Winter 12-19-2017

The Domain-Generality and Durability of Efficient Learning

Christopher Zerr
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological & Brain Sciences

The Domain-Generality and Durability of Efficient Learning

by

Christopher L. Zerr

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

December 2017
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am extremely grateful to my mentor, Kathleen McDermott, for her advice and encouragement throughout all phases of this project. I am especially appreciative of her dedication to her students, and to our weekly meetings that have helped me grow as both a student and researcher. I would like to thank the members of my committee, Ian Dobbins and Mark McDaniel, who have asked thoughtful questions and provided helpful suggestions when presenting these ideas. Additionally, I would like to thank the other members of my lab, including Adrian Gilmore, Jeff Berg, Nate Anderson, Hank Chen, Ruthie Shaffer, Thomas Spaventa, Justin Vincent, and Hannah Becker, who have contributed in various ways and have ultimately helped shape how I think about the ideas in this project.

This material is based upon work supported by a grant from Dart NeuroScience, LLC (awarded to Kathleen McDermott), and by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1745038.

Christopher L. Zerr

Washington University in St. Louis

December 2017

ABSTRACT OF THESIS

The Domain-Generality and Durability of Efficient Learning

by

Christopher L. Zerr

Master of Arts in Psychological & Brain Sciences

Washington University in St. Louis, 2017

Professor Kathleen B. McDermott, Chair

People differ in how quickly they learn information and how long they remember it, and a common finding in the literature is that a quicker rate of learning coincides with better retention for the learned material. Zerr and colleagues (2017) termed the relation between learning rate and retention as *learning efficiency*, with more efficient learning representing both a faster acquisition rate and better memory performance after a delay. Zerr et al. also demonstrated in separate experiments that how efficiently someone learns is stable across a range of days and years. The current thesis includes two experiments addressing additional questions regarding efficient learning. Experiment 1 (N = 119) examined whether efficient learning is generalizable across stimuli, including Lithuanian-English (verbal-verbal) and Chinese-English (visuospatial-verbal) paired associates. Experiment 2 (N = 190) assessed whether faster learners demonstrate better retention at a longer delay of 1 week, and also preliminarily examined whether faster and slower learners demonstrate differential rates of forgetting. These experiments demonstrated that learning efficiency is generalizable across stimuli and that faster learners maintain a retentive advantage at longer delays of 1 week.

Chapter 1: Introduction

1.1 Individual Differences in Learning Rate and Retention

"*By and large, individual differences in learning are reflected in individual differences in retention*" (p. 375, McGeoch & Irion, 1952).

People differ substantially in how quickly they learn information and how long they retain it. Memory researchers have examined the relation between learning rate and retention over the past 120 years or so, and conclusions have differed considerably with regard to how learning rate and retention are related. While some researchers had claimed that information quickly learned was also quickly forgotten—arguing for the notion of "quickly come, quickly go" (p. 81, Strayer & Norsworthy, 1917)—the more consistent finding has been that people who learn more quickly also show better retention for the learned information. For example, Henderson (1903) found those "who learn quickest retain in general a greater percentage of what they have learned" (p. 53); Thorndike (1908) concluded that "it is the quick learners who are the good retainers" (p. 135). Pyle (1911), too, reported that "the most rapid learners showed the highest percentage of retention" (p. 311; also see Norsworthy, 1912; Lyon, 1917; Luh, 1922; Gillette, 1936).

There were problems with these early studies, of course. They failed to equate the degree of learning for the to-be-remembered information which resulted in overlearning; they neglected to ensure that learners had similar prior knowledge of the to-be-learned material; or they relied upon self-reported learning times as an index of learning speed. To counteract some of these limitations, Woodworth (1914) introduced a dropout procedure whereby each time an item was correctly recalled from a list during learning, it was dropped out from subsequent testing. Woodworth, however, only used this dropout procedure to examine item-level learning and recall differences *across* people for Italian-English word pairs, rather than individual differences *between* people. More recently, Nelson et al. (2016) and Zerr et al. (2017) used this dropout procedure to examine individual differences in the rate at which material is learned and how well it is remembered, and found that the rate at which individuals learned Lithuanian-English word pairs was strongly correlated with how well they performed on delayed cued-recall tests at both shorter (5 min) and longer (48-60 hr) delays (also see Kyllonen & Tirre, 1988).

1.2 Learning Efficiency

Zerr et al. (2017) characterized the relationship between speed of learning and goodness of retention as *learning efficiency*: More efficient learning represents quicker, more durable long-term learning. They proposed that this relationship reflects a singular *learning efficiency* "ability" - more efficient learners tend to learn more quickly and better remember what was learned across time. This "ability" was stable (r = .70) for 46 people across an average of 3 years (Experiment 2 in Zerr et al., 2017) in substantially different environments (inside an MRI scanner, inside a laboratory, and online in whatever environment a participant chose) using different Lithuanian-English word lists. It was also stable (r = .68) across days in a larger online-only sample (N = 281; Experiment 1 in Zerr et al., 2017) using different Lithuanian-English word lists.

Why are efficient learners able to learn material in less time and fewer exposures, but still demonstrate substantially better retention at various delays? There are three primary proposed reasons for why faster learners demonstrate better retention:

1. *Attentional control* (Nelson et al., 2016; Zerr et al., 2017). 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). Previous research has shown that slower learners are more adversely affected by the interfering quality of intralist items (Stroud & Carter, 1961; Kyllonen & Tirre, 1988), and poorer attentional control on the part of slower learners may be one reason why.

2. *Differences in strategy use* (Desrochers & Begg, 1987; McDaniel & Kearney, 1984; Resnick & Neches, 1984; Zerr et al., 2017). Another potential reason why faster learners show better retention may be due to strategy use during learning. More efficient learners may employ more effective strategies, such as the keyword method, to better encode material; or perhaps they apply the same strategies as less efficient learners, but either adopt them more quickly in the learning sequence (e.g., have better metacognitive awareness) or simply implement them more effectively (e.g., have more effective keywords to link the two items of each pair together). McDaniel and Kearney (1984) demonstrated that, even in the absence of instructions on which learning strategies to use, better learners were able to spontaneously employ effective strategies for a particular learning task to remember information effectively. MacLeod (2003) emphasized that strategies that improve one person's learning also improve another person's learning, and that the most critical difference is instead, "how people select optimal processes from their repertoire for a particular learning situation" (p. 252).

3. *Differences in prior knowledge* (Kurtz & Zimprich, 2014; Resnick & Neches, 1984; Royer, Hambleton, & Cadorette, 1978; Shuell, 1972). While Zerr and colleagues used Lithuanian-English pairs to reduce participants' reliance on prior knowledge, it is reasonable to suspect that participants with greater vocabulary knowledge or verbal ability had an advantage on the task (cf. Kyllonen, Tirre, & Christal, 1991). For example, if a participant had to learn the word pair UGNIS – FIRE, it would be beneficial if the participant knew the word *ignite* to use as a keyword mediator (i.e., UGNIS looks like *ignite*, which reminds them of FIRE). In addition,

Resnick and Neches (1984) specify that, “domain-specific knowledge relevant to learning also includes differences in knowledge of strategies” (p. 308), so it is possible that having more prior knowledge of English or more experience with acquiring other languages may also inform participants’ strategy use.

Zerr et al. (2017) put forth two questions for future research as it pertains to efficient learning, both of which will be preliminarily addressed in this thesis. The first question is whether learning efficiency is a domain-general or a domain-specific phenomenon (Gagné, 1972; Resnick & Neches, 1984). If an individual is able to both quickly acquire and successfully retain certain kinds of information, such as verbal-verbal paired associates, will this ability generalize to other kinds of materials, such as visuospatial stimuli? This question is addressed in Experiment 1 of this thesis. The second question, considered in Experiment 2, concerns whether the retentive advantage displayed by quicker learners persists for a delay of longer than two or three days—and relatedly—whether fast and slow learners show differential rates of forgetting. However, whether differential forgetting rates exist for fast and slow learners—and the topic of measuring forgetting more broadly—are contentious arguments in the memory literature.

1.3 Experiment 1: Generalizability of Learning Efficiency

Zerr et al. (2017) characterized the positive correlation between learning rate and retention using a multitrial learning and recall procedure (termed the “Learning Efficiency Task” or LET) with Lithuanian-English word pairs. These foreign language paired associates were used for several reasons: First, verbal materials are discrete, easily scored items (Tulving, 1983). Second, a foreign language paired-associate task is one in which the contribution of learner sophistication should be diminished because strategies are not as readily available as in most standardized tests of learning and memory. For instance, the California Verbal Learning Test-

Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) utilizes categorized word lists, and can thus benefit from strategies such as chunking; the Wechsler Memory Scale-Fourth Edition (WMS-IV; Wechsler, 2009) 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 as compared to more common foreign languages (e.g., Spanish), reducing the opportunity for participants to rely on prior relevant knowledge.

To examine whether learning efficiency generalizes to other kinds of material, Experiment 1 utilizes Chinese-English paired associates to represent a visuospatial-verbal relationship (Kang, 2010). Chinese characters are ideographs, which are relatively non-verbalizable to people who read and speak only phonetic languages (Wang & Thomas, 1992). These sort of stimuli may further limit the number of learning strategies available, and presumably make it more difficult to cheat during the learning phase (e.g., writing down or typing the pairs) or to rehearse them during the retention interval. While both Lithuanian-English and Chinese-English stimuli still utilize English pairings, the difference between the Lithuanian words and Chinese characters make it a sufficient starting point for examining how well learning efficiency generalizes.

1.4 Experiment 2: Learning Efficiency at Longer Delays

Another question is whether retention advantages for faster learners remain at delays of longer than two to three days, or if retention differences instead become attenuated over time. In addition, there is a substantial amount of disagreement as to whether faster and slower learners show similar or differential rates of forgetting. The vast majority of studies that examine how learning rate and memory performance interact use a single retention test at a single point in time. Examining forgetting rates, however, requires multiple measures of retention, the simplest being the difference in scores between two retention tests. Forgetting thus defined represents an absolute performance decrement over time independent of initial performance differences (Slamecka & McElree, 1983).

While many researchers disagree about the true relation of learning rate to retention, most researchers tend to agree that forgetting is a process independent from both learning and retention and that regardless of performance on these variables, forgetting will occur at a similar rate across people (Underwood, 1954; Shuell & Keppel, 1970; Kyllonen & Tirre, 1988; though see Loftus, 1985b).

1.4.1 The Measurement of Forgetting

Forgetting is measured either longitudinally or cross-sectionally. Longitudinal studies of forgetting test the same individual at two or more retention intervals, which can result in attrition and—even more problematically—repeated retrieval confounds. Most (if not all) longitudinal studies of forgetting examine retention performance on the same material, which can introduce a testing effect by practicing retrieval for the same material over time (Runquist, 1983). Repeated retrieval of the same material is problematic as studies have demonstrated that testing reduces the rate of forgetting across days (Roediger & Karpicke, 2006; Wheeler, Ewers, & Buonanno, 2003). Cross-sectional studies are less useful for investigating individual differences in forgetting across time, since these studies rely on the immense assumption that different groups properly represent how forgetting progresses in a given individual (Nairne & Pandeirada, 2008).

The current study attempted to assess forgetting in the same individuals by measuring retention for half of the learned material after a delay of 48 hours, and the other half of the learned material after a delay of 1 week. The rate (or slope) of forgetting in this case is simply the difference in scores between retention tests at 48 hours and 1 week. This design reduces the confound of repeated retrieval after the acquisition phase while allowing for the examination of forgetting rates within the same individuals. The only other learning rate study (to the author's knowledge) to test different sets of material at different delays was Norsworthy (1912), who had 83 college students study a minimum of 200 German-English paired associates across 19 days. Students took an immediate test after the study period over 50 of the 200 pairs, and then a month later were tested on a different subset of 50 pairs; she found that scores on both tests were positively correlated (r = .60), and that it was the "rapid learners [who] retain[ed] more than the slow learners" (p. 217). However, Norsworthy relied upon self-reported study times outside of the lab as an index of learning rate, and she did not equate the learners in terms of amount of material studied (some students learned the minimum 200 pairs, others studied as many as over 700 pairs).

1.4.2 Differential Forgetting Rates

Underwood (1954, 1964) emphatically concluded that slow and fast learners demonstrate a single, constant rate of forgetting, and that "individual differences in rate of forgetting are minimal" (p. 21, Underwood, 1972). A number of other studies have supported Underwood's conclusions (Loftus & Bamber, 1990; Postman, 1978; Shuell, 1972; Shuell & Keppel, 1970). Gentile et al. (1982), for example, concluded from their own data that, "if a 'slow' learner can be brought to the same standard of performance (learning criterion) as a 'faster' learner, there is no reason to expect the former to have a different forgetting curve than the latter" (p. 137).

Slamecka and McElree (1983) examined the literature on the effects of degree of learning on retention, and agreed with Underwood (1964) that forgetting is a process that is independent from learning. Across three experiments, Slamecka and McElree (1983) manipulated the degree of learning for categorized lists, paired-associate lists, and sentence lists, and found that the degree of learning for verbal materials did not impact the slope of forgetting functions. A brief overview of the forgetting literature by Nairne and Pandeirada (2008) reiterated the same message: "Variables that affect acquisition – e.g., word frequency, meaningfulness, similarity, and so forth – typically have little, if any, impact on subsequent forgetting rates" (p. 182).

MacDonald, Stigsdotter-Neely, Derwinger, and Bäckman (2006) however, demonstrated reliable heterogeneity in forgetting slopes. Similarly, Kyllonen and Tirre (1998) found that quicker learners forgot (measured as delayed recall and relearning time) fewer name-number paired associates across time compared with slower learners, despite having less exposure to the to-be-remembered information. This pattern remained significant when partialling out cognitive ability measures (short-term memory, reasoning ability, general learning speed ability, and general knowledge), suggesting that the rate of learning may not only influence how well information is retained at a single point in time, but may also influence the rate at which individuals forget over time.

1.5 Problems with Previous Research

Two problems exist for previous experimental work done on this topic. One is with respect to how data are analyzed. With the exception of MacDonald et al. (2006), these previous studies investigating individual differences have analyzed data using a contrastive approach—that is, participants are usually (arbitrarily) analyzed in terms of subgroups, such as the top and bottom 25% of performers, or those who performed 1 standard deviation above and below the mean. While this approach is certainly useful, its exclusive use may overshadow results revealed by more continuous, individual-level modeling techniques as opposed to subgrouping, where the results obtained may be dependent upon which subgroups are selected for the analysis. The second problem has already been mentioned and involves the repeated retrieval of all of the same material for each retention test. Because retrieval practice will slow the rate of forgetting of the material (Roediger & Karpicke, 2006), estimates of the effect of learning rate on forgetting rates in the literature are potentially contaminated by testing effects.

Chapter 2: Experiment 1

Experiment 1 examined how generalizable efficient learning is by using two sets of stimuli, including Lithuanian-English (verbal-verbal) and Chinese-English (visuospatial-verbal) paired associates. If efficient learning is more of a domain-general type of ability, then participants' performance should be correlated across types of learning material.

2.1 Method

2.1.1 Participants

Participants included 201 Amazon Mechanical Turk (MTurk) workers. Informed consent was obtained from all participants in accordance with standard Washington University human research practices, and participants were compensated $15 total for completion of all three sessions of the study. Because MTurk studies take place in uncontrolled environments, at the end of the second session we asked participants in a nonjudgmental way, with no risk to their compensation, whether they had written down any of the words during any of the sessions, and whether they had thought about or studied the words in the intervening time periods; 18 participants were excluded for doing so. In addition, 28 participants were excluded for failing to finish both sessions, 14 for restarting the task after the study portion, 4 for reporting a neurological disorder, 1 for not having normal (or corrected-to-normal) vision, 16 for having prior knowledge of either the Chinese language or a similar East Asian language (e.g., Japanese, Korean, Tibetan, Vietnamese, Filipino; these languages have similar characters), and 1 for having prior knowledge of the Lithuanian language. Of the final sample of 119 participants, 62 (52.1%) were female, with a mean age of 36.7 years (*SD* = 10.1, range = 20-64) and 14.9 years of education (*SD* = 1.9, range = 10-20).

2.1.2 Materials

Learning material consisted of 28 Lithuanian-English word pairs (e.g., KNYGA - BOOK) and 28 Chinese-English word pairs (e.g., 风 - WIND; see **Table A1** for the complete word lists). The Lithuanian-English word pairs were selected from previous norms (Grimaldi, Pyc, & Rawson, 2010; Zerr et al., 2017). English words from both lists were concrete nouns, and were matched as closely as possible for length (M_D = 0.1, range for both = 3-8), log frequency (M_D = 0.1; Lithuanian-English range = 6.8-11.6, Chinese-English range = 8.2-11.7), number of phonemes (M_D = 0.0, Lithuanian-English range = 1-6; Chinese-English range = 2-5), and number of syllables (M_D = 0.1, range for both = 1-2). These measures were calculated using the English Lexicon Project (ELP) database (Balota et al., 2007; http://elexicon.wustl.edu/). Diacritical marks and typographic ligatures were removed from the Lithuanian words to make them more similar to English words, and Lithuanian-English pairs were selected to reduce the incidence of cognates (cf. Nelson & Dunlosky, 1994) and false friends. All word pairs were displayed in capital letters on a white background in 36-pixel (27-point) font; Lithuanian and English words were presented in Roboto font (sans-serif; Arial font family), while Chinese characters were presented in Lora font (serif; PT Serif font family) to preserve character details. The online tasks administered in this thesis were coded using jsPsych (de Leeuw, 2015; http://www.jspsych.org/), an open-source JavaScript library for web-based experiments.

2.1.3 Procedure

This study occurred across two sessions (Figure 2.1). Word lists were blocked and counterbalanced such that participants studied and learned either all of the Lithuanian-English words before moving onto the Chinese-English pairs, or vice versa. In the first session, participants studied either 28 randomly-ordered Lithuanian-English word pairs or 28 randomly-ordered Chinese-English pairs. 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.

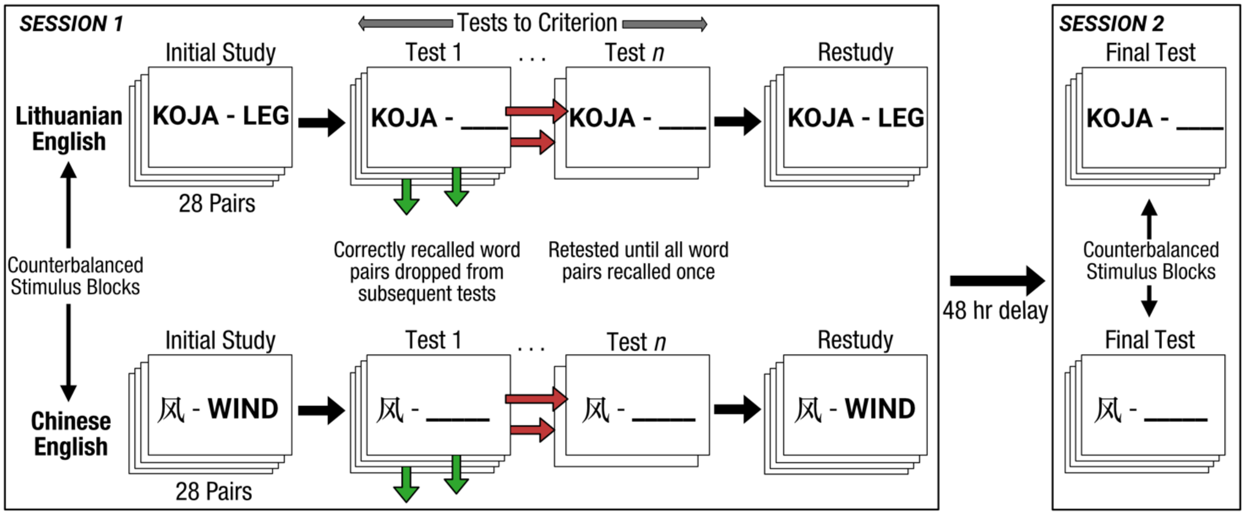


Figure 2.1 Procedure for Experiment 1. In Session 1, participants (*N* = 119) studied either 28 Lithuanian-English word pairs or 28 Chinese-English word pairs (word pairs were blocked and order was counterbalanced across people). They then took an initial cued recall test (Test 1); correctly recalled pairs were dropped from subsequent testing, whereas incorrectly recalled word pairs were presented on the next test, until all 28 Lithuanian-English (or 28 Chinese-English) word pairs were recalled. A final restudy took place on the word pairs, before participants completed the same procedure with the other kind of word pairs. Participants then took a final cued recall test on both types of word pairs 48 hr later.

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 (e.g., “BUGNAS”) or Chinese cue, presented on-screen for 5 s in a random order. Regardless of response accuracy, the correct pairing was displayed for 1 s; this pairing was followed by a 1 s ISI before the next cue appeared. Correctly recalled pairs were dropped from subsequent tests until the final cued recall test at the end of the session. This dropout procedure (Woodworth, 1914) ensured participants recalled each word pair exactly once to minimize overlearning.

Participants repeated this testing process on unrecalled word pairs until all 28 word pairs were 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 28 word pairs (*Tests to Criterion*) was used as an index of learning rate. The number of tests was limited to a maximum of 22 for each type of material in the interest of time. After a participant reached criterion, all word pairs were presented once more in a random order for a final study session, which was identical to the initial study session. Participants then repeated this procedure with the other type of word pairs.

Approximately 2 days later (M = 51.3 hr, SD = 7.2, range = 41.8-78.8), participants took a final cued recall test on the pairs. The Lithuanian-English and Chinese-English pairs were completed in blocks, so participants were first tested on the 28 Lithuanian-English pairs and then the 28 Chinese-English pairs, or vice versa. The ordering of blocks was counterbalanced across participants. Participants had 5 s to type the English target for the randomly-ordered Lithuanian or Chinese cue that was present on screen, before another cue was presented 1 s later. No feedback was provided.

After the first session, participants provided ratings (1 through 5, with 1 being the lowest) of how difficult they thought the task was, how much effort and they expended, and how focused they were on the task.

Based on Nelson et al. (2016) and Zerr et al. (2017), the three measures from this task—Test 1, Tests to Criterion, and Final Test—were combined into an overall metric due to high intercorrelation. Specifically, the standardized z-scores for each individual's Test 1, Tests to Criterion, and Final Test were averaged (Tests to Criterion was multiplied by -1, such that fewer Tests to Criterion was better) to create a single metric of learning and memory performance (*Learning Efficiency Score*). A higher overall Learning Efficiency Score (LE 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).

2.1.4 Analysis

Normality of dependent variables was assessed using the Shapiro-Wilk test; if normality was violated, the non-parametric Wilcoxon Signed Rank test replaced a paired-samples *t*-test and the non-parametric Mann-Whitney *U* (M-W *U*) test replaced an independent-samples *t*-test. The non-parametric equivalent of Pearson's *r*—Spearman's rho (r_s)—was calculated for ordinal data (i.e., Likert ratings). Differences were considered significant if $p < .05$. Reported *p*-values are two-sided unless otherwise noted.

2.2 Results

Neither participants' age nor years of education (YoE) were significantly correlated with overall task performance for either the Chinese-English (Age: $r = -.05$, $p = .585$; YoE: $r = .10$, $p = .296$) or Lithuanian-English (Age: $r = .00$, $p = .982$; YoE: $r = .16$, $p = .076$) stimuli. Difficulty ratings ($M = 4.4$, Median = 5, $SD = 0.71$, range = 2-5) significantly negatively correlated with Chinese-English LE Score performance, $r_s = -.35$, $p < .001$, 95% CI [-.50, -.18], such that participants who did worse on the task rated it as more difficult; this pattern was not significant for Lithuanian-English LE Scores, $r_s = -.14$, $p = .133$, 95% CI [-.31, .04]. Effort ratings were not related to Chinese-English performance, $r_s = .17$, $p = .060$, 95% CI [-.01, .34], or Lithuanian-English performance, $r_s = .08$, $p = .380$, 95% CI [-.10, .26]. Focus ratings were also not related to Chinese-English performance, $r_s = .10$, $p = .267$, 95% CI [-.08, .28], or Lithuanian-English performance, $r_s = .17$, $p = .073$, 95% CI [-.02, .34].

Overall, the Chinese-English stimuli were, as predicted, more difficult than the Lithuanian-English word pairs. The Chinese-English word pairs took longer on average to reach criterion, M_D = 0.6, $t(118) = -2.9$, $p = .002$, and were not recalled as well as Lithuanian-English word pairs on the initial test, M_D = -2.4, $t(118) = 4.4$, $p < .001$, or the delayed final test, M_D = -1.1, $t(118) = 2.2$, $p = .014$. Descriptive statistics for task performance can be found in Table 2.1.

Table 2.1 Descriptive statistics for Experiment 1.

Materials	Measure	*M*	*SD*	Median	Min	Max
Lithuanian-English	Test 1 Score (Max 28)	9.7	5.8	9.0	0.0	26.0
	Tests to Criterion	6.0	2.1	6.0	2.0	13.0
	Final Test Score (Max 28)	13.0	6.3	13.0	1.0	26.0
	LE Score (*z*-scores)	0.0	0.8	0.0	-2.1	1.8
Chinese-English	Test 1 Score (Max 28)	7.3	4.4	7.0	0.0	20.0
	Tests to Criterion	6.6	2.1	7.0	3.0	13.0
	Final Test Score (Max 28)	11.9	5.6	12.0	1.0	24.0
	LE Score (*z*-scores)	0.0	0.8	-0.1	-1.6	2.0

Note. LE Score represents Learning Efficiency Score, which is calculated for each person by taking their average *z*-score for Test 1 Score, Tests to Criterion (reverse-scored, since fewer tests is better), and Final Test Score.

In addition, there were significant order effects based on the stimuli order for participants in Session 1 that affected scores on Test 1 and Tests to Criterion. Specifically, participants who received the Chinese-English materials first performed more poorly on the Chinese stimuli than those who received the Chinese-English materials second for scores on Chinese-English Test 1, M_D = -1.93, M-W U_{117} = 1295, $p = .012$, Chinese-English Tests to Criterion, M_D = 1.5, M-W U_{117} = 2468, $p < .001$, and overall Chinese-English Learning Efficiency Score, M_D = -0.4, $t(117) = -2.8$, $p = .006$. This order effect did not significantly affect Final Test scores for either the Chinese-English, M_D = -0.4, $t(117) = -0.42$, $p = .677$, or Lithuanian-English stimuli, M_D = -2.0, $t(117) = -1.74$, $p = .084$. The order of stimulus presentation in Session 2 did not affect Final Test scores for either stimulus type (*p*s > .165).

The finding that learning rate (Tests to Criterion) and retention (Final Test scores at a 48 hr delay) are correlated was replicated for both stimuli, including Lithuanian-English pairs, *r* = -

.28, *p* = .002, 95% CI [-.44, -.11], and Chinese-English pairs, *r* = -.42, *p* < .001, 95% CI [-.56, -.26]. Additional correlations for task measures can be found in Table 2.2.

Table 2.2 Correlation matrix for LET measures for both types of stimuli.

		1	2	3	4	5	6	7	8
	Lithuanian								
1	Test 1								
2	Criterion	-.59*							
3	Final Test	.28*	-.28*						
4	LE Score	.81*	-.81*	.68*					
	Chinese								
5	Test 1	**.38***	-.27*	.46*	.49*				
6	Criterion	-.29*	**.37***	-.41*	-.46*	-.55*			
7	Final Test	.17	-.20*	**.57***	.41*	.54*	-.42*		
8	LE Score	.34*	-.35*	.59*	**.55***	.85*	-.80*	.80*	

Note. *$p < .05$. Criterion represents Tests to Criterion. Bolded values represent generalizability correlations for the same measures with different stimuli. Correlations within triangles represent correlations for the same stimuli.

2.2.1 Generalizability

As an indicator of generalizability across stimulus type, performance significantly correlated for each of the LET submeasures, including Test 1, *r* = .38, *p* < .001, 95% CI [.21, .52] the number of tests to reach criterion, *r* = .37, *p* < .001, 95% CI [.21, .52], and Final Test scores, *r* = .57, *p* < .001, 95% CI [.43, .68]. The average *z*-scores of these submeasures (LE Score) significantly correlated across stimulus type (Lithuanian-English and Chinese-English), *r* = .55, *p* < .001, 95% CI [.41, .67], which is representative of a large effect size (Cohen, 2009). See Figure 2.2 for scatterplots depicting performance for each of these measures across stimulus type.

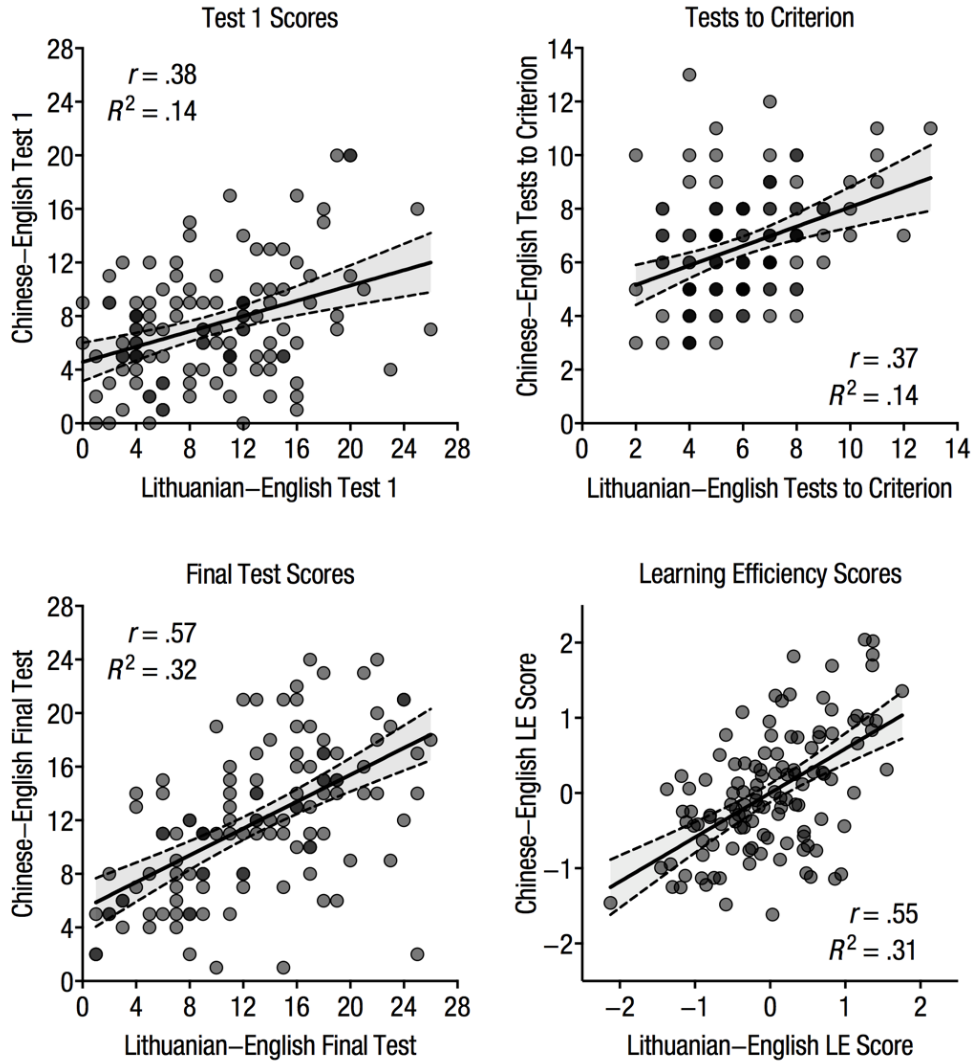


Figure 2.2 Scatterplots (with best-fitting regression lines and 95% confidence intervals) showing that performance was generalizable across types of learning stimuli (Lithuanian-English and Chinese-English paired associates). Overlapping data points are darker.

2.2.2 IntraClass Correlation

When using tasks for individual differences research, it is most desirable to have large between-subject variance and minimal within-subject variance. Hedge, Powell, and Sumner (2017) recommend examining reliability (or generalizability) of measures by using an IntraClass Correlation (ICC), which represents the correlation between repeated measures on the same subject by scaling the data with a pooled mean and standard deviation. An ICC ranges from 0 (large within-person variability and small between-person variability) to 1 (small within-person variability and large between-person variability), and can be thought of as a measure of the percentage of total variation that is attributable to between-person variation. A two-way random ICC for assessing absolute agreement amongst average scores can be represented via Equation 2.2 (Field, 2005; Hedge et al., 2017):

Equation 2.2

$$ICC = \frac{Between\ subject\ variance - Error\ variance}{Between\ subject\ variance + (Within\ subject\ variance - \frac{Error\ variance}{Number\ of\ tasks})}$$

An ICC was calculated (using Equation 2.2) for Learning Efficiency Scores from the Lithuanian-English and Chinese-English tasks. The two-way random ICC was used because LE Scores represent an average score of several measures. The ICC for LE scores between tasks was .71, indicating that LE Scores demonstrated good reliability (or in this case, generalizability) across Lithuanian-English and Chinese-English stimuli. This ICC was significantly greater than 0, $F(2, 118) = 3.46$, $p < .001$, 95% CI [.59, .80]; it was also significantly greater than a large effect size of .5, $F(2, 118) = 1.74$, , $p = .001$, 95% CI [.59, .80]. Another way to describe the ICC of .71 is that there was greater LE Score variability *between* participants on each task than *within* participants across each task; in other words, there was more observed variability between people and less variability within individuals, making it a compelling tool for studying individual differences.

2.2.3 Experiment 1 Summary

Experiment 1 found that efficient learning was generalizable across Lithuanian-English and Chinese-English paired associates, and that there was greater between-person variability on the task than within-person variability. Subjective ratings of difficulty were negatively correlated only for overall performance on the Chinese-English stimuli.

Chapter 3: Experiment 2

The purpose of Experiment 2 was to assess whether the relation between learning rate and retention holds for a longer retention interval of 1 week. Additionally, to mitigate the effect of repeated retrieval for multiple retention tests, half of the learned material was tested at 48 hours, and the other half was tested at 1 week. A related question was whether learning rate is related to differential forgetting rates, or if forgetting is constant regardless of how quickly material was learned in the acquisition phase.

3.1 Method

3.3.1 Participants

Participants included 298 MTurk workers. Informed consent was obtained from all participants in accordance with standard Washington University human research practices, and participants were compensated $15 total for completion of all three sessions of the study. Because MTurk studies take place in uncontrolled environments, at the end of the third session we asked participants in a nonjudgmental way, with no risk to their compensation, whether they had written down any of the words during any of the sessions, and whether they thought about or studied the words; 35 participants were excluded for doing so. In addition, 57 participants were excluded for failing to finish all three sessions, 6 for restarting the task after the study portion, 4 for reporting a neurological disorder, 4 for failing to answer any post-experiment questions, and 2 for having prior knowledge of the Lithuanian language. Of the final sample of 190 participants, 111 (58.4%) were female, with a mean age of 35.8 years (*SD* = 9.2, range = 20-62) and 15.4 years of education (*SD* = 2.2, range = 11-23).

3.1.2 Materials

Learning material consisted of 50 Lithuanian-English word pairs (**Table A2**) selected from previous norms (Grimaldi, Pyc, & Rawson, 2010; Zerr et al., 2017). As in Experiment 1, diacritical marks and typographic ligatures were removed from the Lithuanian words to make them more similar to English words, and Lithuanian-English pairs were selected to reduce the incidence of cognates (cf. Nelson & Dunlosky, 1994) and false friends. For this list, Lithuanian words varied in length between 4 to 9 characters ($M = 6.3$) and 1 to 4 syllables ($M = 2.4$). English words, all of which were concrete nouns, ranged from 3 to 8 characters ($M = 4.6$) and 1 to 2 syllables ($M = 1.2$). The combined Lithuanian-English word pairs ranged from 9 to 16 characters ($M = 10.9$) and 2 to 5 syllables ($M = 3.6$). Word pairs were displayed in all capital letters on a white background in black, 36-pixel (27-point) Roboto font (sans-serif; Arial font family).

3.1.3 Procedure

The procedure for Experiment 2 was similar to the procedure for Experiment 1, except it occurred across three sessions (Figure 3.1) and only used Lithuanian-English word pairs.

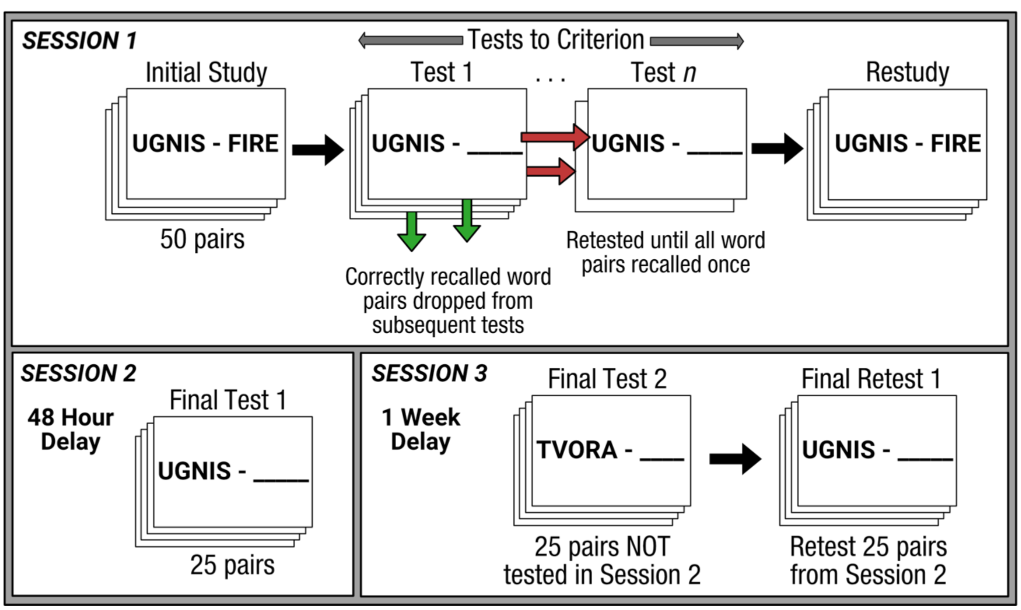


Figure 3.1 Procedure for Experiment 2. In Session 1, participants studied 50 Lithuanian-English word pairs, took an initial cued recall test (Test 1), and repeated this testing procedure with dropout for correct items (and corrective feedback). Once all 50 word pairs had been recalled once each, participants restudied the words. Session 2 took place 48 hr later and included a final cued recall test for only half of the 50 items from Session 1. Session 3 took place 1 week later and first included a final test for the words that had not been tested in Session 2 followed by a final test of the 25 words they had been tested on in Session 2.

In Session 1, participants studied 50 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 ISI. Participants were instructed to learn each of the word pairs for a later cued recall test. Participants then took an initial cued recall test (Test 1), where they had 5 s to type the English target for the Lithuanian cue. Regardless of response accuracy, the correct pairing was displayed for 1 s; this pairing was followed by a 1 s ISI before the next cue appeared. Correctly recalled pairs were again dropped from subsequent tests until each participant recalled all 50 word pairs correctly exactly once. Each test block was separated by 30 s of simple arithmetic to minimize maintenance of the word pairs in working memory. The number of tests was limited to a maximum of 30 in the interest of time. After a participant reached criterion, all word pairs were presented once more in a random order for a final study session, which was identical to the initial study session. At the end of the first session, participants provided ratings (1 through 5, with 1 being the lowest) of how difficult they thought the task was, how much effort and they expended, and how focused they were on the task.

Session 2 took place two days after the first session (M = 2.0 days, SD = 0.3, range = 1.3-2.9). Each participant was given a final cued recall test on only 25 of the 50 (50%) Lithuanian-English word pairs, which were randomly ordered. Two lists of 25 word pairs were created that were matched for length, frequency, and difficulty (Grimaldi et al., 2010; Zerr et al., 2017), and these two lists were counterbalanced across participants for Sessions 2 and 3. There was no effect of list on any final test scores at either delay (ps > .948). Participants had 5 s to type the appropriate English target to the given Lithuanian cue before the next cue was presented 1 s later. The final cued recall test did not provide feedback.

Session 3 took place 1 week after the first session (*M* = 7.0 days, *SD* = 0.3, range = 6.4-8.8). Each participant was first given a final cued recall test on the other 25 ("new") items they had not been tested on in the second session; they were then tested on the same 25 ("old") items they had been tested on in the second session. The third session's testing parameters (cue presentation time, ISI, no feedback) were the same as the second session. At the end of the third session, participants were asked to provide ratings (1 through 5, with 1 indicating "Never" and 5 indicating "Always") about their strategy usage. Specifically, participants were asked 12 questions about specific strategies, which are depicted in Section 3.2.6 of this thesis; these strategy questions and ratings were modified from McDaniel and Kearney (1984).

3.1.4 Analysis

Normality of dependent variables was assessed using the Shapiro-Wilk test; if normality was violated, the non-parametric Wilcoxon Signed Rank test replaced a paired-samples *t*-test and the non-parametric Mann-Whitney *U* replaced an independent-samples *t*-test. The non-parametric equivalent of Pearson's *r*—Spearman's rho (r_s)—was calculated for ordinal data (i.e., Likert ratings). For response time data, the non-parametric Kolmogorov-Smirnov test (K-S test) was used for two independent samples and the Kruskal-Wallis H test (K-WH test) was used for more than two independent samples. Differences were considered significant if $p < .05$. Reported *p*-values are two-sided unless otherwise noted. For multiple comparisons, *p*-values were adjusted (p_{adj}) using either Dunn-Bonferroni or Holm-Bonferroni corrections to reduce the familywise error rate.

3.2 Results

Participants' age and YoE did not significantly relate to LE Score (Age: $r = -.04$, $p = .627$; YoE: $r = .05$, $p = .478$). Session 1 difficulty ratings negatively correlated with overall task performance (LE Score), $r_s = -.28$, $p < .001$, and also negatively correlated with participants' years of education, $r_s = -.20$, $p = .007$; however, difficulty ratings did not correlate with participants' age, $r_s = .04$, $p = .584$. Neither Session 1 effort ratings nor focus ratings significantly correlated with LE Score (Effort: $r_s = .13$, $p = .078$; Focus: $r_s =.00$, $p =.980$), years of education (Effort: $r_s = -.01$, $p = .932$; Focus: $r_s = -.10$, $p =.186$), or age (Effort: $r_s = .09$, $p = .228$; Focus: $r_s = .08$, $p =.277$). See Table 3.1 for descriptive statistics from Experiment 2, and Table 3.2 for a correlation matrix of how performance measures were related.

Table 3.1 Descriptive statistics for Experiment 2.

Measure	*M*	*SD*	Median	Min	Max
Test 1 Score (50 max)	8.7	6.9	7.0	0.0	38.0
Tests to Criterion	10.5	4.4	9.5	4.0	26.0
Final Test 1 (25 max; 48 hr delay)	12.2	5.2	12.0	0.0	25.0
Final Test 2 (25 max; 1 week delay)	6.8	5.0	6.0	0.0	25.0
Final Retest 1 (25 max; 1 week delay)	10.9	5.5	11.0	0.0	25.0
Learning Efficiency Score (*z*-scores)	0.0	0.8	0.0	-2.0	2.6

Table 3.2 Correlation matrix for the LET measures.

	Test 1	Criterion	Final Test 1	Final Test 2
Test 1	1			
Criterion	-.54	1		
Final Test 1	.34	-.58	1	
Final Test 2	.44	-.34	.60	1
Final Retest 1	.42	-.58	.82	.71

Note. All correlations are significant at $p < .001$. Criterion refers to Tests to Criterion.

3.2.1 Do Faster Learners Show a Retention Advantage at 1 Week?

Yes (see Table 3.3 and Figures 3.2 and 3.3). When the sample was divided into 4 equal quartiles (each $n = 47$) on the basis of learning rate (Tests to Criterion), a one-way ANOVA for Final Test 1 scores was significant, $F(3,184) = 34.65$, $p < .001$, $\eta^2 = .36$. Post hoc tests with Holm-corrected p-values indicated that all of the four groups significantly differed from one another, all $p_{adj} < .033$. The one-way ANOVA remained significant for Final Test 2 scores at a 1-week delay, $F(3,184) = 11.34$, $p < .001$, $\eta^2 = .16$. Post hoc tests indicated that the only non-significant differences between retention after 1 week were between the 2nd and 3rd learning rate quartiles, $p_{adj} = .926$.

A one-way ANOVA was also significant for Final Retest 1 at 1 week, $F(3,184) = 35.93$, $p < .001$, $\eta^2 = .37$. Post hoc tests revealed the only non-significant differences were between the 2nd and 3rd quartile, $p_{adj} = .087$.

Table 3.3 Performance on Final Tests binned by Tests to Criterion (learning rate).

Learning Rate Quartile	Criterion *M* (*SD*)	Final Test 1 *M* (*SD*)	Final Test 2 *M* (*SD*)	Final Retest 1 *M* (*SD*)
1	16.6 (3.2)	7.5 (3.7)	3.9 (3.5)	6.1 (3.5)
2	11.2 (1.0)	11.7 (4.0)	6.9 (4.7)	10.3 (4.8)
3	8.4 (0.6)	13.5 (4.0)	7.0 (3.9)	11.9 (4.5)
4	5.8 (1.0)	16.0 (4.9)	9.2 (5.5)	15.4 (4.6)

Note. Learning Rate Quartile represents the Tests to Criterion quartiles (1 = slowest 25% of learners, 4 = fastest 25% of learners). Criterion represents the number of tests required to recall each of the word pairs once.

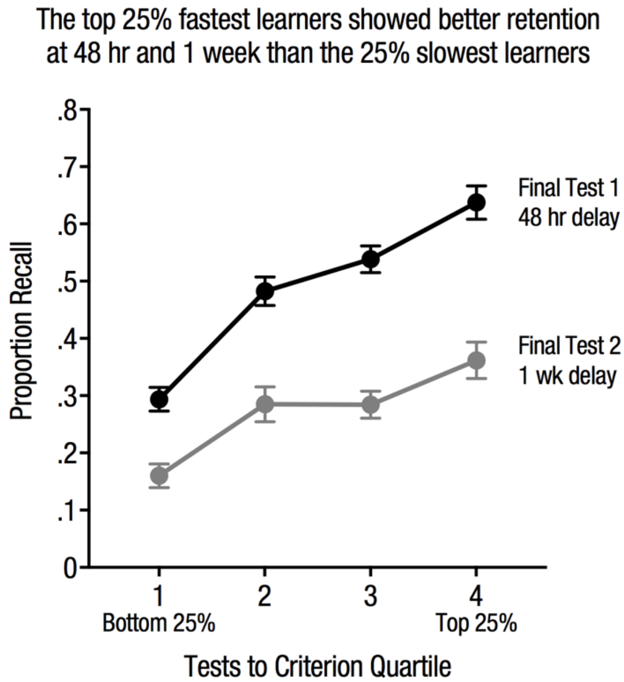


Figure 3.2 Retention performance on Final Test 1 (48 hr delay) and Final Test 2 (1 week delay) binned by Tests to Criterion (learning rate).

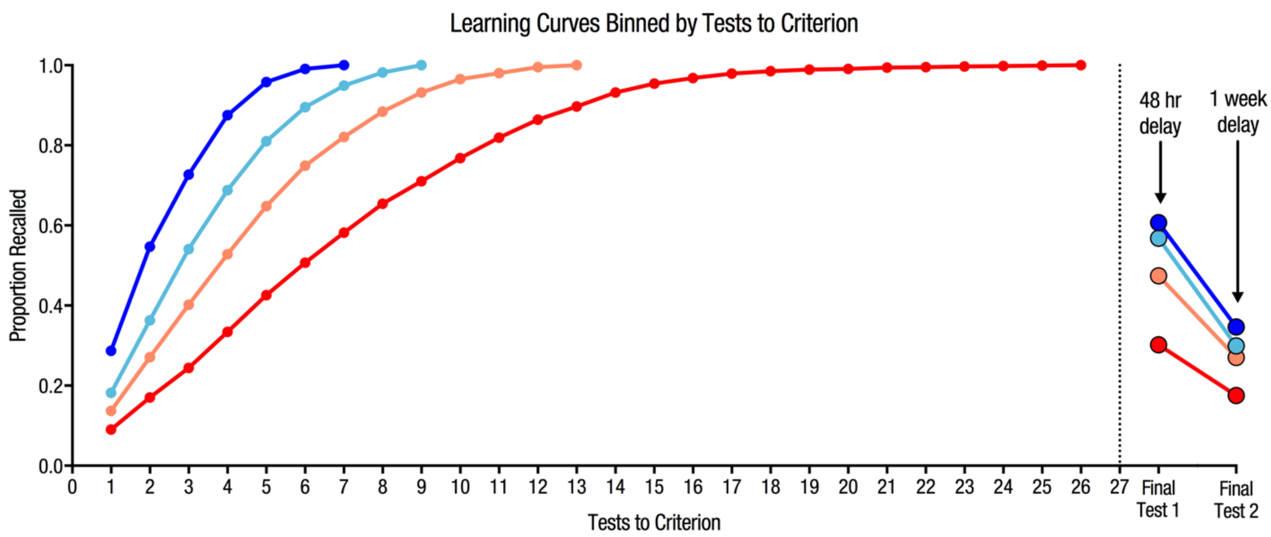


Figure 3.3 Learning curves and forgetting slopes binned by Tests to Criterion performance (learning rate). The graph represents each quartile's mean correct proportion recall for each test block. The 25% of participants (blue) who reached criterion more quickly showed better retention at delays of both 48 hr and 1 week than the 25% of participants who reached criterion the slowest (red).

The Top 25% fastest learners showed better retention of 25 word pairs after a delay of 1 week than the Bottom 25% slowest learners did at a delay of 48 hours, $t(92)=1.80$, one-sided p = .038. Top 25% fastest learners (n = 47) had an average proportion of .36 recall after 1 week (SD = .22), whereas the bottom 25% learners (n = 47) had an average proportion of .29 recall after 48 hr (SD = .14).

The number of tests required for participants to reach criterion significantly negatively correlated with retention performance at the 48 hr delay (r = -.58, p < .001) and for a different set of 25 Lithuanian-English words at the 1 week delay (r = -.34, p < .001), though the relationship was slightly attenuated at the longer retention delay (see Figure 3.4).

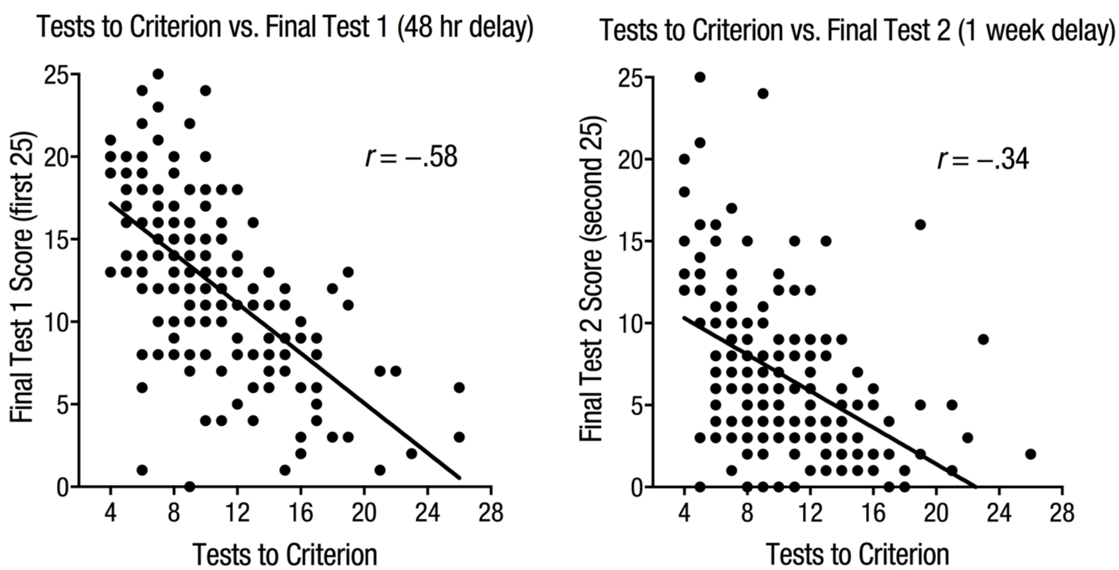


Figure 3.4 Scatterplots representing the significant relationship between Tests to Criterion (fewer tests represents a quicker learning rate) and Final Test performance at delays of 48-hours (left) and 1-week (right) for different sets of 25 Lithuanian-English word pairs.

3.2.2 Multilevel Modeling: Learning Rate Predicts Heterogeneous Forgetting Slopes

Multilevel models were conducted to assess the contribution of learning rate (Tests to Criterion) to subsequent forgetting slopes (the difference in scores between Final Test 1 and Final Test 2). Multilevel models are similar to regression, but differ in a few ways: First, parameters are allowed to vary (random effects) as opposed to a fixed value estimate from the overall sample (fixed effects) used in regression. Additionally, there is no assumption regarding the homogeneity of regression slopes, no assumption of independent errors, and missing data are not as problematic (though this last point was not relevant for the present dataset). Importantly, typical regressions assume no individual differences and thus treat all individuals as being the same (i.e., a single regression equation for all). However, multilevel models can account for individual differences as well as differences in learning rates.

R (version 3.4.2) and the *neml* package were used to fit equations to the forgetting data, and restricted maximum likelihood (REML) was used to estimate robust standard errors. Model fit was assessed via deviance comparisons between models, as well as Akaike's Information Criterion (AIC); smaller deviance and AIC values represent better model fit to the data. Two-level longitudinal multilevel models were conducted whereby each retention test (Final Test 1 at 48 hr and Final Test 2 at 1 week) was nested within each participant, and scores on Final Test 1 acted as each participants' "baseline" retention performance with Final Test 2 performance as the dependent variable. The multilevel slope parameter thus reflect each participant's individual forgetting rate between the two measures and can be treated as random or fixed.

The relationship between forgetting rate (differences between Final Test 1 and 2) and learning rate (number of tests required to reach criterion) showed significant variance in intercepts across participants, $SD = 5.45$, 95% CI [3.88, 7.65], $\chi^2(186) = 107.65$, $p < .001$. In addition, the slopes varied across participants, $SD = 2.19$, 95% CI [1.03, 3.17], $\chi^2(186) = 38.87$, $p < .001$. Learning rate significantly predicted the heterogeneous rate of forgetting as defined by the difference score, $b = 0.47$, $t(186) = 5.08$, $p < .001$. Contrary to our predictions and to other research on the rates of learning and forgetting, we found that—statistically—faster learners (those who took fewer tests to reach criterion) showed larger rates of forgetting from delays of 48 hr to 1 week (see Figure 3.5). The same outcome is observed when conducting a simple linear regression on absolute difference scores between Final Tests 1 and 2 using Tests to Criterion as a predictor, $F(1,189) = 17.84$, $p < .001$, $b = 0.31$, $p < .001$. There are many caveats to this conclusion—both statistical and theoretically—based on the present dataset that are addressed in the general discussion.

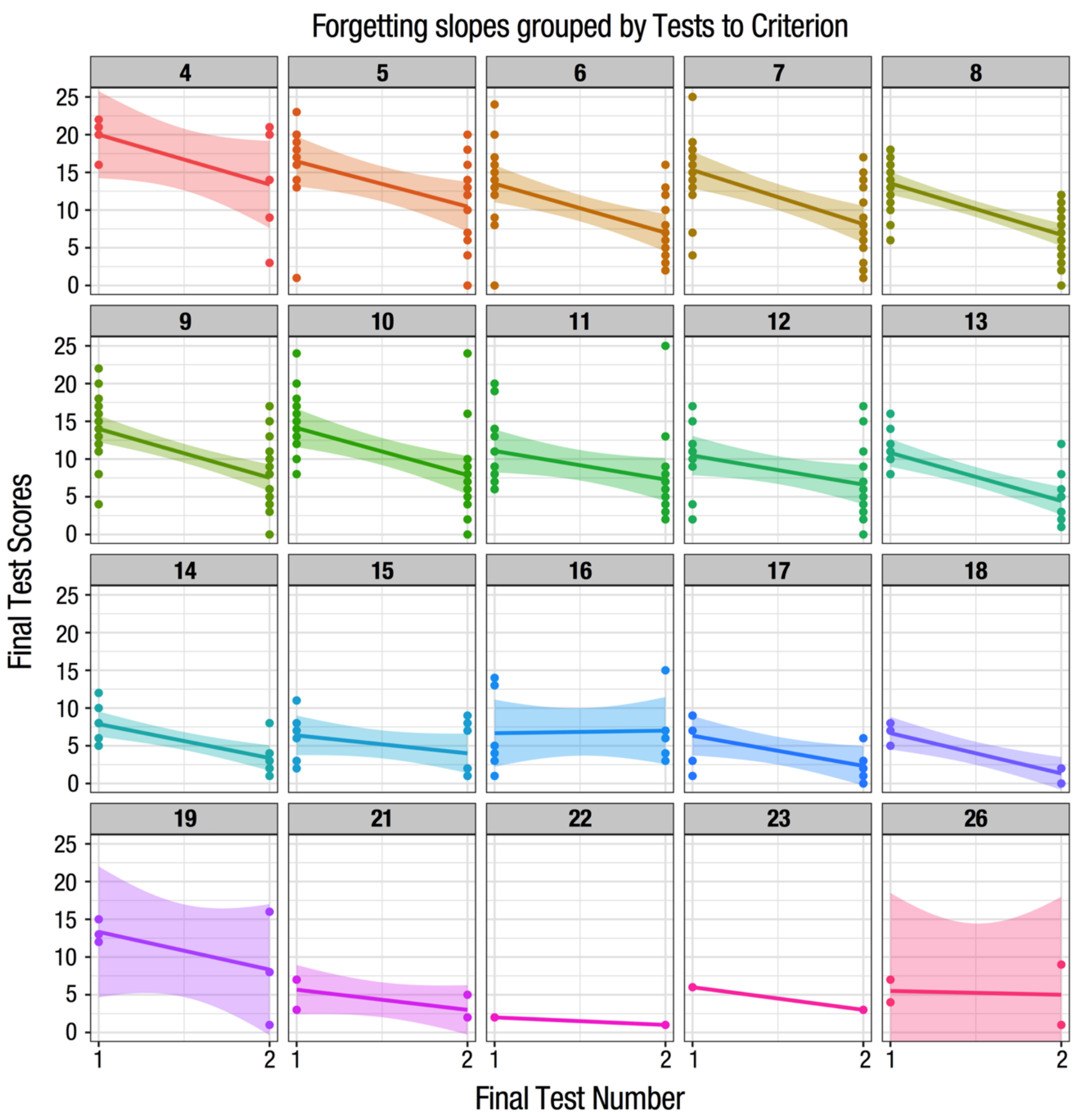


Figure 3.5 Scatterplots representing random-effects slopes (with standard errors) modeled via multilevel modeling for final test scores (y) at a delay of 48 hr (Final Test Number 1; x) and a delay of 1 week (Final Test Number 2; x). These data are subgrouped by Tests to Criterion. The top left grid represents forgetting

slopes from 48 hr to 1 week for the fastest learners (4 Tests to Criterion); the bottom right grid represents forgetting slopes from 48 hr to 1 week for the slowest learners (26 Tests to Criterion). Visually, the slopes become less steep as learning rate decreases (slopes tend to become more flat moving left to right across the columns as well as top to bottom across the rows); however, Final Test scores also tended to decrease as learning rate decreased, as did sample size per bin, so floor effects, restricted range, and asymptotic performance are problematic for interpretations. Multilevel models treated learning rate and individual subjects as random effects, and slopes were created for each participant; however, to make the figure more intuitive, the group average slope (via model fit predictions) is shown.

3.2.3 Multilevel Modeling: Testing the Same Material at Both Delays Produced Homogenous Forgetting Slopes

As mentioned in the introduction, no studies to our knowledge (aside from a poorly-controlled experiment by Norsworthy, 1912) that have investigated the effect of learning rate on retention and forgetting have had participants retrieve *separate* subsets of the material at different time delays. Rather, these studies have asked participants to repeatedly retrieve the same information at separate delays, which is problematic as retrieval practice minimizes forgetting of the retrieved information. This practice would especially be problematic in assessing forgetting as it relates to learning rate, as faster and slower learners could, theoretically, differentially benefit from the effects of testing.

The relationship between forgetting slopes for material retrieved at a 48 hr delay and the same material re-retrieved at a 1 week delay (difference scores between Final Test 1 and Retest 1) and learning rate (number of tests required to reach criterion) showed significant variance in intercepts across participants, SD = 3.94, 95% CI [1.56, 6.21], $\chi^2(186) = 83.25$, $p < .001$. Importantly, however, the slopes were homogenous and did not vary across participants, SD = 0.18, $\chi^2_{Change}(186) = 0.49$, $p = .484$. Tests to Criterion (learning rate) did not significantly predict the homogenous slope of forgetting, $b = -0.04$, $t(188) = -0.94$, $p = .348$ when examining differences between material tested at 48 hr and retested at 1 week (see Figure 3.6). This finding is contradictory to the one obtained for delayed tests on different material (see Figures 3.7 and 3.8), though likely results from restricted range in the difference scores.

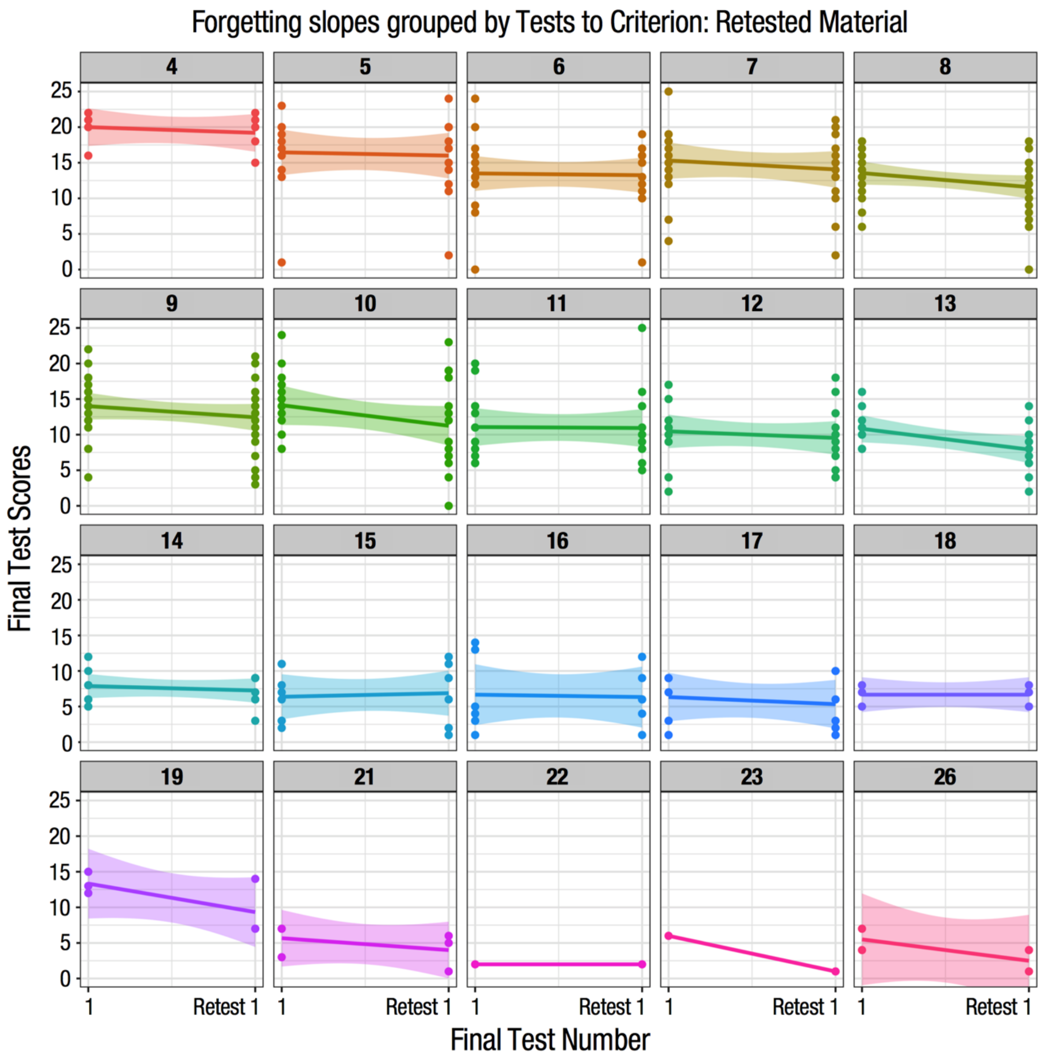


Figure 3.6 Scatterplots representing random-effects slopes (with standard errors) modeled via multilevel modeling for final test scores (*y*) at a delay of 48 hr (Final Test Number 1; *x*) and on the retested material at a delay of 1 week (Retest 1; *x*). These data are subgrouped by Tests to Criterion. Compared to Figure 3.5, testing the same material at both delays resulted in homogenous forgetting slopes.

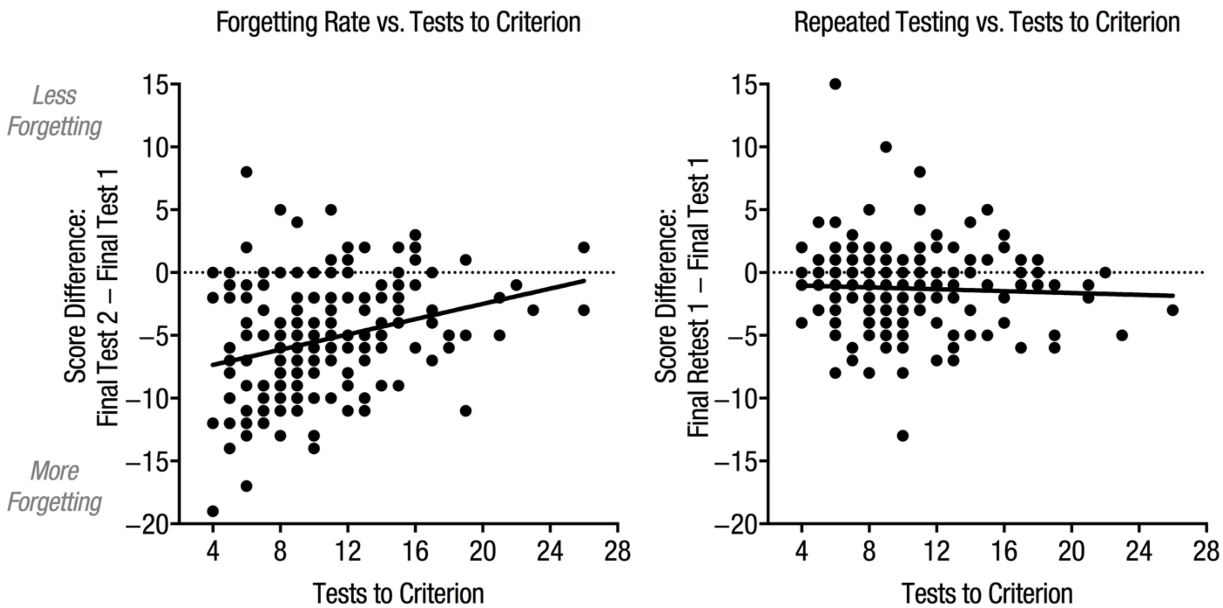


Figure 3.7 Simple difference scores between Final Test 1 and Final Test 2 scores (left) and Final Test 1 and Final Retest 1 scores (right) plotted by Tests to Criterion. Smaller *y*-values imply greater forgetting, while smaller *x*-values indicate faster learning.

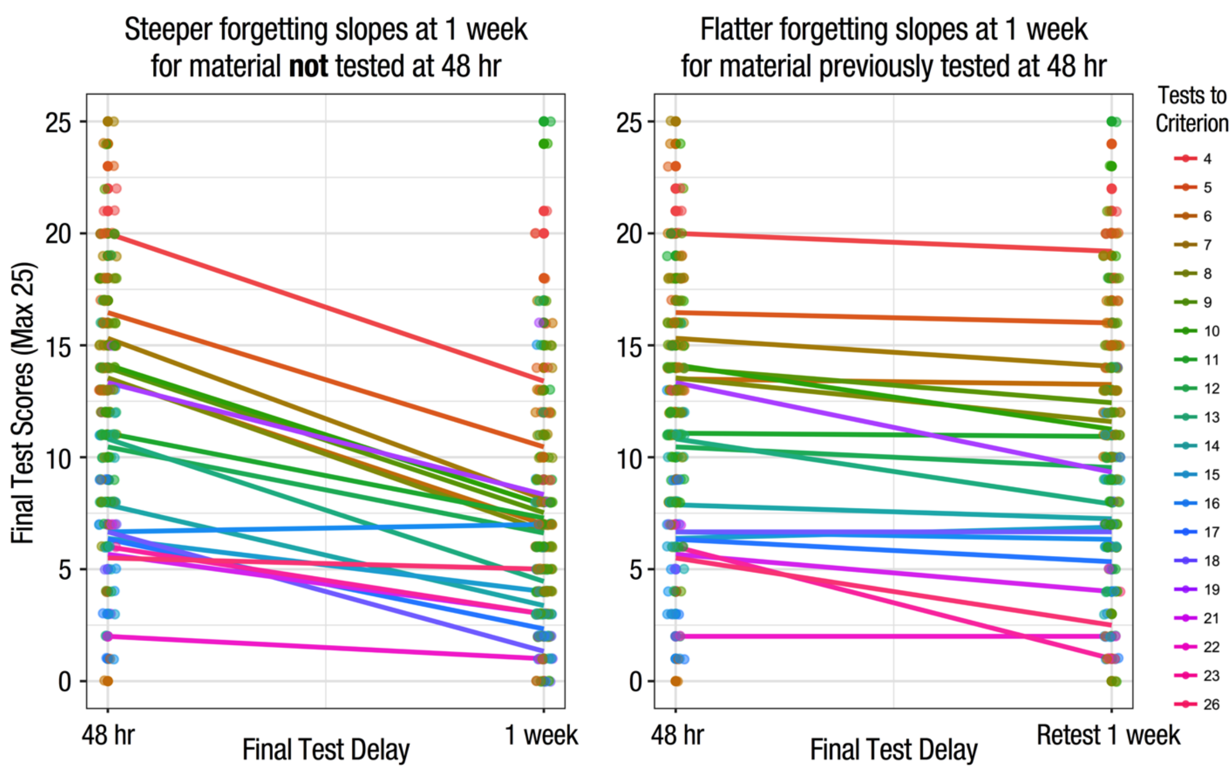


Figure 3.8 Random slope models of forgetting at 48 hr and 1 week. The figure on the left represents heterogeneous forgetting slopes for a **different** 25 items tested at 48 hr and 1 week, while the figure on the right represents homogenous forgetting slopes for the **same** 25 items tested at 48 hr and 1 week.

3.2.4 Two-Way Mixed ANOVAs Reiterate Results from Multilevel Modeling

Another way to support the results from the multilevel modeling approach and indicate its consistency via the typical contrastive method, two-way mixed ANOVAs were conducted to test main effects and interactions.

Tests to Criterion Quartiles vs. Final Test 1 (48 hr) and Final Test 2 (1 week)

A 4 x 2 mixed ANOVA was conducted with Tests to Criterion quartiles (1, 2, 3, and 4) as the between-subjects variable and retention test (Final Test 1 at 48 hr vs. Final Test 2 at 1 week) as the within-subjects factor. There was a significant main effect of Tests to Criterion quartile, $F(3,184) = 28.11$, $MSE = 27.68$, $p < .001$, $\eta_p^2 = .31$, and a main effect of retention test, $F(1,184) = 285.14$, $MSE = 9.70$, $p < .001$, $\eta_p^2 = .61$. There was a significant interaction between Tests to Criterion quartile and retention test, $F(3,184) = 5.48$, $MSE = 53.14$, $p = .001$, $\eta_p^2 =.08$.

Tests to Criterion Quartiles vs. Final Test 1 (48 hr) and Final Retest 1 (1 week)

A 4 x 2 mixed ANOVA was also conducted for Tests to Criterion quartiles (1, 2, 3, and 4) as the between-subjects factor and retention tests for the same material (Final Test 1 at 48 hr vs. Final Retest 1 at 1 week) as the within-subjects factor. There was again a significant main effect of Tests to Criterion quartile, $F(3,184) = 5.48$, $MSE = 53.14$, $p = .001$, $\eta_p^2 = .08$, and a significant main effect of retention test, $F(1,184) = 28.35$, $MSE = 5.23$, $p < .001$, $\eta_p^2 = .13$. However, there was no longer a significant interaction between Tests to Criterion quartile and retention test, $F(3,184) = 0.85$, $MSE = 5.23$, p = .468, $\eta_p^2 = .01$, suggesting that there was no discernable effect of learning rate (Tests to Criterion) on forgetting slopes (Final Test 1 vs. Final Retest 1) for repeatedly tested material.

3.2.5 Does the Learning Rate for Individual Word Pairs Predict Recall Probability?

Woodworth (1914) examined the learning rate and retention of individual Italian-English word pairs, ultimately concluding that words that were learned more quickly were also better retained after an interval of 2 to 20 hours. Is the same true of the present dataset?

Each Lithuanian-English word pair was fit via a logistic function for each participant. The probability of recall on a final cued recall test (*Y*) after a 48 hr delay can be modeled by the test number on which a particular word was first recalled (or "learned") for each participant (*X*). Specifically, logistic regression was performed to predict the probability of recall (1 = recalled; 0 = not recalled) after 48 hr given each word pair's learning rate. In its simplest form, we can write an equation (Equation 3.2) for the relation between the probability of a given item being correctly recalled contingent upon the test number it was initially "learned" on:

Equation 3.2

$$\Pr(Y|X) = \frac{1}{1 + e^{-(b_0 + b_1 X_1)}}$$

where $P(Y)$ represents the probability of successful recall after 48 hr and X_1 represents learning rate. The logistic regression resulted in Equation 3.3, which is plotted in Figure 3.9.

Equation 3.3

$$\Pr(Recall|X_{Tests\ to\ Criterion}) = \frac{1}{1 + e^{-(2.07 + .839X)}}$$

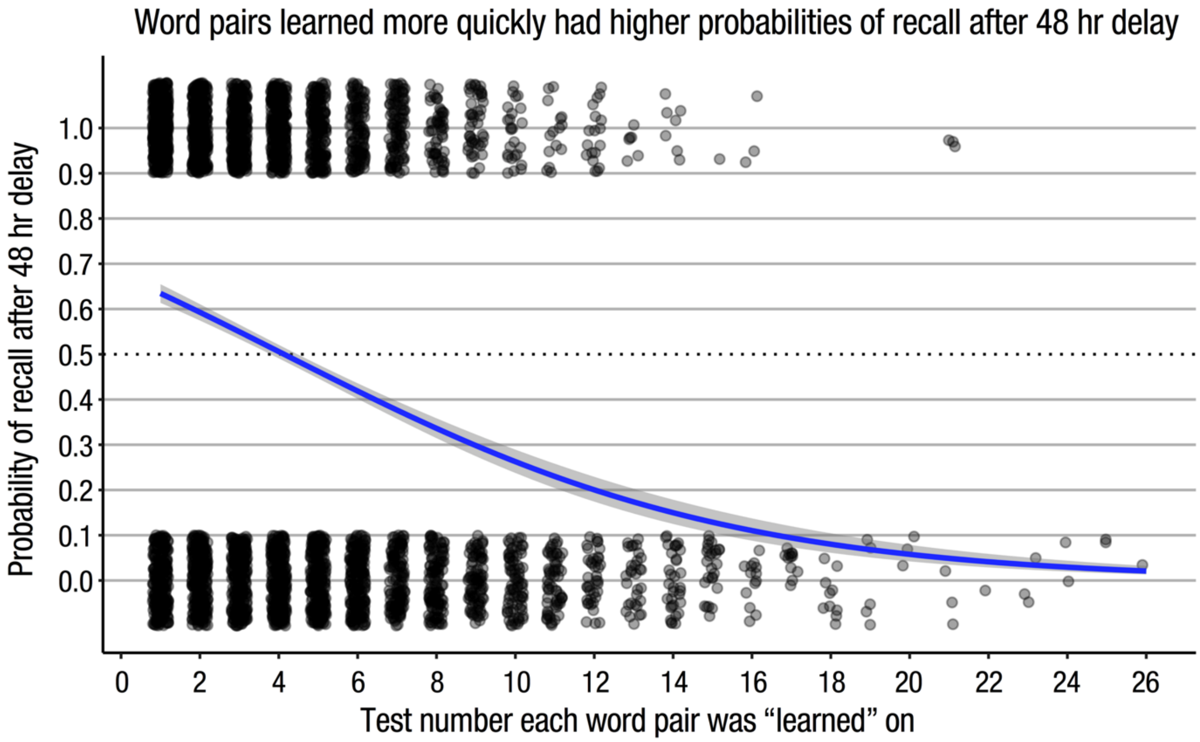


Figure 3.9 Words learned more quickly had a higher probability of recall after a delay of 48 hr. Points represent individual word pairs for each person, and points are jittered around 1 and 0 to reduce overlap. Line represents logistic model of prediction, with standard error.

The learning rate for an individual word pair significantly predicted the probability of recall for that item at a delay of 48 hr, Wald $z = 281.57$, $p < .001$. A χ^2 of 339.27, $p < .001$ indicated goodness of fit of the logistic model for the observed data. An odds ratio of 0.84 indicates that as the number of tests to learn a particular word pair for a particular person increased, the odds of recalling that particular word pair across persons decreased; as can be seen in Figure 3.9, an item learned after the fourth test had a probability of recall below 50%. This is somewhat of a counterintuitive result, as more tests meant more exposure to, more opportunities to study, and more opportunities to be tested on that particular word pair. The learning rate for individual word pairs also significantly predicted the probability of recall for that item at a delay of 1 week, Wald $z = 145.70$, $p < .001$. A χ^2 of 174.64, $p < .001$ indicated goodness of fit of the logistic model for the observed data at 1 week.

3.2.6 Learning Strategies

As mentioned in the introduction, one of the proposed reasons that learning rate and retention are related has to do with strategy use. In an attempt to examine this issue, questions about strategies (adapted from McDaniel & Kearney, 1984) were asked at the end of the third session of the experiment. Specifically, participants were asked to rate how frequently they had employed 12 possible strategies to help them learn and remember the Lithuanian-English word pairs. The complete list of questions and the label used to refer to each question is in Table 3.4. The proportions of responses for each strategy are depicted in Figure 3.10.

Table 3.4 Learning strategy questions asked of participants at the end of the third session.

Label	How often did you:	*M* (*SD*)
Physical	Construct sentences that described what you physically saw?	2.3 (1.1)
Repetition	Repeat the two words together over and over (either in your head or out loud) to commit them to memory?	3.5 (1.1)
Another Language	Think of a word in a different language (e.g., Spanish) to link to the Lithuanian and English word?	1.9 (1.0)
Keyword	Think of an English word that looked similar to the Lithuanian word, and used that similar-looking English word to remember the other English word?	3.3 (1.0)
Keyword Imagery	… by: Forming a picture or image in your mind of both items?	2.9 (1.2)
Keyword Meaning	… by: Thinking about the meaning of both words and how they relate to each other?	2.5 (1.1)
Keyword Personal	… by: Relating both words to a personal experience?	1.9 (1.0)
Keyword Sentence	… by: Making up a phrase or sentence using both words?	2.4 (1.1)
None	How often did you struggle or have difficulty trying to come up a strategy for learning the word pairs?	3.5 (0.7)
Failed	How often did your various strategies **not work** for helping you learn the word pairs?	3.2 (0.7)
Failed Perseverance	If a strategy did not work the first time for a certain word pair, how often did you **keep using that same strategy** for that word pair?	3.1 (0.9)
Failed Switch	If a strategy did not work the first time for a certain word pair, how often did you **switch strategies to something else** for that word pair?	2.8 (0.8)

Note. Ratings included: 1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Usually; 5 = Always. These strategy questions were modified from McDaniel and Kearney (1984).

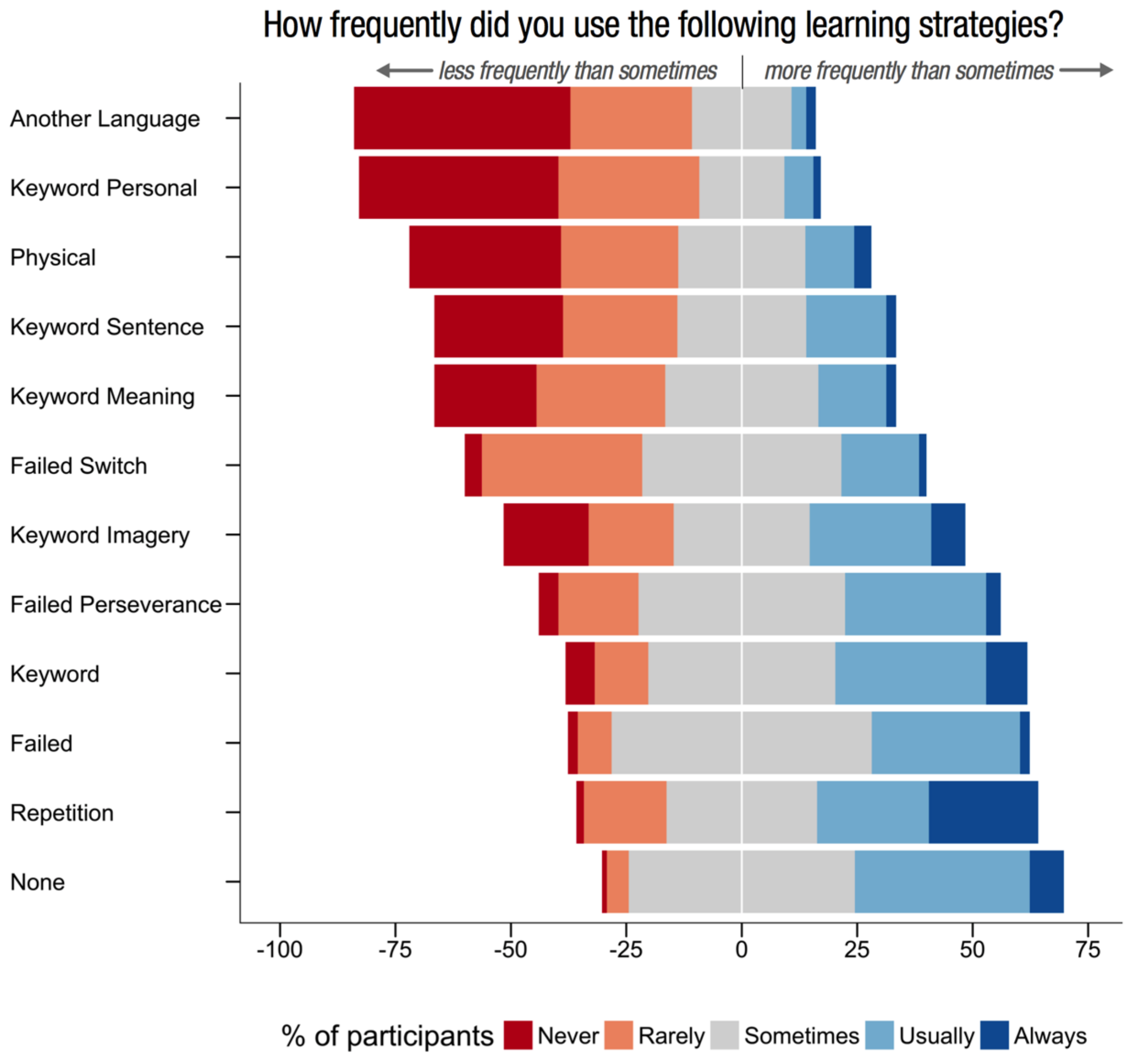
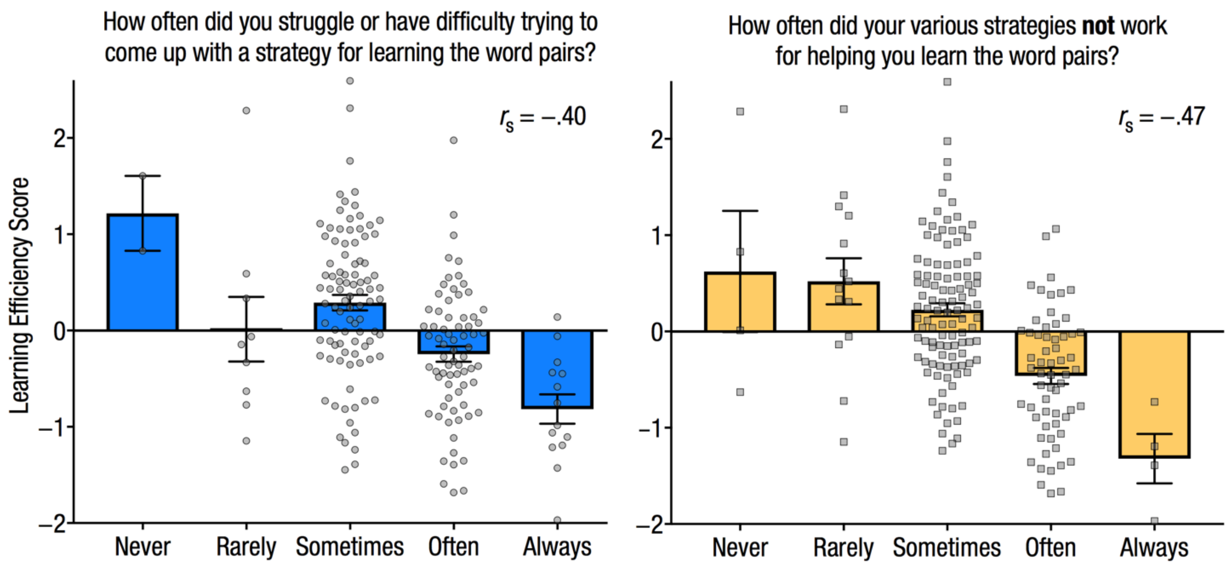


Figure 3.10 Proportion of responses for different strategy questions. Strategies are ordered from least frequently used (top) to most frequently used (bottom). X=0 centers responses at "Sometimes," thus a strategy that has more responses to the left of 0 implies they were less frequently used, whereas a strategy with more responses to the right of 0 implies they were more frequently used.

The only two strategies that were shown to relate to LET performance were the strategies labeled *None* ("How often did you struggle or have difficulty trying to come up with a strategy for learning the word pairs?") and *Failed* ("How often did you various strategies not work for helping you learn the word pairs?"). As can be seen in Figure 3.11, participants who more frequently struggled or had difficulty thinking of a strategy to use also had a tendency to do worse overall on the task, $r_s = -.40$, $p < .001$, 95% CI [-.51, -.27]; in addition, participants whose strategies did not work very often tended to perform worse on the task, r_s = -.47, p < .001, 95% CI [-.57, -.35]. Ratings for these two strategies were significantly correlated, r_s =. 53, p < .001, 95% CI [.41, .62], suggesting that participants who frequently struggled to come up with a strategy also tended to employ ineffective strategies. Another interesting finding is that those who more frequently employed Repetition to learn the word pairs (repeating the words either internally or out loud) tended to require more tests to reach criterion, r_s = .20, p = .007, 95% CI [.06, .33].



3.11 Colored bars represent the mean LE Score binned by strategy rating. Participants who never struggled with or failed at using strategies had higher overall LE Scores on average than participants who did struggle or failed to use strategies. Error bars represent standard error.

The two questions relating to failed strategies—*Failed Switch* (switched to a different strategy if a strategy was unhelpful) and *Failed Perseverance* (kept using the same strategy even though it was unhelpful)—were negatively correlated, r_s = -.47, p < .001, 95% CI [-.57, -.35], suggesting that (at least for these questions) participants were not providing the same rating for contrary strategy questions. A Spearman's rho correlation matrix for all strategy responses can be found in the Appendix (Table A3).

For reported strategy use, the Top and Bottom 25% of LE Score performers significantly differed for strategies labeled *Failed* ($U = 436$, $Z = -5.54$, $p_{adj} < .001$, $M_{Top25\%} = 2.83$, $M_{Btm25\%} = 3.66$), *None* ($U = 555.5$, $Z = -4.58$, $p_{adj} < .001$, $M_{Top25\%} = 3.02$, $M_{Btm25\%} = 3.77$), and *Repetition* ($U = 817.5$, $Z = -2.25$, $p_{adj} = .025$, $M_{Top25\%} = 3.15$, $M_{Btm25\%} = 3.64$).

3.2.7 Retrieval Speed Differences

Another one of the proposed mechanisms underlying efficient learning is attentional control. As specified in the introduction, people who demonstrate better attentional control also demonstrate more rapid, refined memory search at retrieval (Shipstead, Redick, Hicks, & Engle, 2012; Unsworth & Spillers, 2010). In the current experiment, response time (ms) was measured as the latency between the Lithuanian cue presentation and the first key stroke a participant made. This measure can thus be treated as an index of "retrieval speed" (cf. Hunt, 1978), as most of this time is presumably used by participants to retrieve the appropriate target for each cue, and there is no reason to expect differences between motor response timings for participants (i.e., there should be no significant differences in how long it physically takes participants to begin typing their answer).

Retrieval times binned by top and bottom quartiles for Learning Efficiency Scores are depicted in Figure 3.12. Retrieval times were significantly different across these 4 learning efficiency bins, as revealed by a K-W $H_3 = 12.0$, $p = .007$. Dunn-Bonferroni follow-up pairwise comparisons revealed the only significant difference in retrieval speeds were between Quartile 1 (Bottom 25% of LET performers) and Quartile 4 (Top 25% of LET performers), $Z = 37.65$, $p = .001$, $p_{adj} = .005$. Retrieval speeds were not significantly different across bins for Final Test 2, K-W $H_3 = 5.51$, $p = .273$, nor when the two Final Tests were pooled, K-W $H_3 = 7.82$, $p = .050$.

A two-sample K-S test found that the top 25% and bottom 25% of LE Score performers had significantly different response time distributions for Final Test 1 at the 48 hr delay, K-S Z = 1.71, p = .006, and a Cohen's d of .68 represents a medium to large effect size (Cohen, 2009). Pooled response time distributions from Final Test 1 and Final Test 2 were also significantly different between top and bottom quartiles, K-S Z = 1.40, p = .040.

In addition, the top 50% and bottom 50% of LE Score performers had significantly different response time distributions for Final Test 1, K-S Z = 1.46, p = .028. There were no significant RT differences between the top and bottom 25% of LET performers for Final Test 2, Z = .57, p = .899, or for Final Retest 1, Z = .992, p = .278.

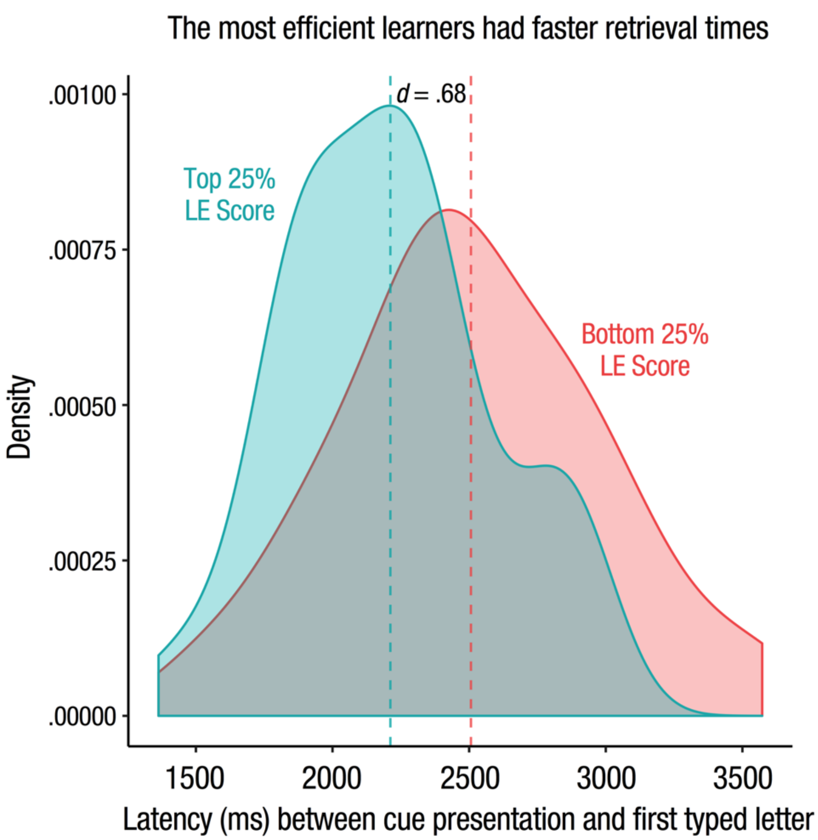


Figure 3.12 Probability density functions for correct-answer response latencies (ms) binned by the Top 25% (green) and Bottom 25% (red) Learning Efficiency Scores for Final Test 1. Dashed lines represent mean response latencies for each group; d represents Cohen's d, the standardized difference between the group means.

3.2.8 Experiment 2 Summary

Experiment 2 found that faster learners still demonstrate better retention than slower learners at a longer retention interval of 1 week. In addition, faster learners showed a greater drop in scores from 48 hours to 1 week for previously-untested material, suggesting that learning rate could be related to differential forgetting. For material tested at 48 hours and retested at 1 week—which is the typical way forgetting is assessed—there was no observable indication of differential forgetting rates. Slower learners had more difficulty coming up with learning strategies, and also demonstrated slower retrieval time for the 48 hour cued recall test. Difficulty ratings were negatively correlated with task performance. At the item-level, words that were learned earlier in the multitrial learning phase had a higher probability of being recalled after both delays, mimicking findings by Woodworth (1914).

Chapter 4: General Discussion

The two goals of this thesis were to demonstrate an initial investigation into the generalizability of learning efficiency across different types of materials, and to assess whether quicker learners maintain a retentive advantage over slower learners at a longer delay of 1 week. Experiment 1 (N = 119) demonstrated that learning efficiency was sufficiently generalizable between verbal-verbal (Lithuanian-English) and visuospatial-verbal (Chinese-English) stimuli. Experiment 2 (N = 190) demonstrated that, not only did the fastest learners retain more than the slowest learners at a delay of 1 week, but the fastest learners showed significantly higher retention after a 1 week retention interval than the slowest learners did at a retention interval of only 48 hours. Additionally, when different material was tested at different delays, heterogeneous forgetting slopes were found across learning rates, which is consistent with conclusions from Kyllonen and Tirre (1988) and MacDonald et al. (2006), though our results were in the opposite direction (faster learners showed greater forgetting slopes). However, when the same material was tested at different delays, homogenous forgetting slopes were found across learning rates, which is consistent with earlier work (Gentile et al., 1982; Gentile et al., 1995; Postman, 1978; Shuell, 1972; Shuell & Keppel, 1970; Slamecka & McElree, 1983, Underwood, 1954; Underwood, 1964). There are a number of caveats concerning these findings in the Limitations section.

One reasonable conclusion from these experiments is that more efficient learners simply find the task to be easier than less efficient learners. Subjective ratings of difficulty showed this tendency for Experiments 1 and 2 as well as for previous experiments (Zerr et al., 2017), so perhaps the safest assumption is that “efficient learning” is merely a matter of task difficulty for different people. Perhaps the task is, as Robert Bjork suggested at a conference, simply an undesirable difficulty for some learners and a desirable difficulty for others. These points then beg the question, why is this task easier for some and not for others? The three suggested explanations in the introduction for why individuals show differences in learning efficiency—attentional control, strategy use, and prior knowledge—are thus far supported by the literature, and were partially supported by the current experiments in the form retrieval speeds and questions relating to strategy usage.

4.1 Limitations and Future Directions

There were several limitations and caveats to these experiments and to the conclusions one can draw from them. One limitation for Experiment 1 concerns the similarity of the generalizability stimuli. Both the Lithuanian and Chinese stimuli were paired with English words, which may have increased the generalizability of learning efficiency due to a common underlying factor or ability (i.e., English vocabulary knowledge or verbal ability). In unpublished lab data, raw vocabulary scores from the WASI-II (where participants are asked to define a series of vocabulary words; Wechsler, 2011) significantly correlated with the number of tests to reach criterion, $r = -.30$, in a sample of 92 people. This negative correlation suggested higher vocabulary knowledge coincided with a quicker rate of learning Lithuanian-English paired associates. However, when vocabulary scores were statistically controlled for (i.e., partialled out), the zero-order correlation between the number of tests to reach criterion and final test scores after a 48 hour delay remained unchanged (r_{Y1} and $r_{Y1.2} = -.71$). Future work should attempt to use wider varieties of stimuli (e.g., visuospatial-visuospatial pairings) and even different types of memory tests aside from cued recall using paired associates (e.g., recognition, free recall, etc.; cf. Underwood, Boruch, & Malmi, 1978).

The two most problematic issues for Experiment 2 relate to theoretical issues concerning both learning and forgetting. While the experiments in this thesis were designed to produce equal "learning" of paired-associates for every person, *learning* cannot be observed directly and must instead be inferred based upon measures of *performance* (for review, see Soderstrom & Bjork, 2015). In the present experiments, "learning" was inferred to be equal when everyone demonstrated equal performance (i.e., recall each word pair exactly once during the learning phase); Underwood (1954) argued that this sort of approach does not actually produce equal learning, and that faster learners benefit more (i.e., have a greater association strength) for recall of each item during learning. Data consistent with this conclusion can be seen in Zerr et al (2017), who showed that a short 5 min delay (with a video game distractor) after reaching criterion already permitted observable differences in recall performance, which was predicted by learning speed. This situation would subsequently produce differential retention favoring faster learners, which is a consistent finding in the literature. As Kyllonen and Tirre (1988) suggested, it may simply be the case that "fast learners appear to remember better because they learned better in the first place" (p. 394). In an attempt to mitigate this challenge, we attempted to measure a simple forgetting slope for each person, defined as the difference between two delayed retention tests.

The measurement of forgetting produces its own difficulties. While one analysis approach seemed to suggest the counterintuitive finding that faster learners (those who required fewer tests to reach criterion) actually forgot at faster rates between two retention tests with different material, scaling and floor effects were an issue. Faster learners still demonstrated better retention at both retention intervals, and slower learners demonstrated asymptotic performance near floor; it may be possible that faster learners and slower learners do show similar rates of forgetting when assessing different (or more) time points. In addition, scaling is a common problem in the study of forgetting (Loftus, 1978, 1985; Slamecka 1985; Slamecka & Katsaiti, 1988; Wixted, 1990), such that vertically-parallel slopes do not necessarily suggest that underlying forgetting rates are equivalent. How can we say that—psychologically—a person scoring 90% at 48 hr and 50% at 1 week is the same as someone going from 60% at 48 hr to 20% at 1 week? They both represent a 40% change, but inferring the two to be equivalent forgetting is dubious (cf. Nairne & Pandeirada, 2008).

If we are to assume that faster learners did learn better, or to a higher degree, then our results parallel those of Joinson and Runquist (1968) and Runquist and Snyder (1969). Both of these studies manipulated degrees of learning and examined forgetting. In the former, absolute losses between estimated immediate recall and 1-week delayed recall indicated less forgetting at the lowest-learning level; however, retention was near floor after 1 week for the lowest-learning level group, suggesting floor effects may have artificially constricted difference scores (or "forgetting slopes"). It is entirely possible, and perhaps even likely, that if the forgetting slope had been calculated between retention tests immediately after testing and after 48 hr, that the opposite pattern—slower forgetting for faster learners—would appear.

One goal for future research is to increase the number of items learned, and test the same individuals using a similar procedure at more points in time; for the most accurate measure of forgetting curves, it has been recommended to have at least five retention tests (Rubin & Wenzel, 1996). For example, if participants learned 100 Lithuanian-English pairs, what would their forgetting curve look like if five 20-item retention tests were provided at different retention intervals, such as immediately after acquisition, an hour after acquisition, 24 hr after acquisition, 48 hr after acquisition, and 1 week after acquisition? What would retention look like at a month, or 3 months, or 6 months? What about a year later? Modeling trajectories such as these for the same individuals over time may help provide a clearer answer. This would allow us to fit power curves to analyze individual subject forgetting functions (Wixted & Ebbesen, 1997), and to compare forgetting curves horizontally as opposed to vertically (Loftus, 1985a).

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445-459.

Cohen, J. (2009). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Psychology Press.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1-12. doi:10.3758/s13428-014-0458-y

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *The California Verbal Learning Test—Second Edition.* San Antonio, TX: The Psychological Corporation.

Desrochers, A., & Begg, I. (1987). A theoretical account of encoding and retrieval processes in the use of imagery-based mnemonic techniques: The special case of imagery-based mnemonic techniques. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 56-77). New York, NY: Springer-Verlag.

Gagné, R. M. (1972). Domains of learning. *Interchange*, *3*(1), 1-8. doi:10.1007/bf02145939

Gentile, J. R., Monaco, N., Iheozor-Ejiofor, I. E., Ndu, A. N., & Ogbonaya, P. K. (1982). Retention by "fast" and "slow" learners. *Intelligence*, *6*, 125-138. doi:10.1016/0160-2896(82)90010-1

Gentile, J. R., Voelkl, K. E., Pleasant, J. M., & Monaco, N. M. (1995). Recall after relearning by fast and slow learners. *Journal of Experimental Education*, *63*(3), 185–197.

Gillette, A. L. (1936). Learning and retention: A comparison of three experimental procedures. *Archives of Psychology*, *198*, 5-56.

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

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1-21. https://doi.org/10.3758/s13428-017-0935-1

Henderson, E. N. (1903). A study of memory for connected trains of thought. *The Psychological Review*, *5*, 1-94. doi:10.1037/h0093003

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review*, *85*, 109-130.

Joinson, P. A., & Runquist, W. N. (1968). Effects of intralist stimulus similarity and degree of learning on forgetting. *Journal of Verbal Learning and Verbal Behavior*, *7*, 554–559.

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition*, *38*(8), 1009–1017. http://doi.org/10.3758/MC.38.8.1009

Kurtz, T., & Zimprich, D. (2014). Individual differences in criterion-based dropout learning in old age: The role of processing speed and verbal knowledge. *European Journal of Ageing*, *11*, 183-193. doi: 10.1007/s10433-013-0293-5

Kyllonen, P. C., & Tirre, W. C. (1988). Individual differences in associative learning and forgetting. *Intelligence*, *12*, 393-421. doi:10.1016/0160-2896(88)90004-9

Kyllonen, P. C., Tirre, W. C., & Christal, R. E. (1991). Knowledge and processing speed as determinants of associative learning. *Journal of Experimental Psychology: General*, *120*(1), 57–79.

Loftus, G. R. (1985a). Consistency and confoundings: Reply to Slamecka. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 817-820.

Loftus, G. R. (1985b). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(2), 397-406.

Luh, C. W. (1922). The conditions of retention. *Psychological Monographs*, *31*, 1-87. doi:10.1037/h0093177

Lyon, D. O. (1917). *Memory and the learning process*. Baltimore, MD: Warwick & York, Inc.

MacLeod, C. M. (2003). Individual differences in learning and memory. In J. H. Byrne, H. Eichenbaum, H. L. Roediger, III, & R. F. Thompson (Eds.), *Learning & Memory* (2nd ed., pp. 251-253). New York, NY: Macmillan.

McDaniel, M. A., & Kearney, E. M. (1984). Optimal learning strategies and their spontaneous use: The importance of task-appropriate processing. *Memory & Cognition*, *12*(4), 361–373.

McGeoch, J. A., & Irion, A. L. (1952). *The psychology of human learning*. New York: Longmans, Green.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Forgetting. In R. Menzel & J. H. Byrne (Eds.), *Learning and memory: A comprehensive reference* (Vol. 2, pp. 179-194). Elsevier.

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

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

Postman, L. (1978). Picture-word differences in the acquisition and retention of paired associates. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 146–157.

Pyle, W. H. (1911). Retention as related to repetition. *The Journal of Educational Psychology*, *6*, 311-321. doi:10.1037/h0073818

Resnick, L. B., & Neches, R. (1984). Factors affecting individual differences in learning ability. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 275-323). Hillsdale, NJ: Erlbaum.

Roediger, H. L., III., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210.

Royer, J. M., Hambleton, R. K., & Cadorette, L. (1978). Individual differences in memory: Theory, data, and educational implications. *Contemporary Educational Psychology*, *3*, 182-203. doi:10.1016/0361-476x(78)90026-7

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*, 734-760.

Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, *11*, 641-650.

Runquist, W. N., & Snyder, R. A. (1969). Forgetting of verbal paired associates after low degrees of learning. *Journal of Verbal Learning and Verbal Behavior*, *8*, 512–517.

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

Shuell, T. J. (1972). *Individual differences in learning and retention* (Final Report No. 0-0341). Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Shuell, T. J., & Keppel, G. (1970). Learning ability and retention. *Journal of Educational Psychology*, *61*, 59-65.

Slamecka, N. J. (1985). On comparing rates of forgetting: Comment on Loftus (1985). *Journal of Experimental Psychology*, *11*, 812-816.

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 384–397.

Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 716–727.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*, 176-199.

Strayer, G. D., & Norsworthy, N. (1917). *How to teach*. New York: The Macmillan Company. doi:10.1037/14866-000

Stroud, J. B., & Carter, L. J. (1961). Inhibition phenomena in fast and slow learners. *Journal of Educational Psychology*, *52*(1), 30–34. http://doi.org/10.1037/h0041411

Tulving, E. (1983). *Elements of episodic memory*. New York, NY: Oxford University Press.

Thorndike, E. L. (1908). Memory for paired associates. *Psychological Review*, *15*, 122-138. doi:10.1037/h0073570

Underwood, B. J. (1954). Speed of learning and amount retained: A consideration of methodology. *Psychological Bulletin*, *51*, 276–282.

Underwood, B. J. (1964). Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, *3*, 112–129.

Underwood, B. J. (1972). Are we overloading memory? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston & Sons.

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General*, *107*(4), 393–419.

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

Wang, A. Y., & Thomas, M. H. (1992). The effect of imagery-based mnemonics on the long-term retention of Chinese characters. *Language Learning*, *42*, 359-376.

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—Second Edition*. San Antonio, TX: Pearson.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571-580.

Wixted, J. T. (1990). Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 927–935.

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, *25*(5), 731–739.

Woodworth, R. S. (1914). A contribution to the question of "quick learning, quick forgetting." *Psychological Bulletin*, *11*, 58-59.

Zerr, C. L., Berg, J. J., Nelson, S. M., Fishell, A. K., Savalia, N. K., & McDermott, K. B. (2017). *Learning efficiency: Identifying individual differences in learning rate and retention in healthy adults.* Manuscript resubmitted for publication.

Appendix

Table A1. Paired associates used for Experiment 1.

Lithuanian	*English*	*Chinese*	*English*
Obuolys	Apple	箭	Arrow
Tvartas	Barn	豆	Bean
Vonia	Bath	鸟	Bird
Tiltas	Bridge	车	Car
Pastatas	Building	椅	Chair
Pyragas	Cake	牛	Cow
Puodelis	Cup	污	Dirt
Durys	Door	火	Fire
Bugnas	Drum	叉	Fork
Akis	Eye	友	Friend
Zuvis	Fish	金	Gold
Plaukas	Hair	草	Grass
Raktas	Key	手	Hand
Riteris	Knight	心	Heart
Koja	Leg	马	Horse
Turgus	Market	王	King
Pienas	Milk	刀	Knife
Burna	Mouth	月	Moon
Nafta	Oil	山	Mountain
Augalas	Plant	钉	Nail
Lietus	Rain	河	River
Ziedas	Ring	衫	Shirt
Kambarys	Room	天	Sky
Muilas	Soap	曲	Song
Laiptelis	Stair	店	Store
Gatve	Street	风	Wind
Stalas	Table	狼	Wolf
Vanduo	Water	木	Wood

Table A2. 50 Lithuanian-English word pairs used in Experiment 2.

Obuolys - Apple	Raktas - Key
Maisas - Bag	Riteris - Knight
Tvartas - Barn	Laiskas - Letter
Pliazas - Beach	Turgus - Market
Knyga - Book	Burna - Mouth
Smegenys - Brain	Adata - Needle
Sepetys - Brush	Tinklas - Net
Pastatas - Building	Dazai - Paint
Mygtukas - Button	Lietus - Rain
Pyragas - Cake	Ziedas - Ring
Kede - Chair	Kambarys - Room
Miestas - City	Kilimelis - Rug
Puodelis - Cup	Mokykla - School
Sokis - Dance	Batas - Shoe
Purvas - Dirt	Kriaukle - Sink
Durys - Door	Muilas - Soap
Bugnas - Drum	Krautuve - Store
Tvora - Fence	Gatve - Street
Ugnis - Fire	Stalas - Table
Zuvis - Fish	Arbata - Tea
Sakute - Fork	Traukinys - Train
Auksas - Gold	Medis - Tree
Zole - Grass	Vanduo - Water
Plaukas - Hair	Langas - Window
Namas - House	Vilkas - Wolf

Table A3. Spearman rank correlation matrix for reported strategy use.

		1	2	3	4	5	6	7	8	9	10	11
1	Physical	1										
2	Repetition	.00	1									
3	Another Language	.09	.02	1								
4	Keyword	.29**	.03	-.05	1							
5	Keyword Imagery	.34**	-.09	-.03	.53**	1						
6	Keyword Meaning	.19*	.09	.12	.37**	.30**	1					
7	Keyword Personal	.38**	.07	.10	.26**	.41**	.38**	1				
8	Keyword Sentence	.56**	-.06	.02	.38**	.47**	.23*	.32**	1			
9	None	-.14	.12	.04	-.07	-.13	-.06	-.11	-.15*	1		
10	Failed	-.17*	.17*	-.02	-.13	-.15*	-.09	-.03	-.22*	.53**	1	
11	Failed Perseverance	-.03	.08	-.12	.22*	.15*	.10	-.02	.09	.02	.07	1
12	Failed Switch	.14	.09	.08	.01	.05	-.12	.02	.06	.13	.08	-.47**

Note. *p < .05, **p < .001. Theoretically related strategies are grouped together within triangular borders to make interpretation easier.