

# Learning Efficiency: Identifying Individual Differences in Learning Rate and Retention in Healthy Adults



Psychological Science  
2018, Vol. 29(9) 1436–1450  
© The Author(s) 2018  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0956797618772540  
www.psychologicalscience.org/PS



Christopher L. Zerr<sup>1</sup>, Jeffrey J. Berg<sup>1</sup>, Steven M. Nelson<sup>2,3,4</sup>,  
Andrew K. Fishell<sup>5</sup>, Neil K. Savalia<sup>3</sup>, and  
Kathleen B. McDermott<sup>1,5</sup>

<sup>1</sup>Department of Psychological & Brain Sciences, Washington University in St. Louis; <sup>2</sup>VISN 17 Center of Excellence for Research on Returning War Veterans, Waco, Texas; <sup>3</sup>Center for Vital Longevity, School of Behavioral and Brain Sciences, University of Texas at Dallas; <sup>4</sup>Department of Psychology and Neuroscience, Baylor University; and <sup>5</sup>Department of Radiology, Washington University in St. Louis

## Abstract

People differ in how quickly they learn information and how long they remember it, yet individual differences in learning abilities within healthy adults have been relatively neglected. In two studies, we examined the relation between learning rate and subsequent retention using a new foreign-language paired-associates task (the learning-efficiency task), which was designed to eliminate ceiling effects that often accompany standardized tests of learning and memory in healthy adults. A key finding was that quicker learners were also more durable learners (i.e., exhibited better retention across a delay), despite studying the material for less time. Additionally, measures of learning and memory from this task were reliable in Study 1 ( $N = 281$ ) across 30 hr and Study 2 ( $N = 92$ ; follow-up  $n = 46$ ) across 3 years. We conclude that people vary in how efficiently they learn, and we describe a reliable and valid method for assessing learning efficiency within healthy adults.

## Keywords

learning efficiency, memory, learning rate, individual differences, open data, open materials

Received 3/31/17; Revision accepted 3/31/18

People vary in their ability to both learn and retain information. How does the rate at which people initially acquire information relate to their ability to remember it after a delay? Do quicker learners save time at the expense of long-term retention? Little is known about individual differences in these capacities or how they relate to each other, in part because of a lack of validated measures of learning abilities within healthy young adults. Most psychometrically validated tests of learning and memory were developed for neuropsychological purposes, such as detecting cognitive impairments in clinical samples or deficits associated with aging. These tests tend to lack the sensitivity necessary for assessing individual differences within healthy adults—especially younger adults, such as undergraduate students—who often score at or near maximum performance, resulting in undesirable ceiling effects

(Uttl, 2005; Uttl, Graf, & Richter, 2002). These ceiling effects have been observed in popular neuropsychological measures, including the second edition of the California Verbal Learning Test (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000), the fourth edition of the Wechsler Memory Scale (WMS-IV; Wechsler, 2009), and the Rey Auditory-Verbal Learning Test (Rey, 1964). The aforementioned ceiling effects restrict the range in performance within healthy young adults, making it difficult to observe individual differences within that population.

## Corresponding Author:

Kathleen B. McDermott, Washington University in St. Louis,  
Department of Psychological & Brain Sciences, 1 Brookings Dr.,  
Campus Box 1125, St. Louis, MO 63130  
E-mail: kathleen.mcdermott@wustl.edu

In an attempt to examine neural correlates of individual differences in long-term memory within healthy young adults, S. M. Nelson and colleagues (2016) developed a learning task designed to reduce ceiling effects and generate enough variability to begin to understand how neural activity during encoding differs as a function of a person's long-term memory abilities. In this task, participants learned Lithuanian-English word pairs across multiple study-test trials. These materials offer several advantages. First, verbal materials are discrete, easily scored items (Tulving, 1983). Second, a foreign-language paired-associates task is one in which strategies are not as readily available as in most standardized tests of learning and memory. For instance, the CVLT-II utilizes categorized word lists and can thus benefit from strategies such as chunking; the WMS-IV Logical Memory task involves story recall and thus provides a meaningful structure. Pairing a foreign-language word with its English equivalent also diminishes the semantic encoding one might use with arbitrarily selected English noun-noun pairs (Papagno, Valentine, & Baddeley, 1991). Third, English-speaking participants in the United States are less likely to have been exposed to Lithuanian than to more common foreign languages (e.g., Spanish), reducing the opportunity for participants to rely on prior relevant knowledge.

S. M. Nelson et al. (2016) used this task in a sample of 86 healthy young adults between the ages of 18 and 31 years and sought to obtain a "raw" learning measure thought to be relatively unconfounded with learner sophistication (in terms of practiced strategies and metacognitive understanding). They found that younger participants tended to perform better on the task. Most notable, however, was that the magnitude of neural activity during initial study of the paired-associates task was significantly related to later memory performance across the participants. Specifically, S. M. Nelson et al. found greater suppression of the *default mode network* during encoding for people who subsequently performed better on the task. Brain regions that comprise the default mode network have previously been shown to deactivate (or suppress activity) when one directs one's attention to the external environment (Gusnard & Raichle, 2001), leading S. M. Nelson et al. to hypothesize that greater deployment of attentional resources contributed to enhanced learning of the word pairs. The suppression was observed across all trial types (i.e., for instances of successful as well as unsuccessful learning) and thus was preliminarily interpreted as reflecting traitlike interpersonal differences in the ability to focus attention in service of learning complex, novel information. Whether this pattern of results truly reflects a stable, traitlike measure (as opposed to resulting from different mental states adopted by participants during

the task) has not been established. More generally, can we quantify a stable construct of learning ability, or is performance characterized by marked day-to-day variability? In the present work, we addressed these questions by examining the stability of one's learning performance across days and years and, in the process, sought to better understand how individual differences in the rate of acquisition of to-be-remembered information relate to individual differences in long-term memory performance for multiple time delays. Finally, we provide preliminary evidence regarding construct validity and consider potential mechanisms behind the learning differences.

## Study 1

### Method

**Participants.** To examine reliability of the task in a large representative sample, we conducted an alternate-form reliability study on Amazon's Mechanical Turk (MTurk) with 498 participants. Informed consent was obtained from all participants in accordance with standard Washington University human research practices, and participants were compensated \$12 for completion of the study. Because MTurk studies take place in uncontrolled environments, at the end of the second day we asked participants in a nonjudgmental way, with no risk to their compensation, whether they had written down any of the words during one or both sessions; 108 participants were excluded for doing so. In addition, 46 participants were excluded for failing to reach the learning criterion in the maximum 16 allotted tests, 28 were excluded for failing to finish both sessions, 21 were excluded for restarting the task after the study portion, 11 were excluded for reporting a neurological disorder, and 3 were excluded for having prior knowledge of the Lithuanian language. These selection criteria for participant inclusion were made a priori, and these participants were removed before any analysis of the data. The final sample of 281 participants included 162 females (57.7% of the sample) with a mean age of 33.0 years ( $SD = 10.3$ , range = 18–66) and 15.4 years of education ( $SD = 2.3$ , range = 4–25), which is more representative than a typical undergraduate sample. None of the remaining participants reported any learning disability diagnoses (e.g., attention-deficit disorder, dyslexia).

**Materials.** Learning material consisted of two lists of 45 Lithuanian-English word pairs (Lists A and B; see Tables S1 and S2, respectively, in the Supplemental Material) selected from previous norms (Grimaldi, Pyc, & Rawson, 2010); the two word lists were created to be equivalent in difficulty, defined as the mean recall performance for each item across three trials. Diacritical marks and typographic

ligatures were removed from the Lithuanian words to make them more similar to English words. For both lists, Lithuanian words varied in length from 3 to 9 characters ( $M = 6.1$ ) and 1 to 4 syllables ( $M = 2.3$ ). English words, all of which were concrete nouns, ranged from 3 to 8 characters ( $M = 4.6$ ) and 1 to 3 syllables ( $M = 1.2$ ). The combined Lithuanian-English word pairs ranged from 7 to 16 characters ( $M = 10.7$ ) and 2 to 7 syllables ( $M = 3.6$ ). Lithuanian-English pairs were selected to reduce the incidence of cognates (cf. T. O. Nelson & Dunlosky, 1994) and false friends. Word pairs were displayed in all capital letters on a white background in black, 48-point Arial font.

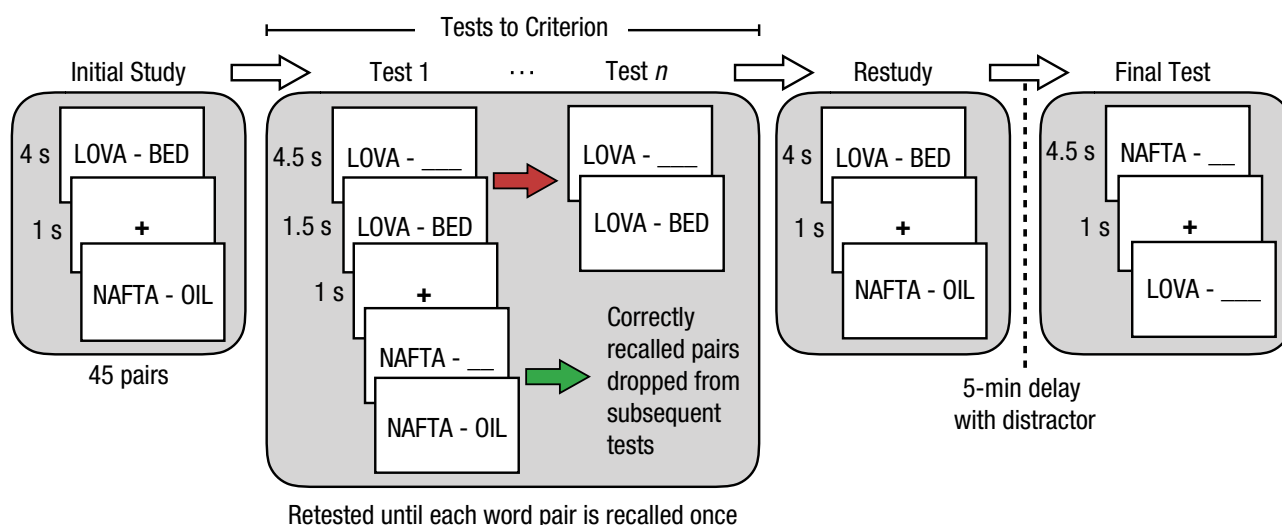
**Procedure.** This study occurred across 2 days (Fig. 1). On the first day, participants studied 45 Lithuanian-English word pairs presented in a different random order for each person. Pairs were presented one at a time for 4 s each and were separated by a 1-s interstimulus interval (ISI). Participants were instructed to learn each of the word pairs for a later cued-recall test.

After participants studied each word pair once, they took an initial cued-recall test (Test 1), which required them to type the English equivalent (e.g., “DRUM”) for the Lithuanian cue (e.g., “BUGNAS”) presented on screen for 4.5 s; test words were shown in a random order. Participants received immediate feedback: Regardless of response accuracy, the correct pairing was displayed for 1.5 s. An ISI of 1 s preceded the next

test cue. Correctly recalled pairs were dropped from subsequent tests until the final cued-recall test at the end of the session. This dropout procedure ensured that participants recalled each word pair exactly once to minimize overlearning.

Participants repeated this testing process on unrecalled word pairs until all 45 word pairs had been correctly recalled once. Each test block (set of previously unrecalled test pairs) was separated by 30 s of addition and subtraction mathematics problems (e.g.,  $7 - 13 = ?$ ) to prevent maintenance of the word pairs in working memory. The number of tests required for each participant to learn all 45 word pairs (tests to criterion) was used as an index of learning rate. The number of tests was limited to a maximum of 16 in the interest of time. After a participant reached criterion, all word pairs were presented once more in a new random order for a final study session, which was otherwise identical to the initial study session. Participants then played a video-game distractor (Tetris) for 5 min before being tested on all 45 word pairs for a final cued-recall test (final test), which was identical to Test 1 but without corrective feedback.

Participants underwent the same procedure approximately 1 day later ( $M = 30.4$  hr,  $SD = 13.5$ , range = 12.7–72.4) with a different set of 45 Lithuanian-English word pairs. Lists were counterbalanced across subjects. After each session, participants provided ratings (1–5, with 1 being the lowest) of their interest in the task,



**Fig. 1.** Procedure for Study 1, which took place across 2 days. On Day 1, participants ( $N = 281$ ) initially studied 45 Lithuanian-English word pairs once each (initial study) before taking an initial cued-recall test (Test 1) requiring them to provide the English word for a Lithuanian cue. Correctly recalled pairs were dropped from subsequent testing, whereas incorrectly recalled pairs were presented on the next test; this process was repeated for up to 16 tests until a participant correctly recalled each of the 45 word pairs exactly once (tests to criterion). In between each test, participants completed math problems for 30 s. A final restudy session took place (restudy) before a 5-min Tetris distractor task, after which participants took a final cued-recall test (final test). Participants repeated this process approximately a day later with a different set of 45 word pairs.

how difficult they thought the task was, and how much effort they expended when completing the task.

As will be seen, the three measures from this task—Test 1, tests to criterion, and final test—were intercorrelated and thus combined into an overall metric. Specifically, the standardized  $z$  scores for each person's Test 1, tests to criterion, and final test were averaged (tests to criterion was multiplied by  $-1$ , so fewer tests to criterion were better) to create a single metric of learning and memory performance (learning-efficiency score). A higher overall learning-efficiency score implies a quicker rate of learning (higher Test 1 scores and fewer tests to criterion) and better retention of the word pairs (higher final-test scores).

**Analysis.** Normality of dependent variables was assessed using the Shapiro-Wilk test; if normality was violated, the nonparametric Wilcoxon signed-rank test was used in place of a paired-samples  $t$  test. The nonparametric equivalent of Pearson's  $r$ —Spearman's rho ( $r_s$ )—was calculated for ordinal data. Effect sizes were calculated using Cohen's  $d$  for repeated measures (Cohen, 2009). Differences were considered significant if  $p$  was less than .05.

## Results

**Performance across days.** The learning curves for participants on the first day can be found in Figure 2 (top panel). Here one can see considerable variability in performance in all three measures (Test 1, tests to criterion, and final test) across participants, as well as across quartiles when binned by overall task performance (Fig. 2, bottom panel).

For Test 1, after studying the items once, participants on Day 1 recalled an average of 9.4 English words ( $SD = 6.6$ ) and on Day 2 an average of 11.1 words ( $SD = 8.2$ ). To reach criterion, participants took an average of 8.3 tests ( $SD = 2.9$ ) on Day 1 and 7.6 tests ( $SD = 2.8$ ) on Day 2. The average cued-recall score on the final test was 33.4 words ( $SD = 7.9$ ) on Day 1 and 33.2 words ( $SD = 8.7$ ) on Day 2. For Day 1, the entire task (including informed consent, directions, and the 5-min delay) took an average of 50.3 min to complete ( $SD = 13.2$ , range = 28.8–119.0), whereas Day 2 took an average of 45.7 min ( $SD = 14.0$ , range = 26.6–115.3). Additional descriptive statistics are in Table 1.

Participants who performed better on the initial test reached criterion more quickly (i.e., required fewer tests to criterion) on Day 1,  $r = -.60$ ,  $p < .001$ , 95% confidence interval (CI) =  $[-.67, -.52]$ , and Day 2,  $r = -.63$ ,  $p < .001$ , 95% CI =  $[-.69, -.55]$ . Participants who reached criterion quickly also had better retention of the word pairs after a delay (i.e., better final-test scores) on Day 1,  $r = -.57$ ,  $p < .001$ , 95% CI =  $[-.64, -.49]$ , and

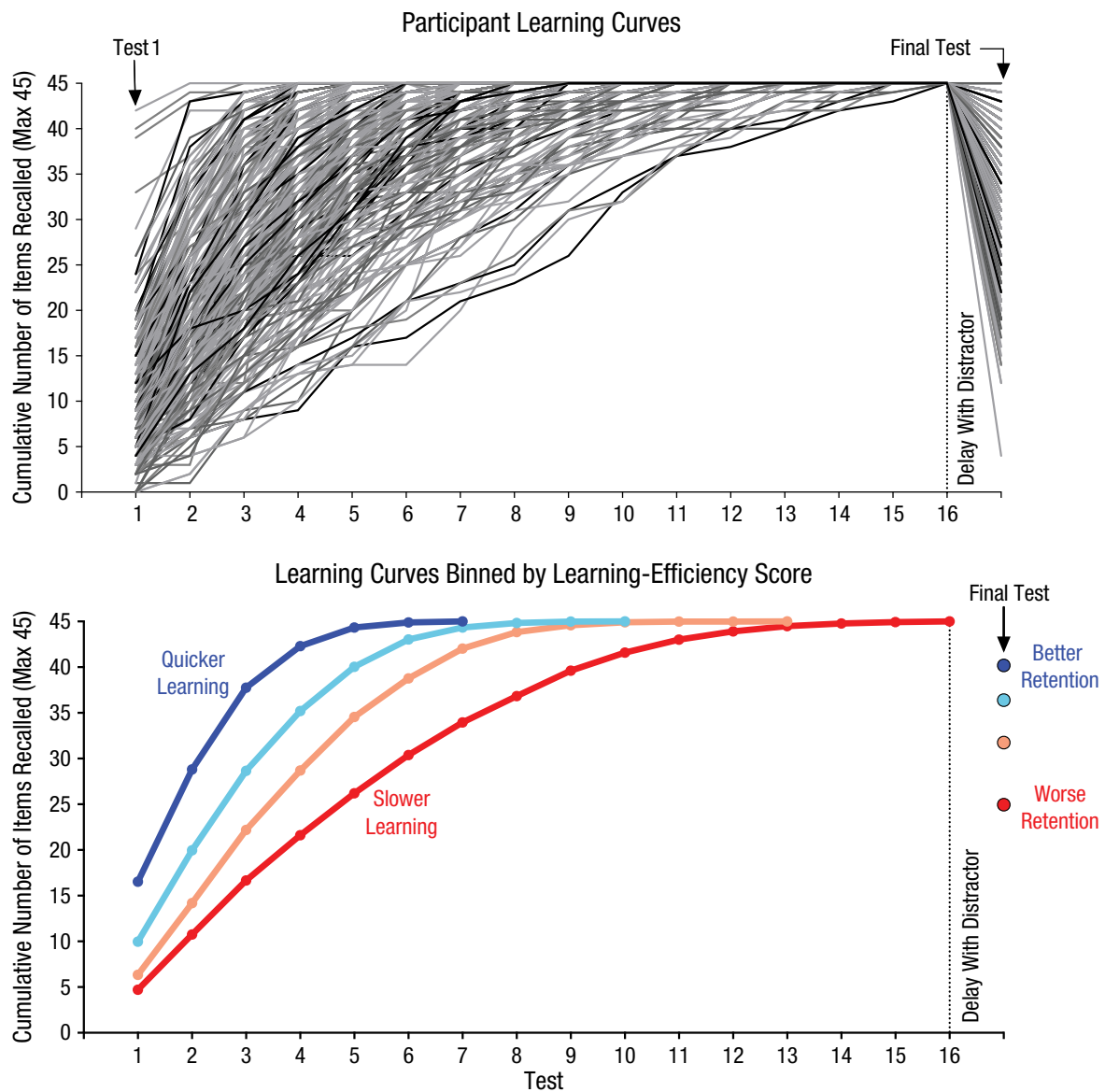
Day 2,  $r = -.48$ ,  $p < .001$ , 95% CI =  $[-.56, -.38]$ . People who performed better on the initial test also remembered more on the final test on Day 1,  $r = .26$ ,  $p < .001$ , 95% CI =  $[-.15, .37]$ , and Day 2,  $r = .18$ ,  $p = .002$ , 95% CI =  $[-.07, .29]$ . As a result of the strong intercorrelations among the dependent measures (initial test, speed of learning, and long-term retention), we refer to the task as the *learning-efficiency task* from here forward.

Performance on the learning-efficiency task significantly correlated across days for participants (Table 2), including scores on Test 1,  $r = .56$ ,  $p < .001$ , 95% CI =  $[-.47, .63]$ , tests to criterion,  $r = .68$ ,  $p < .001$ , 95% CI =  $[-.61, .73]$ , and final test,  $r = .68$ ,  $p < .001$ , 95% CI =  $[-.61, .74]$ . When the three individual measures were converted to  $z$  scores and combined into a single metric (learning-efficiency score, which is a composite of initial test, learning speed, and final retention), the correlation across days was also high,  $r = .68$ ,  $p < .001$ , 95% CI =  $[-.61, .74]$ .

**Performance across lists.** Performance overall was not significantly different between List A and List B for Test 1,  $t(280) = 0.89$ ,  $p = .375$ ,  $d = 0.05$ ; tests to criterion,  $t(280) = -1.94$ ,  $p = .054$ ,  $d = -0.12$ ; final test,  $t(280) = 0.42$ ,  $p = .673$ ,  $d = 0.03$ ; or learning-efficiency score,  $t(280) = 1.57$ ,  $p = .118$ ,  $d = 0.09$ , suggesting that List A and List B could be reasonable parallel forms for practical uses.

**Subjective ratings.** For the first day, self-reported interest ( $r_s = .18$ ,  $p = .004$ , 95% CI =  $[-.06, .29]$ ) and difficulty ( $r_s = -.43$ ,  $p < .001$ , 95% CI =  $[-.52, -.33]$ ) significantly correlated with overall performance (learning-efficiency score), whereas effort ( $r_s = -.09$ ,  $p = .118$ , 95% CI =  $[-.21, .02]$ ) did not. For the second day, only difficulty ratings significantly correlated with learning-efficiency score ( $r_s = -.43$ ,  $p < .001$ , 95% CI =  $[-.53, -.33]$ ), whereas interest level ( $r_s = .09$ ,  $p = .125$ , 95% CI =  $[-.03, .21]$ ) and effort ( $r_s = .04$ ,  $p = .518$ , 95% CI =  $[-.08, .16]$ ) did not.

**Internal consistency.** Internal consistency of test items is typically calculated using Cronbach's  $\alpha$  (Cronbach, 1951), which is the average of all possible combinations of split-half reliabilities and is the lower-bound reliability for single or repeated test administration (Novick & Lewis, 1967). Most researchers suggest that an  $\alpha$  of .70 is a reasonable lower bound (Nunnally & Bernstein, 1994). Item-level data were examined across participants for Test 1 and the final test for both Lists A and B. Test 1 and final test were the only tests examined for each administration because they were the only tests in which all participants were tested on all 45 word pairs. Cronbach's  $\alpha$  for List A was .91 for Test 1 and .94 for the final test; List B had a Cronbach's  $\alpha$  of .87 for Test 1 and .93 for the final test.



**Fig. 2.** Results from Study 1 ( $N = 281$ ). The top panel shows individual learning curves for Day 1. Participants completed Test 1, tests to criterion, and a final test on the first day and then completed the same procedure 30 hr later with a different list of 45 words. A learning-curve figure for Study 2 can be found in S. M. Nelson et al. (2016). The bottom panel shows learning curves after participants were placed into quartiles on the basis of their learning-efficiency scores. The graph represents each quartile's mean correct recall for each test block. The 25% of participants with the highest overall learning-efficiency scores (blue) learned all 45 word pairs more quickly and retained them better on average than the 25% of participants with the lowest overall learning-efficiency scores (red).

## Discussion

Study 1 used a large representative sample of participants across a variety of age ranges and education levels, and it established that across both sessions, participants who learned word pairs more quickly tended to demonstrate better retention of those word pairs on a delayed cued-recall test. Performance on the three measures composing the learning-efficiency task—Test 1, tests to criterion, and final test—and the composite

learning-efficiency score were stable across 30 hr with alternate word lists.

## Study 2

### Method

In Study 2, we examined reliability over a long time period (approximately 3 years) and examined preliminary aspects of construct validity. People who had

**Table 1.** Descriptive Statistics for Learning and Memory Metrics in Study 1 ( $N = 281$ )

Measure and day	<i>M</i>	<i>SD</i>	Minimum	Lower 25%	<i>Mdn</i>	Upper 25%	Maximum
Test 1 score							
Day 1	9.4	6.6	0	5	8	12	42
Day 2	11.1	8.2	0	5	9	15	42
Tests-to-criterion score							
Day 1	8.3	2.9	2	6	8	10	16
Day 2	7.6	2.8	2	6	7	9	16
Final-test score							
Day 1	33.4	7.9	4	29	34	40	45
Day 2	33.2	8.7	2	28	35	40	45

participated in a 100-person functional MRI study using this task (reported in S. M. Nelson et al., 2016) were contacted via e-mail and asked whether they would be willing to participate in an online follow-up study, in which we administered a highly similar task using alternate forms.

**Participants.** Participants consisted of 100 people from the greater St. Louis area recruited via Craigslist; they were compensated \$25 per hour. Eight total participants were excluded: 6 for failing to comply with directions and 2 for not completing the study. This resulted in a final sample size of 92 people.<sup>1</sup> Of these 92, 41 were female (44.6% of the sample), with a mean age of 24.7 years ( $SD = 3.7$ , range = 18–31) and a mean of 15.2 years of education ( $SD = 2.3$ , range = 10–22). None of the participants reported a history of neurological or psychiatric illness or learning disability diagnoses.

**Materials and procedure.** The first phase of this study occurred across 3 days and was similar to the procedure

for Study 1 with a few exceptions. Most notably, the first day of Study 2 took place in an MRI scanner, where participants studied and were tested on word pairs from List A (see Table S1). Pairs were presented one at a time for 3.5 s each in a random order, separated by a variable (jittered) fixation cross presented for 1.5 to 6.5 s. Pairs were displayed in all capital letters on a black background in white, 48-point Arial font. Participants completed the initial study phase, Test 1, tests to criterion, and final study session in the MRI scanner. Prior to leaving the study, participants were instructed not to rehearse or think about the word pairs they had learned before the final cued-recall test (final test) 2 days later. The final test took place in the laboratory approximately 2 days after the first study session ( $M = 43.0$  hr,  $SD = 3.0$ , range = 36.5–49.5) and was nearly identical to Test 1. However, in the final test, the Lithuanian cue was presented on screen for 8 s (during which participants typed their response); cues were shown in a random order with a 1-s ISI, and no feedback was provided.

After the final test, participants rated how difficult they found the task and how much effort they expended to

**Table 2.** Correlations Between Day 1 and Day 2 Performance on the Learning-Efficiency Task in Study 1 ( $N = 281$ )

Day and measure	1	2	3	4	5	6	7
Day 1							
1. Test 1	—						
2. Criterion	-.60***	—					
3. Final test	.26***	-.57***	—				
4. Learning-efficiency score	.77***	-.90***	.76***	—			
Day 2							
5. Test 1	<b>.56***</b>	<b>-.45***</b>	<b>.09</b>	<b>.45***</b>	—		
6. Criterion	<b>-.39***</b>	<b>.68***</b>	<b>-.35***</b>	<b>-.58***</b>	-.63***	—	
7. Final test	<b>.23***</b>	<b>-.48***</b>	<b>.68***</b>	<b>.58***</b>	.18**	-.48***	—
8. Learning-efficiency score	<b>.50***</b>	<b>-.68***</b>	<b>.47***</b>	<b>.68***</b>	.77***	-.89***	.70***

Note: Boldface indicates alternate-forms test-retest reliability across a 30-hr delay. Criterion is the number of tests required for participants to reach criterion.

\*\* $p < .01$ . \*\*\* $p < .001$ .

complete it, as well as answered questions regarding what sort of strategies they used to learn the paired associates. They were then tested on a wide range of cognitive and personality batteries to provide a preliminary investigation into construct validity and ceiling effects.

**Cognitive and personality batteries.** All 92 participants completed the cognitive and personality batteries. Cognitive batteries included the fourth edition of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008), second edition of the Wechsler Adult Scale of Intelligence (WASI-II; Wechsler, 2011), Trail Making Test (Parts A and B; Reitan, 1958), CVLT-II (Delis et al., 2000), computation span task (Salthouse & Babcock, 1991), consonant/vowel odd/even switching (CVOE) task (Duchek et al., 2009), and the Stroop task (Hutchison, Balota, & Duchek, 2010; Stroop, 1935). Personality inventories, including the Zimbardo Time Perspective Inventory (Zimbardo & Boyd, 1999), Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984), Narcissistic Personality Inventory (Ames, Rose, & Anderson, 2006), Ten-Item Personality Inventory (Gosling, Rentfrow, & Swann, 2003), and Vividness of Visual Imagery Questionnaire (Marks, 1973), were also administered to participants at the end of the session. A full description of these batteries and how they were administered can be found in the Supplemental Material available online.

**Three-year follow-up.** To assess long-term reliability of the learning-efficiency task, we conducted a 3-year follow-up study ( $M = 2.8$  years,  $SD = 0.5$ , range = 2.2–4.2) online. Of the original 92 participants, 69 had agreed to be contacted again for future studies; of these 69 people who were contacted for the 3-year follow-up study, 46 agreed to participate for \$60 compensation. Participants ( $n = 46$ ) completed the same task as before with 45 different norm-matched Lithuanian-English words from List B (see Table S2). Sources of measurement error were time (3-year delay), list used (List A, List B), and testing

environment (inside an MRI scanner for Test 1 and tests to criterion for Time 1, inside the lab for final test for Time 1, online in whatever environment was chosen by the participant for both sessions of the 3-year follow-up at Time 2). Lithuanian cues were presented for 3.5 s during study trials with an ISI of 1.5 s. For testing trials (Test 1, tests to criterion, final test), Lithuanian cues were presented for 5 s with an ISI of 1 s. Each testing trial was separated by 30 s of arithmetic problems with the exception of the final test, which took place approximately 2 to 3 days later ( $M = 60.3$  hr,  $SD = 27.3$ , range = 37.8–164.8). The entire task (including informed consent and directions) took an average of 43.4 min to complete ( $SD = 10.1$ , range = 29.7–76.0).

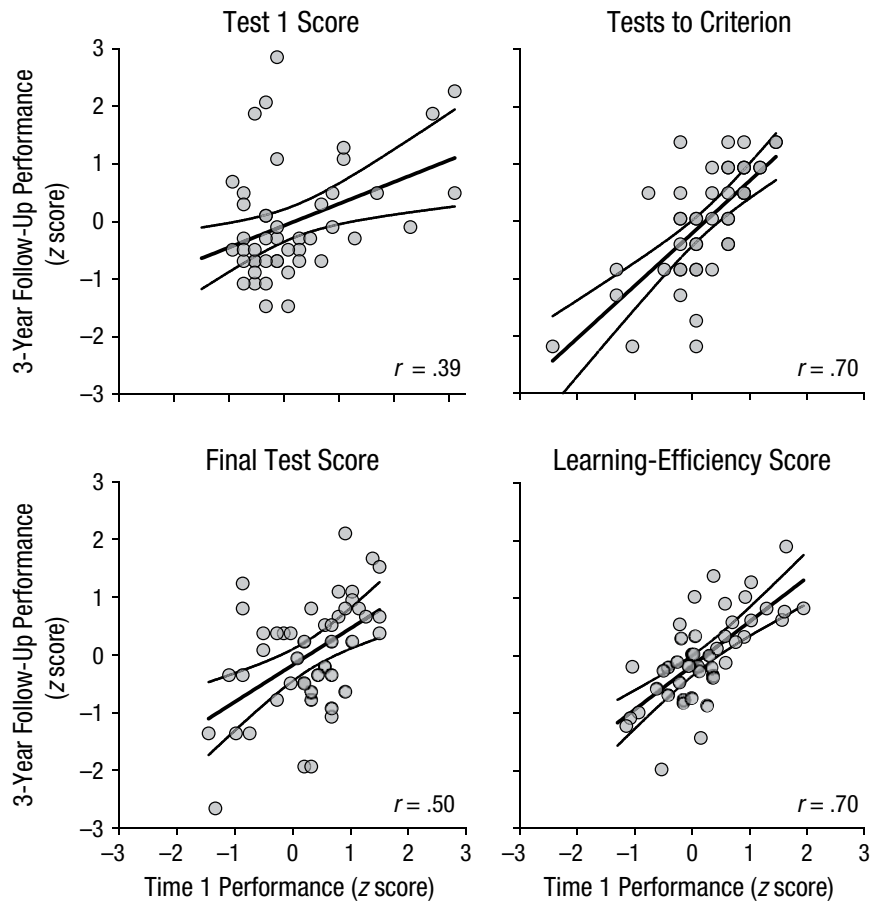
**Analysis.** Comparisons for scores at Time 1 and Time 2 were conducted using a paired-samples  $t$  test (two-tailed), unless the data were not normally distributed, in which case a nonparametric Wilcoxon signed-rank test was used. Effect sizes were calculated using Cohen's  $d$  for repeated measures (Cohen, 2009).

## Results

**Time 1 performance.** S. M. Nelson et al. (2016) reported behavioral data for 86 of the 92 participants (93.5% of the sample) from Study 2; the following sections report results for the complete 92-participant sample as well as follow-up results. On Test 1, after studying the items once, participants recalled an average of 6.7 English words when prompted with the Lithuanian cue ( $SD = 5.0$ ). It took an average of 9.3 tests for subjects to learn the 45 word pairs ( $SD = 3.6$ ), and on a final test, participants recalled 30.3 words on average ( $SD = 8.5$ ). Participants had an average IQ of 115.5 ( $SD = 14.0$ , range = 87–146) as per WASI-II Full Scale IQ-2 (FSIQ-2) composite scores, placing them in the 77th percentile ( $SD = 22.7$ , range = 19–100) in terms of mean intelligence. Additional descriptive statistics for task performance are in Table 3.

**Table 3.** Descriptive Statistics for Study 2 Task Performance for Time 1 ( $N = 92$ ) and the Follow-Up Sample at Time 2 ( $n = 46$ ) Approximately 3 Years Later

Measure and time	$M$	$SD$	Minimum	Lower 25%	$Mdn$	Upper 25%	Maximum
Test 1 score							
Time 1	6.7	5.0	0	3	5	9	22
Time 2	8.5	5.1	1	5	7	11	23
Tests-to-criterion score							
Time 1	9.3	3.6	4	7	9	11	22
Time 2	8.1	2.2	5	6	8	10	13
Final-test score							
Time 1	30.3	8.5	10	24	32	37	45
Time 2	23.4	6.9	5	19	25	29	38



**Fig. 3.** Results from Study 2: scatterplots showing reliability of measures on the learning-efficiency task across 3 years ( $n = 46$ ). Solid lines indicate best-fitting regressions, and dashed lines indicate 95% confidence intervals. Performance was stable across each metric, including similar performance for an initial cued-recall test (Test 1); the number of tests required to learn all 45 word pairs (tests to criterion), which represents learning speed; and a delayed cued-recall test (final test) approximately 2 to 3 days after the final study session, which represents long-term retention. The  $z$  scores for each of the three measures (Test 1, inverse of tests to criterion, and final test) were averaged to create an overall learning-efficiency score (bottom right) for each participant.

**Three-year reliability.** The 3-year follow-up study revealed stable performance for the various submeasures included in the learning-efficiency task (refer to Fig. 3 and the boldface values in Table 4). Correlations between performance at Time 1 and Time 2 for two different word lists were significant for Test 1,  $r = .39$ ,  $p = .008$ , 95% CI = [.11, .61]; tests to criterion,  $r = .70$ ,  $p < .001$ , 95% CI = [.52, .83]; final test,  $r = .50$ ,  $p < .001$ , 95% CI = [.25, .69]; and learning-efficiency score,  $r = .70$ ,  $p < .001$ , 95% CI = [.51, .82]. Many of the submeasures of the learning-efficiency task remained significantly intercorrelated across the 3-year span. For example, learning rate at Time 1 remained negatively correlated with final test performance at Time 2.

Test 1 scores,  $z = 0.74$ ,  $p = .459$ ,  $d = -0.17$ , 95% CI = [-2.50, 1.00], and tests to criterion,  $z = -0.85$ ,  $p = .398$ ,  $d = 0.16$ , 95% CI = [-0.50, 1.00], did not significantly differ across the 3-year delay, and the difference

scores had trivial effect sizes (Cohen, 2009). Final-test scores did significantly differ between Time 1 and Time 2,  $t(45) = 9.22$ ,  $p < .001$ ,  $d = 1.36$ , 95% CI = [7.22, 11.26]. This difference could in part be due to significant differences in the delay between the final study session and final test at Time 1 ( $M = 42.6$  hr,  $SD = 2.9$  for the  $n = 46$  subset) and Time 2 ( $M = 60.3$  hr,  $SD = 27.3$ ),  $t(45) = -4.28$ ,  $p < .001$ ,  $d = -0.63$ , 95% CI for the mean difference = [-26.02, -9.35].

**Strategy usage.** Participants at Time 1 were asked how frequently they employed specific strategies to learn the material on a scale from 1 to 5 (1 = *never*, 2 = *rarely*, 3 = *sometimes*, 4 = *usually*, 5 = *always*). The strategies were (a) “tried to come up with a single word to associate with the pairs” ( $M = 3.4$ ,  $SD = 1.1$ ,  $Mdn = 4$ , mode = 4), (b) “constructed sentences that described what you physically



**Table 4.** Correlations Between Time 1 ( $N = 92$ ) and Time 2 ( $n = 46$ ) Components of the Learning-Efficiency Task in Study 2

Time and measure	1	2	3	4	5	6	7
Time 1							
1. Test 1	—						
2. Criterion	-.63***	—					
3. Final test	.57***	-.71***	—				
4. Learning-efficiency score	.84***	-.90***	.88***	—			
Time 2							
5. Test 1	<b>.39**</b>	<b>-.36*</b>	<b>.28</b>	<b>.41**</b>	—		
6. Criterion	<b>-.51***</b>	<b>.70***</b>	<b>-.58***</b>	<b>-.70***</b>	-.60***	—	
7. Final test	<b>.46**</b>	<b>-.40**</b>	<b>.50***</b>	<b>.54***</b>	.26	-.46**	—
8. Learning-efficiency score	<b>.57***</b>	<b>-.61***</b>	<b>.57***</b>	<b>.70***</b>	.78***	-.87***	.73***

Note: Boldface indicates alternate-forms test-retest reliability across a 3-year span. Criterion is the number of tests required for participants to reach criterion.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

saw" ( $M = 2.4$ ,  $SD = 1.1$ ,  $Mdn = 2$ , mode = 3), and (c) "constructed stories about the pairs" ( $M = 2.3$ ,  $SD = 1.2$ ,  $Mdn = 2$ , mode = 1). None of the strategy frequency ratings correlated with overall performance on the task.

**Subjective ratings.** After completing the learning-efficiency task at Time 2, participants in the follow-up condition provided ratings regarding how difficult they found the task and how much effort they put forth when completing it. Ratings could range from 1 to 100, with 1 being the lowest. Neither self-reported difficulty ( $r_s = -.21$ ,  $p = .154$ , 95% CI =  $[-.48, .08]$ ) nor effort ( $r_s = -.12$ ,  $p = .410$ , 95% CI =  $[-.40, .17]$ ) ratings significantly correlated with overall learning-efficiency score.

**Ceiling effects.** Ceiling effects were present in our healthy young sample on some of the neuropsychological batteries. For instance, approximately one third or more of our sample was at or near ceiling (i.e., at maximum score or maximum score minus 1; Uttil, 2005) on the tests of long-term memory on the CVLT-II, including for short-delay free recall (29.3% of participants at or near ceiling), short-delay cued recall (31.5%), long-delay free recall (32.6%), and long-delay cued recall (38.0%). This is consistent with findings from other studies using neuropsychological tests with healthy young adults (Davis et al., 2003; Travis et al., 2014; Uttil, 2005). For comparison, in the entire combined sample for Studies 1 and 2 ( $N = 610$ ), no participants were at or near ceiling (44 or 45 items) on the learning-efficiency task for Test 1, and only 28 participants ( $< 0.1\%$  of the sample) were at or near ceiling for the final test (27 of which came from Study 1, which had only a 5-min delay).

**Internal consistency.** Cronbach's  $\alpha$  was computed by finding the proportion of participants that correctly

recalled each word pair on Test 1 and the final test for both List A and List B (see Tables S1 and S2), as was done for Study 1. For List A, Test 1 had a Cronbach's  $\alpha$  of .80, and the final-test administration had a Cronbach's  $\alpha$  of .89. For List B, which had 46 participants, Cronbach's  $\alpha$  was .77 for Test 1 and .82 for the final test. These results are similar to those found in Study 1.

**Construct validity.** Construct validation consists of showing the relationship between some measure of interest and other established measures and tests (Cronbach & Meehl, 1955). Two types of construct validation exist: when the measure of interest correlates strongly with other measures supposedly measuring the same construct (convergent evidence) and when it does not correlate strongly with other measures supposedly measuring unrelated constructs (discriminant evidence; Nunnally & Bernstein, 1994). We report here a preliminary step in establishing construct validity.

As a result of the large number of cognitive measures in this study as well as issues with multicollinearity (and ceiling effects in the case of standardized cognitive measures), two composite scores were created. These composite scores included processing speed (S. M. Nelson et al., 2016) and a general-memory-ability factor from the CVLT-II manual (Delis et al., 2000), each of which was computed by averaging the  $z$  scores for each participant across relevant tasks. The processing-speed composite was calculated from switch and nonswitch mean reaction times from the Stroop and CVOE tasks (reverse-scored), the time required to complete Trail Making Tests A and B (reverse-scored), and the Processing Speed Index from the WAIS-IV. A higher processing-speed composite score represented faster reaction times and better overall processing speed. A general-memory-ability component was created from  $z$  scores for

**Table 5.** Correlation Matrix Demonstrating Convergent Evidence Between Measures on the Learning-Efficiency Task and Several Cognitive Batteries in Study 2

Measure	1	2	3	4	5	6
1. Processing speed (Stroop, CVOE, TMT, WAIS-IV)	—					
2. CVLT-II memory	.22*	—				
3. WASI-II FSIQ-2	.45***	.44***	—			
4. Computation Span	.09	.22*	.12	—		
5. CVOE switch accuracy	.33**	.10	.14	.13	—	
6. Stroop switch accuracy	.44***	.16	.30**	.25*	.38***	—
7. Learning-efficiency score	.44***	.43***	.43***	.13	.23*	.25*
8. Test 1	.37***	.33**	.40***	.08	.12	.15
9. Tests to criterion	-.48***	-.46***	-.46***	-.11	-.29**	-.32**
10. Final test	.30**	.33**	.27*	.16	.19	.20

Note: CVOE = consonant/vowel odd/even switching; TMT = Trail Making Test; WAIS-IV = Wechsler Adult Intelligence Scale–Fourth Edition; CVLT-II = California Verbal Learning Test–Second Edition; WASI-II = Wechsler Adult Scale of Intelligence–Second Edition; FSIQ-2 = Full Scale IQ-2.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

CVLT-II standard scores, including Trials 1 to 5 total score, short-delay free and cued recall, long-delay free and cued recall, and recognition discriminability. Additionally, FSIQ-2 composite scores from the WASI-II—computed by summing the  $t$  scores from the Vocabulary and Matrix Reasoning subtests—were used as an index of intellectual ability. A correlation matrix for these composites, as well as the other relevant cognitive variables included in the current study, is depicted in Table 5 as a preliminary demonstration of convergent-validity evidence. Correlation matrices for the processing speed and memory subcomposite measures, as well as factor loadings, can be found in Tables S3 and S4 in the Supplemental Material.

In general, learning-efficiency scores positively correlated with measures of learning and memory from the CVLT-II, with general intellectual ability from the WASI-II, with processing-speed components from the WAIS-IV (such that better performance on the learning-efficiency task coincided with faster response times), and with accuracy measures from switch tasks such as the Stroop and CVOE. Learning-efficiency scores tended to negatively correlate with measures of reaction time

from the Stroop, CVOE, and Trail Making Test, implying that quicker processing speed was related to better learning efficiency.

In regard to discriminant evidence, none of the five personality measures included in the study were significantly correlated with Test 1, tests to criterion, final test, or overall learning-efficiency score (mean  $r_s = .003$ ,  $p_s > .168$ , mean 95% CI =  $[-.20, .21]$ ). This outcome should be expected, as significant correlations could reveal biases in the task or protocol (e.g., if more extraverted participants scored more highly on the metrics).

**Relative contributions of cognitive abilities to efficient learning.** A simultaneous regression model was conducted for overall learning-efficiency scores on the basis of relevant cognitive batteries from Study 2, with the goal of assessing the relative contributions of processing speed, memory ability (as assessed by the CVLT-II), and intellectual ability (as assessed by the WASI-II FSIQ-2) to overall learning efficiency (Table 6). Because these predictor variables were correlated with one another, we calculated each variable's relative weight (Johnson, 2000) to

**Table 6.** Regression Results and Relative Weights for Composite Cognitive Measures in Accounting for Learning Efficiency in Study 2

Measure	Total $R^2$	Adjusted $R^2$	$r$	Raw relative weight	Rescaled relative weight (%)
Learning-efficiency score	.33	.31			
Processing speed			.44	.13	38.1
CVLT-II memory			.43	.12	35.4
WASI-II FSIQ-2			.43	.09	26.5

Note: Raw relative weights sum to the value of  $R^2$  (approximate because of rounding). Rescaled relative weights refer to each measure's contribution to  $R^2$  and thus sum to 100%. CVLT-II = California Verbal Learning Test–Second Edition; WASI-II = Wechsler Adult Scale of Intelligence–Second Edition; FSIQ-2 = Full Scale IQ-2.

determine its individual contribution to overall learning-efficiency scores.

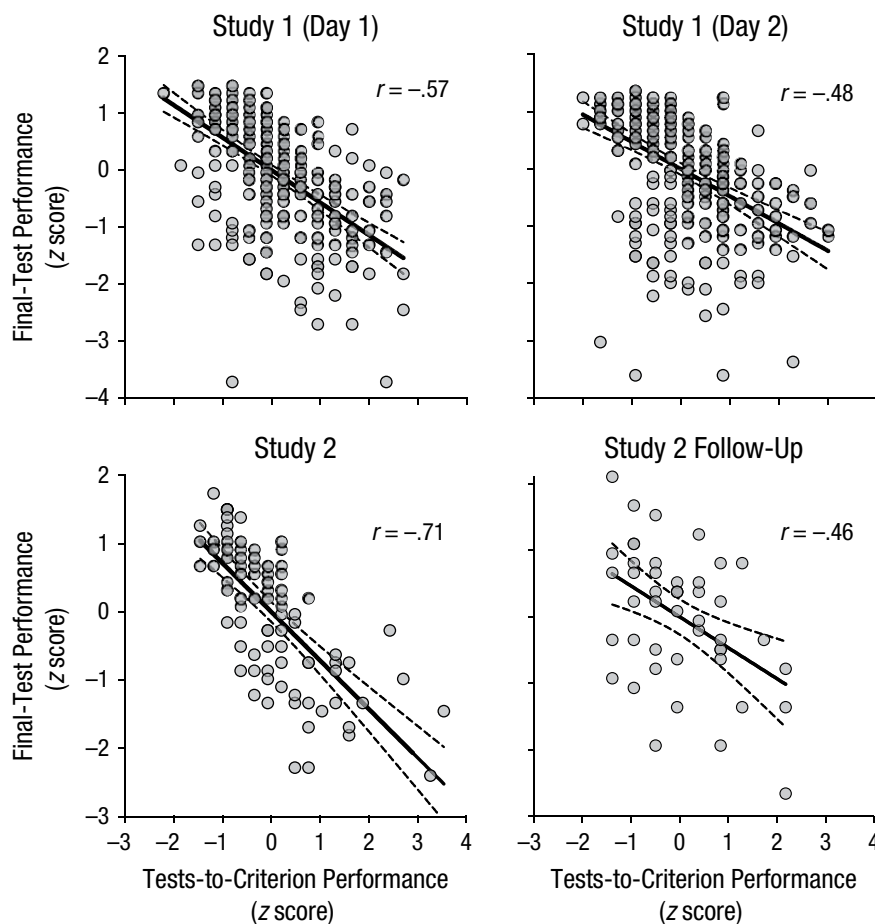
Processing speed and CVLT-II general memory ability accounted for similar percentages of the explained variance (38.1% and 35.4% of  $R^2$ , respectively), with WASI-II FSIQ-2 scores accounting for approximately a quarter of the explained variance (26.5% of  $R^2$ ). These three variables combined accounted for approximately one third ( $R^2 = .33$ , adjusted  $R^2 = .31$ ) of the variance in learning-efficiency scores,  $F(3, 88) = 14.38$ ,  $p < .001$ ; these variables also significantly predicted follow-up learning-efficiency scores 3 years later,  $F(3, 42) = 3.82$ ,  $p = .017$ , accounting for more than a fifth of the observed variance in learning-efficiency scores for the follow-up sample ( $R^2 = .21$ , adjusted  $R^2 = .16$ ).

The individual measures from the learning-efficiency task—Test 1 score, tests to criterion, and final-test score—were also modeled via simple linear regression using the variables from Table 6 as individual predictors. IQ scores as measured by the WASI-II Composite FSIQ-2 accounted for the most variance in Test 1 scores,  $R^2 = .16$ , adjusted  $R^2 = .15$ ,  $F(1, 90) = 16.60$ ,  $p < .001$ ,

and significantly predicted Test 1 performance in the 3-year follow-up,  $R^2 = .12$ , adjusted  $R^2 = .10$ ,  $F(1, 44) = 6.06$ ,  $p = .018$ . The composite processing-speed measure accounted for the most variance in the number of tests required to reach criterion,  $R^2 = .23$ , adjusted  $R^2 = .22$ ,  $F(1, 90) = 26.68$ ,  $p < .001$ , and significantly predicted tests to criterion for the 3-year follow-up,  $R^2 = .11$ , adjusted  $R^2 = .09$ ,  $F(1, 44) = 5.55$ ,  $p = .023$ . Finally, the general-memory composite from CVLT-II measures accounted for most of the variance in final-test scores for participants at Time 1,  $R^2 = .11$ , adjusted  $R^2 = .10$ ,  $F(1, 90) = 10.86$ ,  $p = .001$ , but did not significantly predict final-test performance 3 years later,  $R^2 = .04$ , adjusted  $R^2 = .02$ ,  $F(1, 44) = 1.87$ ,  $p = .178$ .

### ***Learning speed predicts long-term retention***

The single best predictor of final-test scores was tests to criterion—that is, long-term retention scores were best predicted by learning speed (Fig. 4). This was true for all samples across all time periods and both word



**Fig. 4.** Results from Studies 1 and 2: scatterplots showing the relation between tests-to-criterion score and final-test score. Solid lines indicate best-fitting regressions, and dashed lines indicate 95% confidence intervals. Participants who learned 45 word pairs in fewer tests to criterion tended to recall more word pairs on a delayed cued-recall test (higher final-test scores) after delays of 5 min (Study 1) and approximately 48 to 60 hr (Study 2).

lists in both studies; the correlations between tests to criterion and final-test scores ranged from  $r = -.46$  to  $-.71$ , with a mean correlation of  $r = -.56$ . Learning speed at Time 1 even predicted retention performance 3 years later with a different list of words,  $R^2 = .16$ , adjusted  $R^2 = .14$ ,  $F(1, 44) = 8.52$ ,  $p = .006$ . In each case, the quicker learner retained more, even though reaching criterion in fewer tests meant less exposure to—and fewer opportunities to study and be tested on—the to-be-remembered information.

## Discussion

In summary, Study 2 again revealed a tendency for quicker learners to be better retainers, even after a longer delay. Participants who were efficient learners at Time 1 also tended to be efficient learners 3 years later, even in vastly different study and testing environments, suggesting that this learning-efficiency construct is robust and relatively stable. In attempting to better understand what processes might underlie more efficient learning, preliminary aspects of construct validity were examined. Better performance on the learning-efficiency task coincided with better performance on other cognitive measures, including those that measured processing speed (Stroop, CVOE, Trail Making Test, WAIS-IV), switch accuracy (Stroop, CVOE), memory (CVLT-II), and intelligence (WASI-II).

## General Discussion

The present studies demonstrate large variability in learning performance across people; this variability existed for an initial test, the number of tests to reach criterion (learning speed), and a final test (long-term retention). Further, these measures were intercorrelated across people, indicating that quicker learning is often more durable learning; people who learned material in less time displayed better retention across delays ranging from minutes (Study 1) to days (Study 2). We refer to this relation between learning rate and retention as *learning efficiency*. Finally, participants' learning efficiency was highly stable across days (Study 1) and years (Study 2). Hence, this approach to measuring learning efficiency may be useful in settings where individual differences in learning within a healthy young adult population may be of interest or in cases in which one wishes to eliminate ceiling effects produced by popular neuropsychological measures of learning and memory within a healthy population.

## Learning-efficiency mechanisms

What characteristics set an efficient learner apart from an inefficient learner? More efficient learners tended to have

faster processing speed, higher intelligence scores, and better memory performance on various cognitive tasks. One likely mechanism underlying efficient learning is attentional control. As mentioned in the introduction, better learners tended to show greater deactivation of the default mode network at encoding (S. M. Nelson et al., 2016); although the interpretation of default mode network deactivation is still an active topic of research (Raichle, 2015), the pattern could be interpreted as suggesting that more efficient learners better allocate attention while learning novel material. People who are better able to focus their attention on task-relevant information have been shown to demonstrate less susceptibility to proactive interference and less forgetting in memory tasks, as well as more rapid, refined memory search at retrieval (Shipstead, Redick, Hicks, & Engle, 2012; Unsworth & Spillers, 2010). Attentional control is closely tied to working memory capacity, the latter of which can be characterized as the ability to allocate attention in order to engage in active maintenance of task-relevant information (Shipstead, Harrison, & Engle, 2016; Unsworth & Spillers, 2010), and so working memory capacity presumably plays a role in efficient learning as well.

Another possibility is that more efficient learners applied more effective strategies, such as the keyword method, to better encode the material. Or perhaps they applied the same strategies as less efficient learners but either adopted them more quickly in the learning sequence or simply implemented them more effectively (e.g., had more effective keywords to link the two items of each pair together). Although the materials were designed to be less amenable to strategy use than other common materials, such as categorized word lists, it seems reasonable to hypothesize that learners adopted some kind of consistent strategy to master the material, and further work may elucidate strategy differences underlying performance differences. Therefore, although the poststudy survey showed no relation between strategy use and performance, a more thorough investigation into strategies might reveal some intriguing patterns.

Finally, a remaining question concerns how much variance in learning efficiency is accounted for by attentional control, working memory capacity, and strategy use (e.g., which strategy and how well it is applied). Is it the case that differences in learning ability can be largely accounted for by attention, working memory, and learner sophistication, or is any substantial independent variance left over that points to a raw learning difference across people (perhaps mediated by other individual differences variables)?

## Desirable difficulties?

At a surface level, our learning-efficiency results may seem inconsistent with the desirable difficulties

framework. A “desirable difficulty” (Bjork, 1994) refers to a condition during acquisition, such as spacing or retrieval practice, that can slow performance and make learning more difficult yet optimize long-term retention. In our case, the opposite pattern was observed: Quicker learning resulted in better retention. A similar finding was reported by Woodworth (1914) at the level of individual items; items learned more quickly (i.e., the easy items) tended to be retained better. Being a slow learner (or encountering a difficult item) might be classified as an “undesirable difficulty,” although the discussion of desirable difficulties is usually aimed at the condition level rather than as person-level or item-level difficulties; therefore, we suggest that the inconsistency may be more apparent than real. A clear classification of which learning challenges are desirable remains an open question.

### ***Future directions***

One fundamental question is whether learning efficiency is a domain-general or a domain-specific phenomenon. Is the ability to both quickly acquire and successfully retain information dependent on particular materials, such as verbal paired associates, or is efficient learning a skill that is generalizable across a range of materials, such as visuospatial stimuli or prose passages? Earlier studies have tried to examine a similar question (Gillette, 1936; Lyon, 1917), although they failed to control for the degree of prior knowledge of the materials or failed to equate the degree of learning for the to-be-remembered information.

It is also an open question of whether the retentive advantage for quicker learners persists for longer delay intervals, such as weeks or months. If the suggested mechanisms underlying efficient learning do play a significant role, it seems likely that the learning and retention advantages would remain at longer delays.

Finally, and perhaps most importantly, understanding individual differences in how learning and memory interact has broad implications for applied domains, such as educational and clinical settings. Consider the classroom, for instance, where it would be in a student's interest to acquire information in less time and fewer study sessions but retain it well at various delays, such as an upcoming test, an end-of-semester final exam, or even over a lifetime. By further characterizing learning efficiency and better dissecting the processes behind it, it may be possible to teach students to become more efficient learners or better gauge how they might perform on an exam on the basis of the time it takes for them to learn to some criterion. In the clinic, how does this interaction between learning rate and retention change with the progression of memory deficits

associated with aging and disease, such as Alzheimer's disease? Is a test that incorporates both learning speed and retention a more sensitive measure for the detection of deficits than single-trial recall or recognition? By focusing on individual differences in how learning and memory processes interact, we may better understand how each process operates and progresses over time in critical real-world settings.

### ***Implications***

When studying individual differences in memory, it is important to have (a) a reliable task that can produce enough variability in memory performance and (b) some variable or variables that can explain or account for the observed variability at the individual level. The present studies introduced a reliable task that can be used broadly to study individual differences in learning and memory for healthy adults and demonstrated that one explanation for the observed variability in retention is the speed with which individuals learned the to-be-remembered information. Rather than treat individual variability in performance as an error bar in a mean plot, researchers should attempt to better understand this variability and its potential sources; doing so may allow us to generate new theories about how memory operates within individuals. As Melton (1967) argued,

The sooner our experiments and our theory on human memory and human learning consider the differences between individuals . . . the sooner we will have theories and experiments that have some substantial probability of reflecting the fundamental characteristics of those processes. (pp. 249–250)


### **Action Editor**

D. Stephen Lindsay served as action editor for this article.

### **Author Contributions**

S. M. Nelson and K. B. McDermott designed the studies. S. M. Nelson, A. K. Fishell, N. K. Savalia, J. J. Berg, and C. L. Zerr collected the data. J. J. Berg and C. L. Zerr analyzed the data. C. L. Zerr drafted the manuscript, and all authors contributed to revisions. All the authors approved the final manuscript for submission.

### **ORCID iD**

Christopher L. Zerr  <https://orcid.org/0000-0001-9196-5075>

### **Acknowledgments**

We thank Roddy Roediger, Adrian Gilmore, and Rob Tigner for helpful comments and Fan Zou for help with data collection.

## Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

## Funding

This research was funded in part by grants from the McDonnell Center for Systems Neuroscience at Washington University School of Medicine in St. Louis and Dart NeuroScience.

## Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618772540>

## Open Practices



All data and materials have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/rf28u/>. The design and analysis plans for the studies were not preregistered. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618772540>. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

## Note

1. Data from 86 of the 92 participants in this sample were previously reported by S. M. Nelson et al. (2016). They had excluded data for an additional 6 participants resulting from excessive motion within the MRI scanner, which was not an exclusion criterion for the behavioral data. Thus the current study used those additional 6 subjects, resulting in a final sample size of 92.

## References

- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, 40, 440–450. doi:10.1016/j.jrp.2005.03.002
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48, 306–307. doi:10.1207/s15327752jpa4803\_13
- Cohen, J. (2009). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Psychology Press. doi:10.4324/9780203771587
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. doi:10.1007/bf02310555
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. doi:10.1037/h0040957
- Davis, H. P., Small, S. A., Stern, Y., Mayeux, R., Feldstein, S. N., & Keller, F. R. (2003). Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex*, 39, 1063–1091. doi:10.1016/s0010-9452(08)70878-5
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *The California Verbal Learning Test* (2nd ed.). San Antonio, TX: Psychological Corp.
- Duchek, J. M., Balota, D. A., Tse, C. S., Holtzman, D. M., Fagan, A. M., & Goate, A. M. (2009). The utility of intra-individual variability in selective attention tasks as an early marker for Alzheimer's disease. *Neuropsychology*, 23, 746–758. doi:10.1037/a0016583
- Gillette, A. L. (1936). Learning and retention: A comparison of three experimental procedures. *Archives of Psychology*, 198, 5–56.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504–528. doi:10.1016/s0092-6566(03)00046-1
- Grimaldi, P. J., Pyc, M. A., & Rawson, K. A. (2010). Normative multitrial recall performance, metacognitive judgments, and retrieval latencies for Lithuanian-English paired associates. *Behavior Research Methods*, 42, 634–642. doi:10.3758/brm.42.3.634
- Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, 2, 685–694. doi:10.1038/35094500
- Hutchison, K. A., Balota, D. A., & Duchek, J. M. (2010). The utility of Stroop task switching as a marker for early-stage Alzheimer's disease. *Psychology and Aging*, 25, 545–559. doi:10.1037/a0018498
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35, 1–19. doi:10.1207/s15327906mbr3501\_1
- Lyon, D. O. (1917). *Memory and the learning process*. Baltimore, MD: Warwick & York.
- Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64, 17–24. doi:10.1111/j.2044-8295.1973.tb01322.x
- Melton, A. W. (1967). Individual differences and theoretical process variables: General comments on the conference. In R. M. Gagné (Ed.), *Learning and individual differences* (pp. 238–252). Columbus, OH: Merrill.
- Nelson, S. M., Savalia, N. K., Fishell, A. K., Gilmore, A. W., Zou, F., Balota, D. A., & McDermott, K. B. (2016). Default mode network activity predicts early memory decline in healthy young adults aged 18–31. *Cerebral Cortex*, 26, 3379–3389. doi:10.1093/cercor/bhv165
- Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, 2, 325–335. doi:10.1080/09658219408258951

- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, *32*, 1–13. doi:10.1007/bf02289400
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*, 331–347. doi:10.1016/0749-596x(91)90040-q
- Raichle, M. R. (2015). The brain's default mode network. *Annual Review of Neuroscience*, *38*, 433–437. doi:10.1146/annurev-neuro-071013-014030
- Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276. doi:10.2466/pms.8.7.271-276
- Rey, A. (1964). *L'examen clinique en psychologie* [Clinical tests in psychology]. Paris, France: Presses Universitaires de France.
- Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776. doi:10.1037/0012-1649.27.5.763
- Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, *11*, 771–799. doi:10.1037/e528942014-017
- Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory*, *20*, 608–628. doi:10.1080/09658211.2012.691519
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. doi:10.1037/0096-3445.121.1.15
- Travis, S. G., Huang, Y., Fujiwara, E., Radomski, A., Olsen, F., Carter, R., . . . Malykhin, N. V. (2014). High field structural MRI reveals specific episodic memory correlates in the subfields of the hippocampus. *Neuropsychologia*, *53*, 233–245. doi:10.1016/j.neuropsychologia.2013.11.016
- Tulving, E. (1983). *Elements of episodic memory*. New York, NY: Oxford University Press.
- Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392–406. doi:10.1016/j.jml.2010.02.001
- Uttl, B. (2005). Measurement of individual differences: Lessons from memory assessment in research and clinical practice. *Psychological Science*, *16*, 460–467. doi:10.1111/j.0956-7976.2005.01557.x
- Uttl, B., Graf, P., & Richter, L. K. (2002). Verbal Paired Associates tests limits on validity and reliability. *Archives of Clinical Neuropsychology*, *17*, 567–581. doi:10.1093/arclin/17.6.567
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson.
- Wechsler, D. (2009). *Wechsler Memory Scale—Fourth Edition (WMS-IV) technical and interpretive manual*. San Antonio, TX: Pearson.
- Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence* (2nd ed.). San Antonio, TX: Pearson.
- Woodworth, R. S. (1914). A contribution to the question of “quick learning, quick forgetting.” *Psychological Bulletin*, *11*, 58–59.
- Zimbardo, P. G., & Boyd, J. N. (1999). Putting time in perspective: A valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, *77*, 1271–1288. doi:10.1007/978-3-319-07368-2\_2