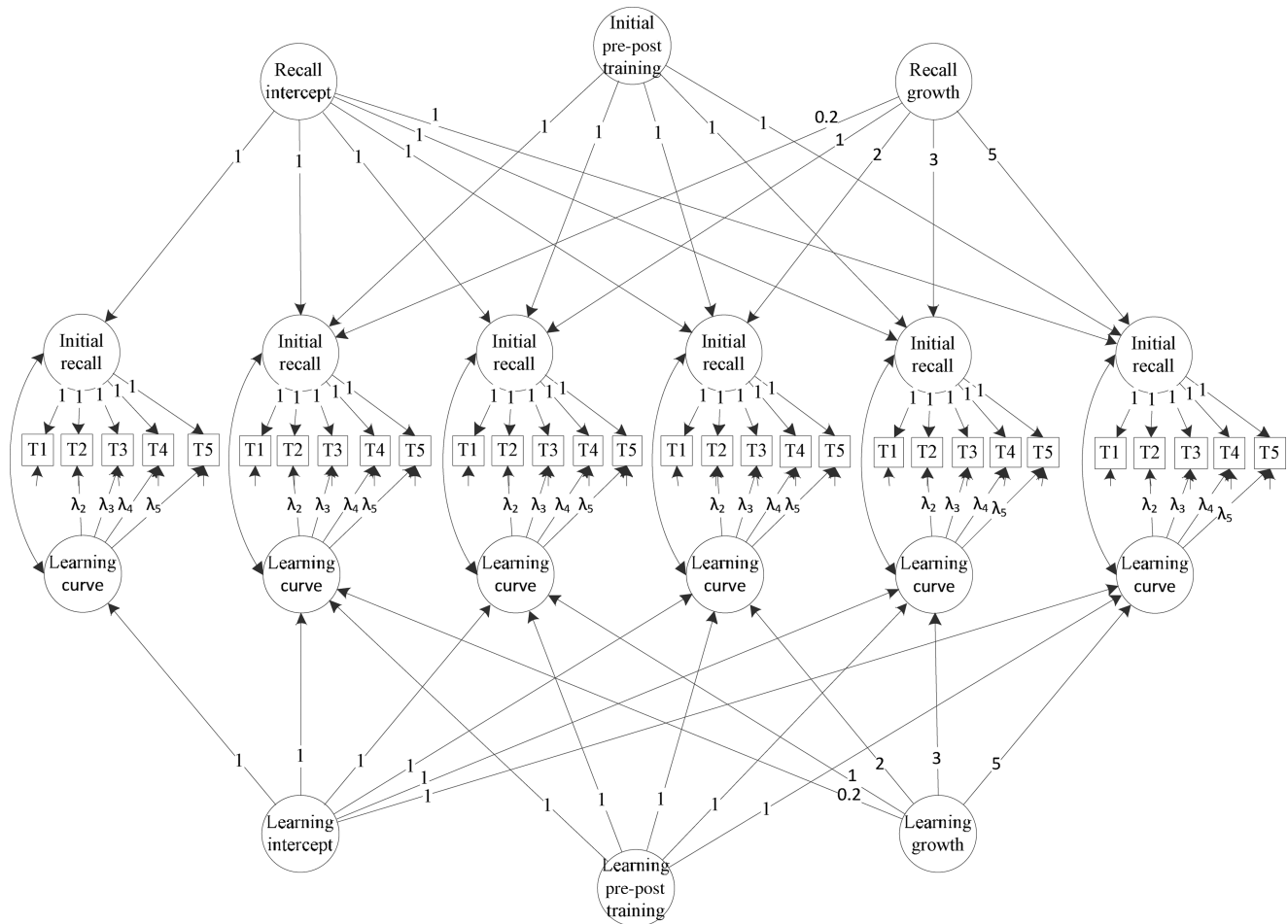
Figure 1. Structural equation model diagram for a second-order latent growth model. Observed trial-level word-recall sums (T1–T5) are shown in squares and latent variables are shown in circles. Initial recall captures recall on the first trial at each study visit. The learning curve captures the number of words recalled between the first and fifth trials, and follows the same approximately logarithmic shape at each study visit. Intercepts for each of these parameters capture baseline values. Pre- and posttraining parameters capture the immediate effects of training and loading at posttraining to fifth-year time points with unit weight. Learning curves capture the annual change in the trial growth parameters, loading with fixed time steps reflecting years from baseline. The structural equation model for the HVLT is identical to this one, except there are only three trials per administration (T1–T3). Latent variable intercepts and slopes capturing change over time were regressed on covariates, which included age, sex, ethnicity, self-rated health, and education. Residual error variances are shown by smaller arrows going toward the observed (boxed) variables. Numbers on arrows going from latent growth parameters to observed time points are factor loadings.

loadings from the learning curve factor to trial scores, representing the proportion of words correctly recalled out of the total on the test, were fixed to values representing different curve shapes, including linear, logarithmic, approximately logarithmic growth, and a freely estimated shape in which all but the first and final trial factor loadings were estimated. The final trial's factor loading was fixed at one in all models so the learning parameter could be interpreted as the proportion of words gained between the first and final trial. In the approximately logarithmic model, the final time step was first freely estimated and then used as the denominator for all other time steps (Jones et al., 2005).

The final longitudinal growth models of AVLT and HVLT trial recall were developed in several steps. First, the curve shape best supported by baseline data was determined by fixing factor-loading paths from the learning-curve factor to trial indicators at values conforming to linear, logarithmic,

approximately logarithmic, and freely estimated curve shapes. We then estimated this final curve shape separately in all follow-up visits in control and memory-trained groups using multiple-group LGM for each study wave. Next, all ACTIVE follow-up visits were entered into a second-order LGM together. The best-fitting curve in the baseline data defined the curve shape for all visits. Longitudinal intercept, pre- and posttraining, and trajectory growth parameters were fit to each growth-curve component (initial recall, learning curve) to characterize longitudinal change in them over time. In the final step of model development, demographic and health-related covariates were added to the model to describe the impact of predictors on level and change in initial recall and learning.

Training effect sizes for parameters were calculated using Cohen's *d* formula as standardized mean differences in parameters between memory-trained and control groups (Cohen, 1988). Mean differences were standardized using

Table 1. Baseline Characteristics and Test Scores of the ACTIVE Sample ($n = 1,401$)

	ACTIVE cohort			
	Memory training ($n = 703$)	Controls ($n = 698$)	Observed range	*p* Value for difference
Age, mean (SD)	73.5 (6.0)	74.1 (6.1)	65, 94	.11
Years of education, mean (SD)	13.6 (2.7)	13.4 (2.7)	4, 20	.14
MMSE score, mean (SD)	27.3 (2.1)	27.3 (2.0)	23, 30	.87
Health status, *n* (%)				.91
Excellent	57 (8.2)	64 (9.3)		
Very good	247 (35.7)	232 (33.7)		
Good	283 (41.0)	287 (41.7)		
Fair	98 (14.2)	100 (14.5)		
Poor	7 (1.0)	6 (0.9)		
Sex, *n* (% female)	537 (76.0)	514 (74.0)		.23
Ethnicity, *n* (% White)	521 (74.0)	500 (72.0)		.30
Strategy clustering scores at immediate posttraining, mean (SD)				
AVLT serial	1.9 (1.1)	1.5 (1.0)	0.1, 3.4	<.001
AVLT subjective	0.9 (0.8)	0.6 (0.8)	−0.5, 2.3	<.001
HVLT semantic	3.5 (1.9)	2.1 (1.7)	−0.5, 5.4	<.001
HVLT subjective	1.3 (1.3)	0.8 (1.1)	−0.8, 3.1	<.001
AVLT scores, mean (SD)				
Baseline				
Trial 1	6.2 (2.0)	6.1 (2.1)	0, 13	.47
Trial 2	9.0 (2.4)	8.8 (2.5)	1, 15	.32
Trial 3	10.6 (2.5)	10.3 (2.6)	2, 15	.03
Trial 4	11.3 (2.5)	11.2 (2.5)	2, 15	.23
Trial 5	11.8 (2.4)	11.7 (2.4)	2, 15	.29
Learning curve (T5–T1)	5.6 (2.1)	5.5 (2.0)	−3, 11	.57
Sum of Trials 1–5	48.8 (10.6)	47.9 (11.0)	8, 73	.11
HVLT, mean (SD)				
Baseline				
Trial 1	7.1 (2.1)	6.9 (2.1)	0, 12	.23
Trial 2	9.2 (2.1)	9.1 (2.1)	0, 12	.37
Trial 3	10.0 (1.9)	9.8 (1.9)	1, 12	.07
Learning curve (T3–T1)	3.0 (1.6)	2.9 (1.5)	−2, 7	.52
Sum of Trials 1–3	26.0 (5.5)	25.7 (5.7)	1, 36	.25

Note. MMSE = Mini-Mental State Exam. Demographic characteristics and baseline AVLT and HVLT word list–learning recall performance are provided for control and memory-trained participants in ACTIVE. Participants in the ACTIVE reasoning and speed of processing training conditions were not used and thus excluded from the table. Learning curves were calculated as the arithmetic difference between the last and first trials.

the estimated standard deviation of each parameter. In the case of the pre- and posttraining parameters for initial recall and learning, which were modeled using a second intercept with zero variance (Ferrer, Salthouse, Stewart, & Schwartz, 2004), the corresponding first intercept's standard deviation was used. This was possible because the variance of the initial intercept reflects model-estimated between-persons variability in initial Trial 1 recall. Because the retest effect has no variance, it is constant for all participants in each intervention group and so the model-estimated between-persons variability at immediate posttraining amounts to the baseline variability mean plus zero.

Analyses were conducted using the MPLUS version 6.1 software package (Muthén & Muthén, 1998–2008). Maximum likelihood estimation procedures that provide robust standard errors and accommodate data missing at random, conditional on observed covariates, were used for parameter estimation (Donders, van der Heijden, Stijnen, & Moons, 2006). By the fifth-year visit, 53% ($n = 749$) of participants were still in the study sample; attrition did not differ by intervention group. There were no significant

covariate-training interactions predicting attrition. Overall model goodness of fit was assessed with standard model fit indices; the root-mean-square error of approximation (RMSEA; Steiger, 1989), comparative fit index (CFI), and the Tucker–Lewis Index (TLI) are reported here (Hu & Bentler, 1999). An RMSEA less than 0.05, CFI greater than 0.90, and TLI greater than 0.90 indicate excellent model fit (Hu & Bentler, 1999). Although the RMSEA is the most popular fit statistic among these, a recent simulation study demonstrated that this statistic fails to detect good-fitting models with small degrees of freedom (Kenny, Kaniskan, & McCoach, under review). As will be seen in the Results section, models of curve shape using baseline data had small degrees of freedom, and so we relied more on CFI and TLI fit statistics than RMSEA to guide model selection.

RESULTS

Descriptive characteristics of the sample are summarized in Table 1. The ACTIVE sample comprises mostly White (73%) women (75%) with a mean age of 73.7 (range 65, 94) at baseline; 40% of participants had at least a high school

Table 2. Latent Growth Model Parameter Estimates for AVLT and HVLT Recall at Baseline: Results from ACTIVE (*n* = 1,401)

	Linear estimate (SE)	Logarithmic estimate (SE)	Approximately logarithmic estimate (SE)	Free time estimate (SE)
AVLT baseline model				
Means				
Initial recall	7.31 (0.12)	6.23 (0.08)	6.22 (0.06)	6.13 (0.08)
Learning curve	4.90 (0.12)	5.66 (0.08)	5.51 (0.05)	5.52 (0.08)
Variances				
Initial recall	3.78 (0.30)	3.74 (0.25)	3.65 (0.17)	3.86 (0.26)
Learning curve	0.59 (0.22)	2.27 (0.24)	2.48 (0.18)	2.51 (0.24)
Proportion of words recalled				
Trial 1	0.00	0.00	0.00	0.00
Trial 2	0.25	0.43	0.47	0.49 (0.01)
Trial 3	0.50	0.68	0.74	0.76 (0.01)
Trial 4	0.75	0.86	0.93	0.91 (0.01)
Trial 5	1.00	1.00	1.00	1.00
Covariance (recall, learning)	0.56 (0.21)	−0.17 (0.19)	−0.30 (0.14)	−0.41 (0.20)
Indicator residual variances	3.30 (0.30)	0.79 (0.15)	0.75 (0.11)	0.51 (0.15)
Model fit statistics				
Degrees of freedom	10	10	10	7
RMSEA	0.359	0.142	0.116	0.101
CFI	0.712	0.955	0.968	0.984
TLI	0.712	0.955	0.968	0.977
HVLT baseline model				
Means				
Initial recall	7.16 (0.08)	7.00 (0.08)	6.93 (0.08)	6.93 (0.08)
Learning curve	2.91 (0.06)	2.96 (0.06)	2.92 (0.06)	2.92 (0.06)
Variances				
Initial recall	3.62 (0.24)	3.86 (0.24)	3.90 (0.24)	3.90 (0.24)
Learning curve	0.23 (0.14)	0.81 (0.14)	0.90 (0.13)	0.90 (0.13)
Proportion of words recalled				
Trial 1	0.00	0.00	0.00	0.00
Trial 2	0.50	0.63	0.73	0.73 (0.01)
Trial 3	1.00	1.00	1.00	1.00
Covariance (recall, learning)	−0.44 (0.14)	−0.71 (0.14)	−0.73 (0.14)	−0.73 (0.14)
Indicator residual variances	1.07 (0.04)	0.80 (0.03)	0.73 (0.03)	0.73 (0.03)
Model fit statistics				
Degrees of freedom	3	3	3	2
RMSEA	0.321	0.169	0.086	0.094
CFI	0.839	0.955	0.989	0.988
TLI	0.862	0.962	0.990	0.988

Note. AVLT = Auditory Verbal Learning Test; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis Index; HVLT = Hopkins Verbal Learning Test. Results from a series of latent growth models of AVLT and HVLT trial recall at the baseline ACTIVE study visit for control group participants. Each column of coefficients comes from a different model that tested different trajectory shapes for the learning curve.

education. Demographic and health characteristics did not differ by training group. Mean values and standard deviations for HVLT and AVLT recall at each study visit are provided in the Appendix.

Shapes of the AVLT and HVLT Learning Curves

Results of curve-fitting exercises using baseline recall are summarized in Table 2 for AVLT and HVLT. In all models, regardless of curve shape, the estimated mean number of words recalled at the first AVLT and HVLT trials was approximately six and seven words, respectively. Participants recalled about 5.5 more words between the first and fifth trials in the AVLT and about three more words between the first and third trials in the HVLT. In the AVLT, as shown by the proportion of words recalled from a model with a

freely estimated curve shape, approximately 49% of the number of words recalled on the final trial was recalled by the second trial. In other words, participants recall almost as many more words between the first and second AVLT trials as they do during the remainder of the trials. For the HVLT, individuals, on average, recall about 73% of their final recall by the second trial. The RMSEA was poor for all models, but this statistic is less trustworthy because the models had small degrees of freedom. As shown by the CFI and TLI model fit statistics, approximately logarithmic growth provided better fit to trial-specific AVLT and HVLT growth than did linear or logarithmic curves, suggesting that shape best characterizes learning across the five AVLT and three HVLT trials (Table 2). In both tests, logarithmic growth systematically underestimated the number of words recalled at each trial.

Table 3. Model Parameter Estimates for Longitudinal Growth in Trial-specific AVLT Recall (n = 1,401)

	Memory-trained group estimate (SE)	Control group estimate (SE)	p Value	Cohen's d for treatment differences
Means				
Initial recall				
Intercept	6.29 (0.08)	6.23 (0.07)	.55	0.05
Pre- and posttraining change	0.15 (0.37)	0.00 (0.00)	.67	0.11
Trajectory	−0.03 (0.02)	−0.08 (0.02)	.03	0.35
Learning curve				
Intercept	5.51 (0.08)	5.38 (0.05)	.18	0.12
Pre- and posttraining change	1.65 (0.42)	0.00 (0.00)	<.001	1.58
Trajectory	−0.10 (0.02)	−0.06 (0.02)	.12	0.40
Variances				
Initial recall				
Intercept	2.38 (0.13)	Same		
Trajectory	0.02 (0.01)	Same		
Learning curve				
Intercept	1.09 (0.09)	Same		
Trajectory	0.01 (0.00)	Same		
Covariances[a]				
Initial I, initial T	0.02 (0.03)	0.04 (0.03)		
Initial I, learning I	0.63 (0.11)	0.60 (0.11)		
Initial I, learning T	−0.05 (0.03)	0.00 (0.04)		
Initial T, learning I	0.02 (0.03)	0.00 (0.02)		
Initial T, learning T	−0.04 (0.03)	0.02 (0.02)		
Learning I, learning T	0.00 (0.01)	0.00 (0.01)		
Initial recall indicator residual variance	1.12 (0.05)	Same		
Trajectory indicator residual variance	1.00 (0.06)	Same		
Model fit statistics				
RMSEA	0.049			
CFI	0.953			
TLI	0.956			

Note. RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis Index. Results from a multiple-group second-order latent growth model of trial-specific AVLT recall over time that assumes approximately logarithmic growth in trial recall during each study visit. The model setup is described in the Method section and shown graphically in Figure 1. The variance of the pre- and posttraining change was fixed to zero because this parameter captures change between only the baseline and immediate posttraining visit, and thus covariances between this parameter and other parameters are not shown because they are also zero.

[a] I refers to intercept, and T refers to trajectory.

Covariances between initial recall and learning curve were significant and negative for the AVLT and HVLT, which is consistent with other studies showing that memory span is inversely correlated with learning. This correlation is not attributable to ceiling effects. To test the assumption that this correlation is attributable to ceiling effects, we reran the models excluding participants who recalled more than 12 words in any AVLT trial (n = 349) or more than 10 words in any HVLT trial (n = 333); inferences were unchanged. The test–retest correlation among control group participants for Trial 1 recall was slightly lower than that for other trials; so different reliabilities may contribute to correlations between initial recall and learning but are unlikely to reverse the association. Indicator residual variances were smaller than initial-recall residual variances, indicating more between-person heterogeneity than within-person variation across trials in initial recall, which might be explained by between-person variables such as age, sex, or education. Findings were similar in models for all follow-up visits, for which CFI and TLI fit statistics were excellent to moderate (all CFI > 0.96; TLI > 0.96).

Longitudinal Changes in AVLT and HVLT Memory and Learning

Longitudinal changes in initial recall and learning in the AVLT and HVLT for the control and memory-trained groups were estimated using second-order multiple-group LGMs. Model fit for the AVLT was excellent (Table 3). Factor loadings on the learning parameters at each study visit conforming to an approximately logarithmic curve shape were constrained to be equal across intervention groups and visits to ensure measurement invariance of these constructs over time (Hayden et al., 2011). Changes in mean levels of words recalled were captured by pre- and posttraining and slope parameters. Results for the AVLT are shown in Table 3 and graphically in Figure 2. The mean number of words recalled during the first trial of the first visit was similar between intervention groups (p = .55; Table 3), and there was a minimal immediate effect of memory training on initial recall (Cohen's d = 0.11, p = .67); but rate of decline in initial trial recall was slower in the memory-trained group (−0.03 words per year on the initial trial) than among
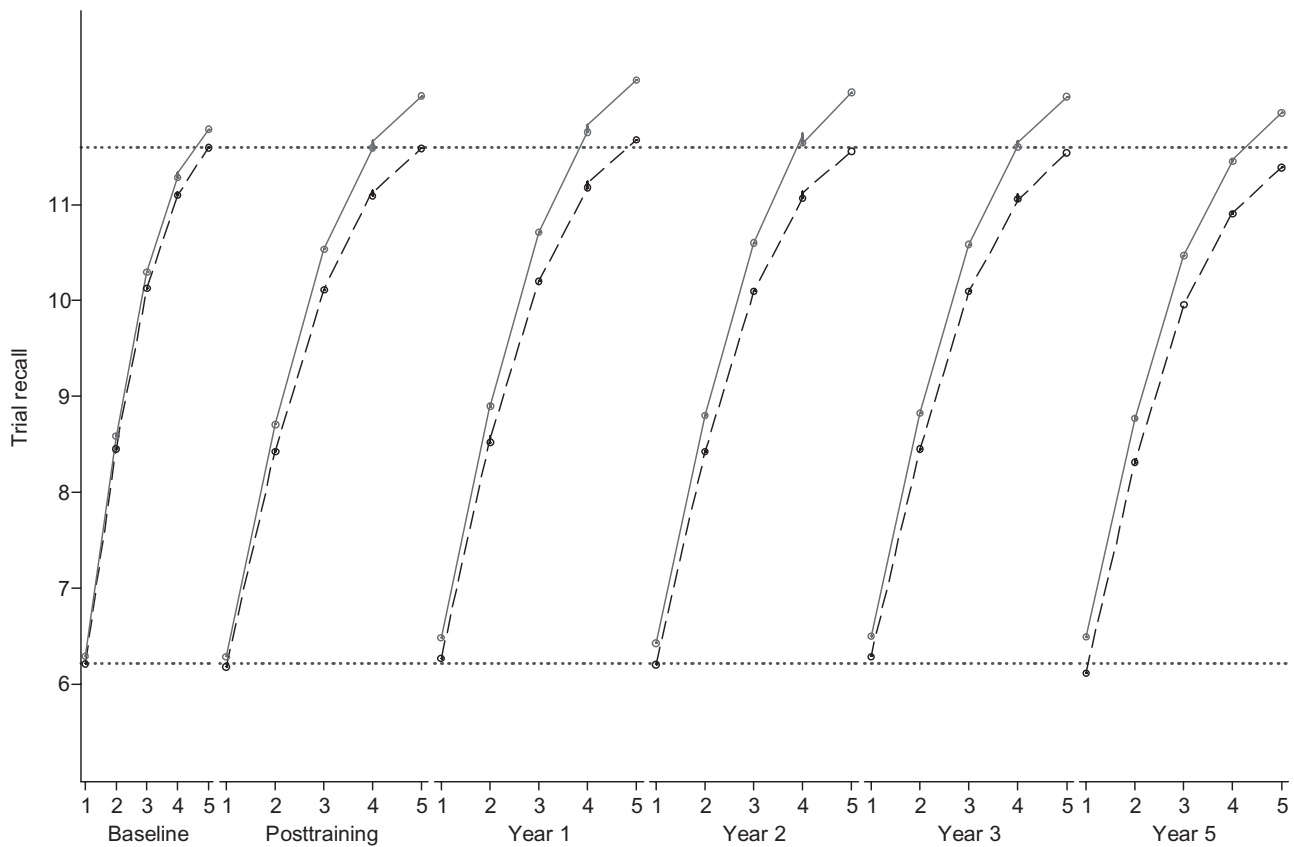
Figure 2. Longitudinal trajectories of AVLT recall and learning: Results from ACTIVE ($n = 1,401$). Graphic results from a multiple-group second-order latent growth model of trial-specific AVLT recall over time. Trial-specific growth was modeled as an approximately logarithmic trajectory. Dashed line: control group; solid line: memory-trained group; dotted lines: reference lines (estimated mean recall for controls on the first and final trials at the baseline study visit).

controls (-0.08 words per year on the initial trial) by an amount corresponding to a small standardized training-effect size ($d = 0.35$, $p = .03$).

For the AVLT, there were no baseline differences in rate of learning ($d = 0.12$, $p = .18$), but the immediate pre- and posttraining change in learning was significantly greater in the memory-trained group than in the control group ($d = 1.56$, $p < .001$). Memory-trained participants recalled an estimated 1.65 more words between the first and final trials after memory training (Table 3). Figure 2 reveals that the learning curve in the memory-trained group was still greater than the control group's through the fifth-year visit. Long-term change in the rate of learning after immediate posttraining did not differ significantly between memory-trained (-0.10 additional words recalled between the first and final trials per year) and control groups (-0.06 additional words; $d = 0.40$, $p = .12$).

Similar models were estimated to characterize growth in HVLT initial recall and learning (Table 4 and Figure 3). Model fit was excellent (Table 4). Similar to the AVLT (Table 3 and Figure 2), the memory-trained group showed less decline in initial recall by 0.06 words annually on the first trial ($d = 0.60$, $p = .01$) and significant pre- and

posttraining gain in rate of learning by 1.7 additional words per year ($d = 3.10$, $p < .001$) relative to the control group. The effect size for the immediate training gain in rate of learning for the HVLT was twice that of the AVLT (3.10 vs. 1.56). Steeper decline in learning rate after the posttraining visit was also observed ($d = 0.90$, $p < .001$, Table 4), although Figure 3 shows that memory-trained participants did not lose their training boost by the fifth year relative to the control group. In a follow-up analysis, we obtained correlations of intercepts and slopes for initial recall and learning parameters from a model that combined AVLT and HVLT growth processes (RMEA = 0.042; CFI = 0.95; TLI = 0.95). Intercepts of initial recall ($r = 0.90$), trajectories of initial recall ($r = 0.49$), intercepts of learning ($r = 0.83$), and trajectories of learning ($r = 0.67$) from the AVLT and HVLT were each highly correlated with their counterparts.

For comparison, longitudinal changes in AVLT and HVLT sum scores across trials were modeled with multiple-group LGMs (results available upon request). Consistent with the results described previously, there were significant pre- and posttraining gains in performance in the memory-trained group with both the AVLT ($d = 0.64$, $p < .001$) and HVLT

Table 4. Model Parameter Estimates for Longitudinal Growth in Trial-specific HVLT Recall ($n = 1{,}401$)

	Memory-trained group estimate (SE)	Control group estimate (SE)	p Value	Cohen's d for treatment differences
Means				
Initial recall				
Intercept	7.13 (0.08)	7.04 (0.07)	.37	0.05
Pre- and posttraining change	0.67 (0.37)	0.00 (0.00)	.07	0.41
Trajectory	−0.07 (0.02)	−0.13 (0.02)	.01	0.60
Learning curve				
Intercept	3.00 (0.06)	2.96 (0.04)	.53	0.07
Pre- and posttraining change	1.70 (0.37)	0.00 (0.00)	<.001	3.10
Trajectory	−0.06 (0.02)	0.03 (0.02)	<.001	0.90
Variances				
Initial recall				
Intercept	2.73 (0.13)	Same		
Trajectory	0.01 (0.00)	Same		
Learning curve				
Intercept	0.30 (0.05)	Same		
Trajectory	0.01 (0.01)	Same		
Covariances[a]				
Initial I, initial T	0.06 (0.03)	0.05 (0.03)		
Initial I, learning I	−0.24 (0.09)	−0.19 (0.09)		
Initial I, learning T	−0.01 (0.03)	0.01 (0.03)		
Initial T, learning I	−0.01 (0.02)	−0.03 (0.02)		
Initial T, learning T	0.01 (0.01)	0.01 (0.01)		
Learning I, learning T	0.01 (0.02)	0.02 (0.02)		
Initial recall indicator residual variance	1.40 (0.05)	Same		
Trajectory indicator residual variance	0.81 (0.06)	Same		
Model fit statistics				
RMSEA	0.060			
CFI	0.950			
TLI	0.954			

Note. RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis Index. Results from a multiple-group second-order latent growth model of trial-specific HVLT recall over time that assumes approximately logarithmic growth in trial recall during each study visit. The model setup is described in the Method section and shown graphically in Figure 1. The variance of the pre- and posttraining change was fixed to zero because this parameter captures the change between only the baseline and immediate posttraining visit, and thus covariances between this parameter and other parameters are not shown because they are also zero.

[a] I refers to intercept, and T refers to trajectory.

($d = 1.21$, $p < .001$). However, the subsequent mean trajectory of overall recall did not differ by intervention status for either AVLT ($d = 0.09$, $p = .49$) or HVLT ($d = 0.03$, $p = .83$). Interestingly, the pre- and posttraining effect from the HVLT was almost exactly double that from the AVLT for both models of trial-specific recall sum scores and sum of recall scores.

Predictors of Initial Recall and Learning

Longitudinal predictors of growth processes for initial recall and learning are provided in Table 5. To facilitate comparisons of the strength of associations between the AVLT and HVLT, coefficients in Table 5 are standardized with respect to the outcome and thus represent differences in the outcome in standard deviation units per unit difference in the predictor.

For both the AVLT and HVLT, younger, female, more highly educated, and generally healthier participants recalled more words on the first trial at baseline, as shown by coefficients for the initial-recall intercept. Greater AVLT serial and subjective clustering, and HVLT semantic and subjective clustering, at immediate posttraining were

associated with better initial recall, as shown by coefficients for the initial recall pre- and posttraining change. Female sex was associated with less immediate pre- and posttraining-related improvement in initial recall, and more HVLT semantic clustering was associated with a larger immediate pre- and posttraining improvement in HVLT initial recall. As shown by coefficients for the initial recall trajectory in Table 5, younger age and White ethnicity were associated with slower loss in initial recall over time for both the HVLT and AVLT, as was better self-rated health and HVLT semantic clustering for the HVLT.

Across the HVLT and AVLT, no demographic variables were consistently associated with differences in growth parameters for the learning curve. More subjective clustering in both the AVLT and HVLT was associated with a 0.9 *SD*s greater pre- and posttraining-related gain in recall, as demonstrated by the corresponding learning curve's pre- and posttraining change coefficients. Greater HVLT semantic clustering also was associated with preservation of long-term changes in HVLT recall by 0.2 *SD*s (Table 5).
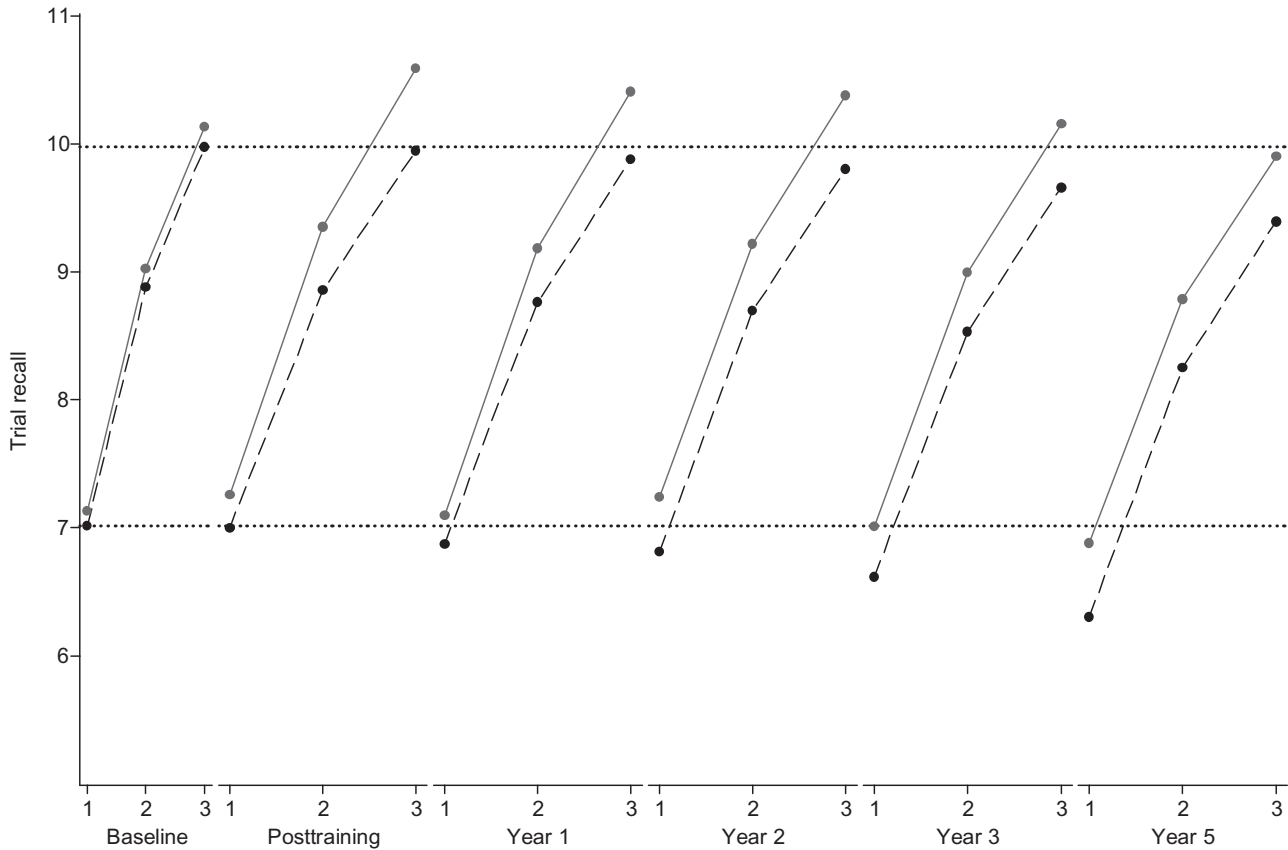
Figure 3. Longitudinal trajectories of HVLT recall and learning: Results from ACTIVE (*n* = 1,401). Graphic results from a multiple-group second-order latent growth model of trial-specific AVLT recall over time. Trial-specific growth was modeled as an approximately logarithmic trajectory. Dashed line: control group; solid line: memory-trained group; dotted lines: reference lines (estimated mean recall for controls on the first and final trials at the baseline study visit).

Discussion

This study investigated longitudinal trajectories of initial recall and learning using two word list–learning tests and explored effects of memory training on these measures. Individuals recall most of their final recall on the first trial of a word list–learning test, after which the rate of word learning follows approximately logarithmic growth. Training in memory strategies is associated with steep pre- and posttraining gains in learning, corresponding to large effect sizes, and slower long-term decline in memory span (initial recall) through up to five years, corresponding to moderate effects. Thus, the ACTIVE memory training produced lasting effects on memory span performance during the test both by slowing the normal (i.e., control group) pace of change in memory span and increasing the number of words recalled on repeat trials (learning). Models that used only trial sum recall scores to parameterize growth detected moderately sized pre- and posttraining–related performance gains that were smaller than effects using trial-specific recall sum scores, and detected no subsequent differences because they do not distinguish change in initial recall from change in learning. It is thus likely that any assessment of the impact of the ACTIVE

cognitive intervention trial that includes memory span performance as part of an outcome variable reflects an overly conservative estimate of training effectiveness.

Results of trial-specific growth are consistent with theories of incremental learning of associations (Estes, 1960) and replicate previous research showing that AVLT recall follows an approximately logarithmic shape (Jones et al., 2005). The present findings expand on previous research by highlighting that a sizable amount of learning takes place during the second recall trial and by uncovering sources of differences in recall attributable to memory training, namely, learning more than memory span. This study is the first to apply LGM methods to the study of learning and recall using the HVLT. The AVLT has twice as many trials as the HVLT, a longer word list, and unrelated words, in contrast to groups of semantically related words. Despite these important differences, trial learning curves follow a similar shape. Importantly, the immediate effect of training on recall was twice as large for the HVLT than the AVLT, probably because the HVLT contains semantically related words, and memory strategies for categorizing related items are more intuitive to implement than strategies for unrelated items (Craik, 1981; Hunt & Love, 1972).

Table 5. Demographic Predictors of AVLT and HVLT Initial Recall and Learning Components ($n = 1,401$)

	Initial recall			Learning curve		
	Intercept	Pre- and posttraining change	Trajectory	Intercept	Pre- and posttraining change	Trajectory
Parameter	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)
AVLT						
Age	−0.5* (0.1)	0.5 (0.5)	−0.8* (0.2)	−0.2* (0.1)	−0.3 (0.5)	0.0 (0.2)
Sex (female)	0.7* (0.1)	−1.1* (0.5)	0.0 (0.2)	0.3* (0.1)	0.4 (0.6)	−0.1 (0.3)
Education	0.2* (0.1)	0.6 (0.4)	0.0 (0.1)	0.0 (0.1)	−0.6 (0.4)	0.0 (0.2)
Self-rated health						
Very good	−0.2 (0.1)	0.6 (0.8)	0.2 (0.3)	0.5* (0.2)	−1.4* (0.7)	0.2 (0.4)
Good	−0.3* (0.1)	0.1 (0.8)	0.1 (0.3)	0.2 (0.2)	−1.0 (0.7)	0.2 (0.4)
Fair	−0.5* (0.1)	1.0 (1.0)	−0.1 (0.3)	0.1 (0.2)	−0.8 (0.9)	−0.2 (0.5)
Poor	−0.8* (0.3)	−1.1 (3.2)	0.8 (0.7)	0.8 (0.8)	−0.8 (3.0)	−3.6* (1.1)
Ethnicity (White)	0.0 (0.1)	0.2 (0.7)	0.5* (0.2)	0.4* (0.1)	−0.5 (0.6)	−0.1 (0.3)
AVLT serial clustering	0.3* (0.1)	0.0 (0.5)	−0.2 (0.1)	0.2* (0.1)	0.3 (0.4)	−0.1 (0.2)
AVLT subjective clustering	0.3* (0.1)	0.7 (0.5)	−0.3 (0.1)	0.1 (0.1)	0.9* (0.4)	−0.3 (0.2)
Coefficient of determination (R^2)	0.55	N/A	0.27	0.24	N/A	0.20
HVLT						
Age	−0.4* (0.1)	0.2 (0.4)	−0.7* (0.1)	0.1 (0.1)	0.0 (0.5)	−0.1 (0.2)
Sex (female)	0.5* (0.1)	−0.5 (0.5)	0.2 (0.2)	−0.2 (0.2)	0.2 (0.7)	−0.4 (0.3)
Education	0.2* (0.1)	−0.2 (0.3)	0.1 (0.1)	0.1 (0.1)	−0.8 (0.5)	0.1 (0.2)
Very good	0.0 (0.1)	−1.2* (0.6)	0.7* (0.3)	0.2 (0.2)	0.3 (0.9)	−0.5 (0.4)
Good	−0.1 (0.1)	−1.1* (0.5)	0.6* (0.3)	0.2 (0.2)	−0.1 (0.9)	−0.5 (0.4)
Fair	−0.3* (0.1)	−0.4 (0.7)	0.2 (0.3)	0.4 (0.3)	−1.0 (1.0)	−0.8 (0.5)
Poor	−0.1 (0.4)	−1.5 (2.2)	−1.1 (0.8)	0.1 (0.7)	−2.6 (3.2)	1.5 (1.4)
Ethnicity (White)	0.2* (0.1)	0.1 (0.5)	0.5* (0.2)	−0.1 (0.2)	0.2 (0.7)	−0.4 (0.3)
HVLT semantic clustering	0.3* (0.0)	0.5* (0.1)	−0.3* (0.1)	0.0 (0.1)	−0.4 (0.3)	0.2* (0.1)
HVLT subjective clustering	0.1* (0.1)	0.1 (0.3)	−0.1 (0.1)	−0.1 (0.1)	0.9* (0.3)	−0.2 (0.1)
Coefficient of determination (R^2)	0.61	N/A	0.35	0.05	N/A	0.20

Note. AVLT = Auditory Verbal Learning Test; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis Index; HVLT = Hopkins Verbal Learning Test. Coefficients represent the standard deviation (SD) difference in the outcome per unit change in the predictor. Age coefficients represent the SD difference in outcome per 10-year difference in age. Education coefficients represent the SD difference in outcome per 4-year difference in education. The reference for self-rated health indicators is excellent health. Coefficients of determination (R^2) represent contributions of all the covariates to the prediction of variance in the outcomes; pre- and posttraining parameters have no values because they have zero variance.

This study found many demographic predictors of memory span, but only older age and ethnicity were associated with faster decline in memory span. Moreover, we found no significant demographic predictors of long-term change in the learning curve but did find that increased strategy use was associated with change in learning. We are aware of no other study that has studied predictors of longitudinal change in these parameters among older adults, so this represents a novel contribution. Previous studies have shown older age and less education are cross-sectionally associated with lower initial recall, or memory span (Bolla-Wilson & Bleeker, 1986; Jones et al., 2005; Mitrushina, Satz, Chervinsky, & D'Elia, 1991), but not rate of learning (Jones et al., 2005; Mitrushina et al., 1991). Jones et al. (2005) also reported that female sex and White ethnicity were associated with greater initial recall. These findings are broadly consistent with the results of this study; although we did find relationships between younger age and female sex with higher initial learning using the AVLT, these results were not replicated with the HVLT.

We are not aware of other studies that have associated objective measures of strategy use with level and change in initial recall and learning. Higher levels of serial and subjective clustering in the AVLT and semantic and subjective clustering in the HVLT were associated with greater initial recall, and subjective clustering also predicted greater pre- to posttraining changes in learning rate consistently across both tests. Subjective clustering can be thought of as an indicator of the use of idiosyncratic strategies after adjusting for serial clustering in the AVLT or semantic clustering in the HVLT because it entails consistency in word recall across trials not explained by other clustering indices. Further, clustering words in semantic units in the HVLT appears to enable participants to learn more over time. These findings are intuitive: The manner in which older adults approach a memory task is a better predictor of performance than demographic characteristics.

Multitrial learning tests afford better measurement of training effects on memory than single-trial memory tests (Royall et al., 2003). One reason is because repeated exposure to stimuli provides more opportunities to apply memory strategies for list learning. Because strategies, measured by clustering scores in this study, mediate training-related changes in everyday functioning as well as memory performance (Gross & Rebok, 2011), learning in

multitrial tests that facilitate ample use of strategies may be more highly correlated with everyday function and may even mediate training-related changes in everyday function. Future research is needed to further investigate this question. Convergent findings from the AVLT and HVLT suggest that even a short word list–learning test like the HVLT, which takes less than half the time to administer than the AVLT, can provide the same information about attention and memory learning. The shallower trajectory of change in initial recall in the memory-trained group is a novel finding and suggests that training helped older adults in part by enabling swifter adoption of relevant strategies at the beginning of the trial rather than during the trial.

In conclusion, this study evaluated the effect of memory training on the level of and changes in initial recall and learning among cognitively normal older adults. Findings highlight the effect of memory training on the rate of learning that occurs during verbal list–learning tests, which can inform future research efforts in learning and evaluations of training. Regardless of the number of trials in a test, memory training effects are most apparent during earlier trials on word list–learning tasks, when most learning takes place, as evidenced by slower longitudinal decline in initial memory span and a large immediate effect on the rate of learning that was maintained for up to five years in this study. Effect sizes for memory training on learning were larger after distinguishing memory span performance from learning, a finding that is probably also true of other determinants of memory besides training. In terms of predictors considered in this study, many demographic variables were associated with level of recall; but memory strategies in the AVLT and HVLT were the strongest determinants of learning.

CORRESPONDENCE

Correspondence should be addressed to Alden L. Gross, PhD, Hebrew SeniorLife, Boston, MA 02131. E-mail: aldengross@hsl.harvard.edu.

REFERENCES

Albano, A. D. (2011). equate: Statistical methods for test equating [Computer software manual]. Available from http://CRAN.R-project.org/package=equate (R package)

Albert, M. (2008). The neuropsychology of the development of Alzheimer's disease. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (4th ed., pp. 97–132). London: Academic Press.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General*, *113*, 518–540. doi:10.1037//0096-3445.113.4.518

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., . . .Willis, S. L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association*, *288*, 2271–2281. doi:10.1001/jama.288.18.2271

Bolla-Wilson, K., & Bleeker, M. (1986). Influence of verbal intelligence, sex, age and education on the Rey auditory verbal learning test. *Developmental Neuropsychology*, *2*, 203–211. doi:10.1080/87565648609540342

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ. Wiley.

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging*, *25*, 767–778. doi:10.1037/a0020683

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The Clinical Neuropsychologist*, *5*, 125–142. doi:10.1080/13854049108403297

Brandt, J., & Benedict, R. H. B. (2001). *Hopkins verbal learning test–revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Buschkuehl, M., Jaeggi, S. M., Hutchison, S., Perrig–Chiello, P., Däpp, C., Müller, M., . . . Perrig, W. J. (2008). Impact of working memory training on memory performance in old–old adults. *Psychology and Aging*, *23*, 743–753. doi:10.1037/a0014342

Butters, N., & Cermak, L. S. (1980). *Alcoholic Korsakoff's syndrome: An information processing approach to amnesia*. New York: Academic Press. doi:10.1080/01688638008403805

Calero, M. D., & Navarro, E. (2007). Cognitive plasticity as a modulating variable on the effects of memory training in elderly persons. *Archives of Clinical Neuropsychology*, *22*, 63–72. doi:10.1016/j.acn.2006.06.020

Caprio-Prevette, M. D., & Fry, P. S. (1996). Memory enhancement program for community–based older adults: Development and evaluation. *Experimental Aging Research*, *22*, 281–303. doi:10.1080/03610739608254012

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Craik, F. I. M. (1981). Encoding and retrieval effects in human memory: A partial review. In J. B. Long (Ed.), *Attention and performance* (Vol. 9, pp. 383–402). Hillsdale, NJ: Lawrence Erlbaum.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (2000). *California verbal learning test* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Donders, A. R., van der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*, 1087–1091. doi:10.1016/j.jclinepi.2006.01.014

Duncan, T. E., Duncan, S. C., & Stryker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Erlbaum.

Ebbinghaus, H. (1895/1964). *Memory: A contribution to experimental psychology*. New York, NY: Dover.

Estes, W. K. (1960). Learning theory and the new "mental chemistry." *Psychological Review*, *67*, 207–223. doi:10.1037/h0041624

Ferrer, E., Salthouse, T. A., Stewart, W. F., & Schwartz, B. S. (2004). Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychology & Aging*, *19*, 243–259. doi:10.1037/0882-7974.19.2.243

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini–mental state": A practical guide for grading the cognitive state of patients

for the clinician. *Journal of Psychiatric Research*, *12*, 189–198. doi:10.1016/0022-3956(75)90026-6

Gross, A. L., Inouye, S. K., Rebok, G. W., Brandt, J., Crane, P. K., Parisi, J. M., . . . Jones, R. N. (2012). Parallel but not equivalent: Challenges and solutions for repeated assessment of cognition over time. *Journal of Clinical and Experimental Neuropsychology*. Epub ahead of print. doi:10.1080/13803395.2012.681628

Gross, A. L., & Rebok, G. W. (2011). Memory training and strategy use in older adults: Results from the ACTIVE cognitive intervention trial. *Psychology & Aging*, *26*, 503–517. doi:10.1037/a0022687

Hastings, E. C., & West, R. L. (2009). The relative success of a self-help and a group-based memory training program for older adults. *Psychology and Aging*, *24*, 586–594. doi:10.1037/a0016951

Hayden, K. M., Jones, R. N., Zimmer, C., Plassman, B. L., Browndyke, J. N., Pieper, C., . . . Welsh-Bohmer, K. A. (2011). Factor structure of the National Alzheimer's Coordinating Centers uniform dataset neuropsychological battery: An evaluation of invariance between and within groups over time. *Alzheimer Disease & Associated Disorders*, *25*, 128–137. doi:10.1097/WAD.0b013e3181ffa76d

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. doi:10.1080/10705519909540118

Hunt, E. B., & Love, T. (1972). How good can memory be? In A. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, DC: V. H. Winston & Sons.

Jobe, J. B., Smith, D. M., Ball, K., Tennstedt, S. L., Marsiske, M., Willis, S. L., . . . Kleinman, K. (2001). ACTIVE: A cognitive intervention trial to promote independence in older adults. *Controlled Clinical Trials*, *22*, 453–479. doi:10.1016/S0197-2456(01)00139-8

Jones, R. N., Rosenberg, A. L., Morris, J. N., Allaire, J. C., McCoy, K. J., Marsiske, M., . . . Malloy, P. F. (2005). A growth curve model of learning acquisition among cognitively normal older adults. *Experimental Aging Research*, *31*, 291–312. doi:10.1080/03610730590948195

Kolen, M., & Brennan, R. (1995). *Test equating: Methods and practices*. New York: Springer.

Langbaum, J. B. S., Rebok, G. W., Bandeen-Roche, K., & Carlson, M. C. (2009). Predicting memory training response patterns: Results from ACTIVE. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *64*, 14–23. doi:10.1093/geronb/gbn026

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York, NY: Oxford University Press.

Li, S. C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging*, *23*, 731–742. doi:10.1037/a0014343

Macartney-Filgate, M. S., & Vriezen, E. R. (1988). Intercorrelation of clinical tests of verbal memory. *Archives of Clinical Neuropsychology*, *3*, 121–126. doi:10.1093/arclin/3.2.121

Magalhães, S. S., & Hamdan, A. C. (2010). Influence of verbal intelligence, sex, age and education on the Rey auditory verbal learning test. *Psychology & Neuroscience*, *3*, 85–91. doi:10.3922/j.psns.2010.1.011

Mahncke, H. W., Bronstone, A., & Merzenich, M. M. (2006). Memory enhancement in healthy older adults using a brain plasticity–based training program: A randomized, controlled study. *Proceedings of the National Academy of Sciences, USA*, *103*, 12523–12528. doi:10.1073/pnas.0605194103

Martin, M., Clare, L., Altgassen, A. M., Cameron, M. H., & Zehnder, F. (2011). Cognition-based interventions for healthy older people and people with mild cognitive impairment. *Cochrane Database of Systematic Reviews*, CD006220. doi:10.1002/14651858.CD006220.pub2

McArdle, J. J., & Bell, R. Q. (2000). Recent trends in modeling longitudinal data by latent growth curve methods. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple–group data: Practical issues, applied approaches, and scientific examples* (pp. 69–108). Mahwah, NJ: Lawrence Erlbaum.

McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumaker (Eds.), *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Erlbaum.

McDaniel, M. A., Einstein, G. O., & Jacoby, L. J. (2008). New considerations in aging and memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (4th ed. pp. 251–310). London: Academic Press.

Mitrushina, M., Satz, P., Chervinsky, A., & D'Elia, L. (1991). Performance of four age groups of normal elderly on the Rey Auditory Verbal Learning Test. *Journal of Clinical Psychology*, *47*, 351–357. doi:10.1002/1097-4679(199105)47:3<351::AID-JCLP2270470305>3.0.CO;2-S

Muthén, B. O. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological methodology* (pp. 453–480). Boston: Blackwell.

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*, 371–402. doi:10.1037//1082-989X.2.4.371

Muthén, L.K., & Muthén, B.O. (1998–2010). *Mplus user's guide: Sixth Edition*. Los Angeles, CA: Muthén & Muthén.

Neely, A. S., & Bäckman, L. (1993). Long-term maintenance of gains from memory training in older adults: Two 3 1/2-year follow-up studies. *Journal of Gerontology*, *48*, 233–237. doi:10.1080/0360127930190202

Nettelbeck, T., Rabbitt, P., Wilson, C., & Batt, R. (1996). Uncoupling learning from initial recall: The relationship between speed and memory deficits in old age. *British Journal of Psychology*, *87*, 593–607. doi:10.1111/j.2044-8295.1996.tb02610.x

Poreh, A. (2005). Analysis of mean learning of normal participants on the Rey Auditory-Verbal Learning Test. *Psychological Assessment*, *17*, 191–199. doi:10.1037/1040-3590.17.2.191

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences*, *9*, 374–380. doi:10.1016/j.tics.2005.06.009

Rebok, G. W., & Balcerak, L. J. (1989). Memory self-efficacy and performance differences in young and old adults: The effect of mnemonic training. *Developmental Psychology*, *25*, 714–721. doi:10.1037//0012-1649.25.5.714

Rebok, G. W., Carlson, M. C., & Langbaum, J. B. S. (2007). Training and maintaining memory abilities in healthy older adults: Traditional and novel approaches. *Journal of Gerontology: Series B*, *62*, 53–61. doi:10.1093/geronb/62.special_issue_1.53

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Royall, D. R., Palmer, R., Chiodo, L. K., & Polk, M. (2003). Decline in learning ability best predicts future dementia type: The Freedom House study. *Experimental Aging Research*, *29*, 285–406. doi:10.1080/03610730303700

Ryan, J. D., & Cohen, N. J. (2004). Processing and short-term retention of relational information in amnesia. *Neuropsychologia*, *42*, 497–511. doi:10.1016/j.neuropsychologia.2003.08.011

Ryan, J. J., Geisser, M. E., Randall, D. M., & Georgemiller, R. J. (1986). Alternate form reliability and equivalency of the Rey Auditory-Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*, *8*, 611–616. doi:10.1080/01688638608405179

Schmidt, M. (2004). *Rey auditory and verbal learning test: A handbook*. Los Angeles: Western Psychological Services.

Scogin, F., & Bienias, J. L. (1988). A three-year follow-up of older adult participants in a memory-skills training program. *Psychology and Aging*, *3*, 334–337. doi:10.1037//0882-7974.3.4.334

Scogin, F., & Prohaska, M. (1992). The efficacy of self–taught memory training for community–dwelling older adults. *Educational Gerontology*, *18*, 751–766. doi:10.1080/0360127920180801

Scoville, W. B. (1968). Amnesia after bilateral mesial temporal-lobe excision: Introduction to case H.M. *Neuropsychologia*, *6*, 211–213. doi:10.1016/0028-3932(68)90020-1

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly Journal of Experimental Psychology*, *22*, 261–273. doi:10.1080/00335557043000203

Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 155–167. doi:10.3758/CABN.3.3.155

Steiger, J. H. (1989). *EZPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: Systat.

Stigsdotter, A., & Bäckman, L. (1989). Multifactorial memory training with older adults: How to foster maintenance of improved performance. *Gerontology*, *35*, 260–267. doi:10.1159/000213035

Stricker, J. L., Brown, G. B., Wixted, J., & Delis, D. C. (2002). New semantic and serial clustering indices for the California Verbal Learning Test–second edition: Background, rationale, and formulae. *Journal of the International Neuropsychological Society*, *8*, 425–435. doi:10.1017/S1355617702813224

Underwood, B. J. (1963). Coding processes in verbal learning. *Journal of verbal learning and verbal behavior*, *1*, 250–257. doi:10.1016/S0022-5371(63)80003-1

Vakil, E., & Blachstein, H. (1997). Rey AVLT: Developmental norms for adults and the sensitivity of different memory measures to age. *The Clinical Neuropsychologist*, *11*, 356–369. doi:10.1080/13854049708400464

Verhaeghen, P., Geraerts, N., & Marcoen, A. (2000). Memory complaints, coping, & well–being in old age: A systemic approach. *Gerontologist*, *40*, 540–548. doi:10.1093/geront/40.5.540

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging*, *7*, 242–251. doi:10.1037//0882-7974.7.2.242

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36–item Short–Form Health Survey (SF–36®): I. Conceptual framework and item selection. *Medical Care*, *30*, 473–483. doi:10.1097/00005650-199206000-00002

Willis, S. L., & Nesselroade, C. S. (1990). Long-term effects of fluid ability training in old-old age. *Developmental Psychology*, *26*, 905–910. doi:10.1037//0012-1649.26.6.905

Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., . . . ACTIVE Study Group. (2006). Long–term effects of cognitive training on everyday functional outcomes in older adults. *Journal of the American Medical Association*, *296*, 2805–2814. doi:10.1001/jama.296.23.2805

Wolk, D. A., Dickerson, B. C., & ADNI Investigators (2011). Fractionating verbal episodic memory in Alzheimer's disease. *Neuroimage*, *54*, 1530–1539. doi:10.1016/j.neuroimage.2010.09.005

Zehnder, F., Martin, M., Altgassen, M., & Clare, L. (2009). Memory training effects in old age as markers of plasticity: A meta–analysis. *Restorative Neurology and Neuroscience*, *27*, 507–520. doi:10.3233/RNN-2009-0491

APPENDIX Descriptive Statistics of HVLT and AVLT Trial Recall at All Study Visits

	Baseline			Immediate posttraining			First annual		
	Memory training ($n = 703$)	Controls ($n = 698$)	p Value for difference	Memory training ($n = 703$)	Controls ($n = 698$)	p Value for difference	Memory training ($n = 703$)	Controls ($n = 698$)	p Value for difference
AVLT scores, mean (SD)									
Trial 1	6.2 (2.0)	6.1 (2.1)	0.47	6.3 (1.9)	6.1 (1.9)	0.18	6.4 (2.0)	6.1 (2.1)	0.05
Trial 2	9.0 (2.4)	8.8 (2.5)	0.32	9.0 (2.4)	8.8 (2.4)	0.15	9.3 (2.3)	9.0 (2.4)	0.01
Trial 3	10.6 (2.5)	10.3 (2.6)	0.03	10.7 (2.5)	10.3 (2.5)	<0.01	11.2 (2.6)	10.6 (2.6)	<0.001
Trial 4	11.3 (2.5)	11.2 (2.5)	0.23	11.6 (2.5)	11.1 (2.5)	<0.001	11.8 (2.4)	11.2 (2.5)	<0.001
Trial 5	11.8 (2.4)	11.7 (2.4)	0.29	12.2 (2.3)	11.6 (2.4)	<0.001	12.3 (2.4)	11.6 (2.4)	<0.001
Learning curve (T5–T1)	5.6 (2.1)	5.5 (2.0)	0.57	5.9 (1.9)	5.5 (1.9)	<0.001	5.8 (2.0)	5.4 (2.1)	<0.01
Sum of Trials 1–5	48.8 (10.6)	47.9 (11.0)	0.11	48.1 (10.9)	46.1 (11.5)	<0.01	47.1 (10.7)	44.6 (10.8)	<0.001
HVLT scores, mean (SD)									
Trial 1	7.1 (2.1)	6.9 (2.1)	0.23	7.2 (2.3)	7.0 (2.2)	0.06	7.1 (2.1)	6.9 (2.1)	0.22
Trial 2	9.2 (2.1)	9.1 (2.1)	0.37	9.6 (2.2)	9.0 (2.2)	<0.001	9.5 (2.1)	9.0 (2.1)	<0.001
Trial 3	10.0 (1.9)	9.8 (1.9)	0.07	10.5 (2.0)	9.9 (2.0)	<0.001	10.5 (1.8)	10.0 (1.9)	<0.001
Learning curve (T3–T1)	3.0 (1.6)	2.9 (1.5)	0.52	3.3 (1.8)	2.9 (1.7)	<0.001	3.4 (1.6)	3.0 (1.6)	<0.01
Sum of Trials 1–3	26.0 (5.5)	25.7 (5.7)	0.25	25.2 (6.0)	23.7 (5.9)	<0.001	27.0 (5.2)	25.9 (5.2)	<0.001

	Second annual			Third annual			Fifth annual		
AVLT scores, mean (SD)									
Trial 1	6.3 (2.0)	6.1 (2.0)	0.17	6.5 (2.0)	6.3 (2.2)	0.20	6.4 (2.1)	6.0 (2.0)	0.01
Trial 2	9.3 (2.4)	8.9 (2.6)	0.02	9.3 (2.3)	8.9 (2.5)	0.01	9.2 (2.4)	8.7 (2.4)	0.01
Trial 3	11.0 (2.6)	10.4 (2.7)	<0.001	10.7 (2.5)	10.3 (2.6)	0.01	10.7 (2.5)	10.1 (2.6)	<0.01
Trial 4	11.7 (2.6)	11.1 (2.7)	<0.001	11.7 (2.4)	11.1 (2.6)	<0.001	11.6 (2.3)	11.0 (2.4)	<0.001
Trial 5	12.2 (2.4)	11.6 (2.5)	<0.001	12.2 (2.3)	11.6 (2.4)	<0.001	11.9 (2.1)	11.4 (2.3)	<0.01
Learning curve (T5–T1)	5.9 (2.1)	5.4 (2.1)	<0.01	5.7 (2.0)	5.3 (1.9)	<0.01	5.5 (1.9)	5.3 (1.9)	0.42
Sum of Trials 1–5	50.3 (10.3)	47.9 (11.2)	<0.001	51.5 (11.0)	49.4 (11.5)	0.01	47.6 (11.4)	45.2 (12.0)	0.01
HVLT, mean (SD)									
Trial 1	7.5 (2.2)	7.1 (2.3)	<0.001	7.2 (2.1)	6.9 (2.1)	0.02	7.2 (2.3)	6.7 (2.2)	0.01
Trial 2	9.8 (2.1)	9.2 (2.1)	<0.001	9.5 (2.1)	9.0 (2.1)	<0.01	9.5 (2.2)	9.0 (2.2)	<0.01
Trial 3	10.5 (1.8)	10.0 (1.9)	<0.001	10.3 (1.9)	9.9 (1.9)	<0.001	10.2 (2.1)	9.7 (2.1)	<0.01
Learning curve (T3–T1)	3.0 (1.5)	2.9 (1.6)	0.82	3.1 (1.6)	3.0 (1.6)	0.49	3.0 (1.5)	3.0 (1.5)	0.81
Sum of Trials 1–3	27.7 (5.8)	25.8 (6.5)	<0.001	28.3 (5.6)	27.1 (5.6)	<0.01	24.6 (6.7)	23.3 (6.5)	0.01

Note. AVLT = Auditory Verbal Learning Test; HVLT = Hopkins Verbal Learning Test. Learning curves were calculated as the arithmetic difference between the last and first trials.