Spring 5-2019

Individual Differences in Verbal and Visuospatial Learning Efficiency

Thomas Spaventa
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychological & Brain Sciences

Individual Differences in Verbal and Visuospatial Learning Efficiency

by

Thomas Spaventa

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

May 2019
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am indebted to my mentor, Kathleen McDermott, for guidance in conceptualizing and designing this project, and for encouraging me to bring it to fruition. This study benefited in innumerable ways, both large and small, from her remarks and suggestions. The members of my lab were also instrumental in realizing this project. I consulted with Chris Zerr extensively, and I am deeply grateful for his insights and statistical expertise. Ruthie Shaffer assisted with task programming and helped me resolve a litany of technical issues, enabling data collection to get off the ground. Additionally, Nate Anderson, Hank Chen, Wenbo Lin, and Abhilasha Kumar offered constructive criticism and support at various stages. Finally, I would like to thank my committee members, David Balota and Mark McDaniel, for their suggestions and for challenging me to think more deeply.

This project was 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.

Thomas Spaventa

Washington University in St. Louis

May 2019

ABSTRACT OF THESIS

Individual Differences in Verbal and Visuospatial Learning Efficiency

by

Thomas Spaventa

Master of Arts in Psychological & Brain Sciences

Washington University in St. Louis, 2019

Professor Kathleen B. McDermott, Chair

There is a great deal of variability in how quickly people learn and how long they remember information. Zerr and colleagues (2018) found a robust and stable relationship between an individual's rate of learning and the durability of their memory, with faster learners tending to retain more after a delay. The relationship between the rapidity and longevity of learning was characterized as learning efficiency. The present study extends these findings by testing whether learning efficiency generalizes across divergent verbal and visuospatial tasks. An ancillary aim was to assess learning efficiency using a continuous measure that can capture fine-grained individual differences in learning. Participants (N = 112) learned and recalled Lithuanian-English word pairs and object locations using a multi-trial cued recall paradigm. Estimates of individuals' learning efficiency generalized across tasks, suggesting that this construct may tap into a domain-general ability. Additionally, the spatial precision of recalled object locations correlated with both the speed and durability of learning, indicating that continuous measures may also be used to evaluate the efficiency of learning.

Chapter 1: Introduction

People differ markedly in their ability to learn and remember information. Whereas some absorb and retain the names of new acquaintances, the finer points of a story, and the dates of upcoming appointments with astonishing facility, others experience frequent memory failures. Such individual differences in memory can not only lead to prosaic day-to-day differences in recall, such as differences in the ability to remember shopping list items at the grocery store, but can also produce differences that have more severe consequences in educational and vocational settings. For example, a student that learns sluggishly may fail class exams, and an employee with poor memory may not be able to meet the demands of their job.

Individual differences in memory have been an area of interest to psychologists for over a century. Ebbinghaus (1885/2013) observed "how differently do different individuals behave in this respect! One retains and reproduces well; another poorly" (p. 155). Since then, theorists and experimentalists have worked to characterize the myriad ways that learners vary and identify sources of this variance (see Bors & MacLeod, 1996; Unsworth, 2019 for reviews). This thesis more specifically addresses individual differences in rate of learning and retention, the relation between these two attributes, and the extent to which this relation generalizes across verbal and visuospatial domains.

1.1 Do Faster Learners Retain More?

A recurring question in memory research has been whether initial speed of learning is related to the amount of information remembered over time (Gillette, 1936; Underwood, 1954; Zerr, 2017). That is, do faster learners retain more than their slower counterparts? The simplicity of this inquiry belies the complexity of addressing it. Gillette (1936) identified three common approaches to investigating the association between learning rate and retention. The first, dubbed the method of Equal Amount Learned, has participants learn a fixed amount of material. Participants are tested on the material repeatedly until they correctly recall all of it on a single test, and the variable of interest is the number of tests required to reach this criterion. A shortcoming of this procedure is that it induces overlearning by reexposing learners to material that has previously been successfully recollected. The second method, the method of Equal Opportunity to Learn, grants participants a fixed time to learn material, and the quantity of material retained after a delay is now the dependent measure. The critical flaw of this method is that it fails to equate initial learning, which artifactually inflates the correlation between learning speed and retention. Finally, the third procedure, and the one espoused by Gillette, is the method of Adjusted Learning.

Pioneered by Woodworth (1914), the method of Adjusted Learning ensures that all tested items are learned, but unlike the method of Equal Amount Learned, it also prevents overlearning by dropping items that are correctly recalled from subsequent tests (see Underwood, 1954 for a dissenting view). Using this paradigm with number-picture pairs, Gillette (1936) found that quicker learners retained more. Half a century later, Kyllonen and Tirre (1988) employed a variation of the method of Adjusted Learning on a large sample (N = 685) of Air Force recruits; echoing Gillette's results, it was once again the fastest learners that retained the most.

1.2 Learning Efficiency

Recently, Zerr et al. (2018) used the method of Adjusted Learning and Lithuanian-English word pairs to examine individual differences in learning ability. Lithuanian words were selected as cues because they are unfamiliar to most English speakers, and because Lithuanian belongs to a different language family from Romance languages that are commonly taught in schools. The novelty of Lithuanian reduces the influence of prior knowledge differences between learners, which have been found to account for variability in associative learning (Hundal & Horn, 1977; Kyllonen, Tirre, & Christal, 1991). Zerr et al. (2018) measured three indicators of learning performance: words recalled on an initial test, the number of tests required to recall all word pairs, and recall on a final, delayed test. All three measures were found to robustly intercorrelate, leading Zerr and colleagues to combine the measures into a composite Learning Efficiency Score, with efficient learning defined as learning that is both fast and enduring. Learning Efficiency Scores were stable across days ($r = .68$) and even over a three-year period ($r = .70$), suggesting that this measure represents a trait-like ability.

In a related study, Zerr and McDermott (2019) investigated whether learning efficiency generalizes to visuospatial material. In addition to learning Lithuanian-English word pairs, participants also learned Chinese character-English word paired associates. Chinese characters are logograms, making it especially difficult for English speakers to form verbalizable associations with them. Nevertheless, those who learned the Chinese-English pairs fastest tended to also retain the most, and performance on this task positively correlated with Lithuanian-English performance. This outcome suggests that learning efficiency generalizes beyond verbal-verbal paired associates, and that more efficient learners may have a retentive advantage even with material that does not readily lend itself to mnemonic strategies.

The present study aimed to further establish the generalizability of learning efficiency by using an even more divergent visuospatial task: object location learning. On this task, participants view the locations of common household objects within a circle and later attempt to recall these locations precisely. This object locations task was selected for two principle reasons. First, remembering object locations is important for everyday functioning (e.g., finding keys in a home or a car in a large parking lot). Thus, this task may capture differences in visuospatial learning that relate to real-world behavior. Second, recording the spatial precision of recalled locations enables a continuous rather than a binary accuracy measure to be used. Because memories are not just recollected in an 'all-or-none' manner but can vary in their fidelity, continuous measures provide a fine-grained index of the quality of recollection (Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016; Richter, Cooper, Bays, & Simons, 2016).

1.3 Precision of Visual Long-term Memory

Subjectively, we all share the experience of being able to recollect memories with intricate detail in some instances ("I left my wallet on the top left corner of the coffee table") and imprecisely in others ("My umbrella is somewhere in the house"). This intuition is supported by a body of research showing that memories do indeed vary in their precision. Research on the visual precision of memories originated in the working memory community. In an early experiment, participants were shown colored squares and, after a brief delay, asked to match the color of a specific square using a color wheel (Zhang & Luck, 2008). Using this procedure in combination with mixture models to decompose error arising from guesses and imprecise recollections, it was found that the precision of visual working memory varies independently of working memory capacity.

Over the past decade, continuous-report tasks have been adapted to study the precision of visual long-term memory. It has been known since at least the 1970s that the capacity of visual long-term memory is vast (Standing, 1973). More recently, Brady, Konkle, Alvarez, and Olivia (2008) tested the long-term storage capacity of memory for visual details. Participants viewed pictures of 2,500 objects with the goal of remembering them for a future test. Afterwards, they were shown these target images alongside closely matched foils that subtly differed (e.g., a half and a quarter of a melon) and attempted to identify the image they had seen previously.

Astonishingly, target images were correctly recognized 87% of the time. In a follow-up study, Brady, Konkle, Gill, Olivia, and Alvarez (2013) more directly tested the fidelity of visual long-term memory. Using a similar continuous-report procedure to Zhang and Luck (2008), but with real-world objects rather than squares, they found that long-term memory has a similar lower-bound to its precision as does working memory. Although recollection accuracy for object color dropped precipitously after a delay, the color precision of objects that were correctly recalled was comparable to the precision measured in a working memory condition.

This line of research has been extended to examine the precision of spatial memories. Harlow and Donaldson (2013) created a "positional response accuracy" task wherein participants learned to associate words with locations on the circumference of a circle. They found that, as is the case for memories of color, spatial recollection accuracy is separable from precision. Moreover, introspective judgments of spatial fidelity track objective measures of spatial error on a trial-by-trial basis, which bolsters the theoretical validity of precision being a distinct construct (Harlow & Yonelinas, 2016). Employing a similar task in a multi-trial learning paradigm, Lew, Pashler, and Vul (2016) found that precise object location memories developed quickly and endured even after a one-week retention interval. Collectively, these results indicate that memory for precise details is capacious, quick forming, robust to delays, and empirically dissociable from recall or recognition success.

1.4 The Present Study

As reviewed above, people dramatically differ in how efficiently they learn and retain information. The primary goal of the present study is to test whether learning efficiency generalizes across verbal and visuospatial learning. This experiment tested individuals' ability to learn and remember verbal and visuospatial paired-associates. Participants learned English translations of Lithuanian words on the verbal task and the locations of objects within a circle on the visuospatial task. A multi-trial, iterative cued-recall paradigm with dropout of correctly recalled items (i.e., the method of Adjusted Learning) was used in both cases. It is hypothesized that if learning efficiency reflects domain-general processes, performance across the two tasks will positively correlate.

Chapter 2: Main Experiment

2.1 Method

2.1.1 Participants

Two-hundred and sixteen participants were recruited from the Amazon Mechanical Turk (MTurk) marketplace and consented in accordance with the guidelines of the Washington University Human Research Protection Office. To incentivize completion of the entire study, participants received a flat rate of $12 for successfully completing both study tasks or for exceeding 25 test blocks on either task, at which point the study was terminated prematurely. A total of 104 participants were excluded from analyses, including 31 for failing to complete both tasks, 17 for having prior knowledge of or exposure to the Lithuanian language, 3 for reporting a learning disability, 2 for exhibiting no learning on at least 4 consecutive blocks at the beginning of a task, 1 participant who reported a neurological condition, and 1 for refreshing the webpage during a task. At the end of each task, participants were asked whether they had written down any information or taken pictures of the stimuli to help on the memory tests, and it was emphasized that receiving compensation was not contingent on their responses. An additional 49 participants were excluded for reporting doing so.

The final sample of 112 participants (47 female) included in analyses were between 19 and 66 years of age (M = 34.7, SD = 9.9) and had completed between 12 and 24 years of education (*M* = 14.7, *SD* = 2.1). Ninety-one participants self-reported being Caucasian, 8 Asian, 7 Black/African American, 4 multiracial, and 1 American Indian/Alaska Native; of these, 10 were Hispanic. All participants had learned English before age 5, reported normal or corrected-to-normal vision, and resided in the continental U.S. or a U.S. territory (Fig. A1).

Recently, concern over the integrity of studies conducted using MTurk samples has grown in the research community (Dennis, Goodson, & Pearson, 2018). Specifically, two threats have been identified: contamination from automated "bots" that respond to surveys, and the use of virtual private servers to mask the true location of participants, which undermines conventional geographic screening methods. To ensure data integrity, IP addresses and geolocations of all participants were checked for duplication, a telltale sign of virtual private server usage. Furthermore, responses to open-ended questions in the post-task questionnaires were carefully screened for signs of automation, such as being irrelevant, incoherent, or overly vague. Using these criteria, no participants were flagged as users of virtual private servers or automated software.

2.1.2 Design

The experiment assessed the degree to which learning efficiency generalizes across learning domains (verbal and visuospatial associative learning). Participants completed two tasks sequentially: a Lithuanian-English task that involved learning the English translations of Lithuanian words, and a task that required learning the locations of objects. Task order was counterbalanced across participants.

2.1.3 Stimuli

In the Lithuanian-English task, stimuli included 28 Lithuanian-English paired associates, a subset of items used in prior norms (Grimaldi, Pyc, & Rawson, 2010; Zerr et al., 2018). Each pair consisted of a Lithuanian noun and its English translation (e.g., LIETUS – RAIN; refer to Table A1 for the complete set). Lithuanian is an ideal language to use to investigate paired-associate learning because it is unfamiliar to most native English speakers, belongs to a separate language family from English and Romance languages that are commonly taught in school (thus minimizing the occurrence of cognates and false friends), and contains the same alphabet as English, obviating transliteration difficulties (Nelson & Dunlosky, 1994; Zerr et al., 2018). All typographic ligatures and diacritical marks were removed from Lithuanian words to ensure that they could be encoded with English phonology (Nelson et al., 2016). Selected English words had a concreteness rating between 500 and 700 per the MRC Psycholinguistic Database (Coltheart, 1981). Additionally, the English words ranged from 3-8 characters in length (M = 4.5), 1-2 syllables (M = 1.2), and 6.8-11.6 logarithmic frequency (M = 10.1) as computed by the English Lexicon Project Database (Balota et al., 2007). Lithuanian words ranged from 4-9 characters in length (M = 6.2) and contained 2-4 syllables (M = 2.6). Word pairs were presented in all capitalized black letters using sans-serif, 27-point font on a white background.

Stimulus presentation code was adapted with permission from Zerr and McDermott (2019) and was written using jsPsych, a Javascript library for running behavioral experiments in a web browser (de Leeuw, 2015).

In the object locations task, images of 28 everyday objects were presented within a circle (Fig. 1; Fig. 2A). To mitigate confusability, objects were chosen to be semantically and perceptually distinct. Images were obtained from a stock image website, www.freeimages.com, and from Google Images. They were exported as 60 $\times$ 60 pixel JPEGs and cropped tightly to reduce excess white space at the periphery. For each participant, the center x- and y-coordinates of objects were randomly generated, with the constraint that objects not overlap with each other, the circumference of the circle, or a 50 $\times$ 50 pixel fixation cross at the circle center. The circle measured 900 pixels in diameter. The object locations task program was modified from custom Javascript code provided by Timothy Lew that was used in Lew, Pashler, and Vul (2016).

Figure 1. The 28 objects used in the object locations task. From top left to bottom right: boot, die, hat, chair, camera, fan, clock, key, bowl, comb, teapot, glasses, bag, lamp, bike, toaster, suitcase, mailbox, scissors, helmet, book, coin, umbrella, headphones, cake, plant, sponge, apple.

2.1.4 Procedure

Participants completed the Lithuanian-English and object locations tasks (hereby abbreviated to LET and OLT, respectively) in sequence, with task order counterbalanced across participants. The LET consisted of three phases: initial study, iterative cued-recall tests, and a delayed final test. Participants were first informed that they would be presented with a list of 28 Lithuanian words paired with their English translations and were instructed to learn these for later cued-recall tests. Subsequently, during the initial study phase, the word pairs were displayed one at a time for 5 s each and separated by a 1 s interstimulus interval (ISI). Presentation order of the word pairs was randomized for each participant. After each had been shown once, participants completed an iterative series of cued-recall tests. The word pairs were rerandomized before each test block to negate serial memory processing and item order effects. A Lithuanian cue was provided (e.g., VANDUO) and the corresponding English translation needed to be typed (e.g., WATER) within 5 s. Responses were deemed correct if either the full English word or at least the first three correct letters (but no incorrect ones) were provided. For instance, if VANDUO were the cue, WAT, WATE, or WATER would be marked correct, but not SWAT or

WATT. After 5 s, irrespective of the accuracy of the response, the correct English translation was shown for 1 s. Each trial within a test block was padded with a 1 s ISI. Lithuanian cues that were correctly answered were dropped from subsequent test blocks, and testing proceeded until every word pair had been correctly recalled exactly once. Thus, as more words were translated accurately, future test blocks became shorter and tested only those word pairs that had been previously missed. This single-trial dropout procedure features a crucial quality that makes it desirable for multi-trial testing. Namely, it ensures that each item is correctly recalled precisely one time, thereby preventing overlearning, which is a manipulation that is known to boost retention (Driskell, Willis, & Cooper, 1992). Moreover, it equates participants in number of correct recalls for each test item (but see Kyllonen & Tirre, 1988 for a discussion of the complexities of equating learning across people).

Participants performed 30 seconds of simple arithmetic problems involving two operands (e.g., 43 – 12 = ?) between test blocks to occupy working memory and prevent rehearsal of the word pairs. Iterative testing terminated once criterion was reached (i.e., all word pairs had been recalled once). To limit the maximum task duration and avoid severe cognitive fatigue, participants who failed to reach criterion within 25 test blocks automatically exited the study prematurely and were compensated for their time. Following the recall tests, participants who had reached criterion restudied all 28 word pairs for 5 s each as in the initial study phase, although in a newly randomized order. They then played 60 s of the puzzle game Tetris before completing a final cued-recall test that contained all 28 Lithuanian-English pairs. This final test was identical to the first test block in all respects. Finally, participants responded to a post-task Likert-type questionnaire that assessed subjective task difficulty, effort, focus, and strategy use.

To encourage honest responding, it was repeatedly emphasized that answers would not affect compensation.

The object locations task (OLT) was structured analogously to the LET, with an initial study phase, iterative cued-recall tests until criterion was reached, and a final test. In the study phase, participants viewed 28 object images located within a circle in sequence. They were instructed to remember each object location, with the name of each object displayed in the top left (Fig. 2A). Each object was presented for 5 s. Pilot testing showed that the OLT generally took longer to complete than the LET, and so ISIs were omitted from the OLT in the interest of time. In the main testing phase, participants were cued to recall the location of each object indicated by an image and name in the top left (Fig. 2B). To respond, they clicked a location within the circle and a 50-pixel diameter crosshair immediately appeared at the selected location. Participants were granted 5 s to respond to each object. Response accuracy was assessed by whether the crosshair overlapped with the object image. Because the objects were square 60×60 pixel images and the crosshair was modeled as a round object, the distance threshold for correct responses varied depending on the position of the clicked location relative to the object. If the clicked location was perfectly orthogonal to the center of the object image, the threshold was 55 pixels; if the clicked location was perfectly diagonal to the object center, the threshold was approximately 67 pixels. After a location was clicked, the true location of the object appeared for 1 s. Feedback for response accuracy was conveyed via the color of the crosshair, which turned red for incorrect and blue for correct responses (Fig. 2C and 2D).

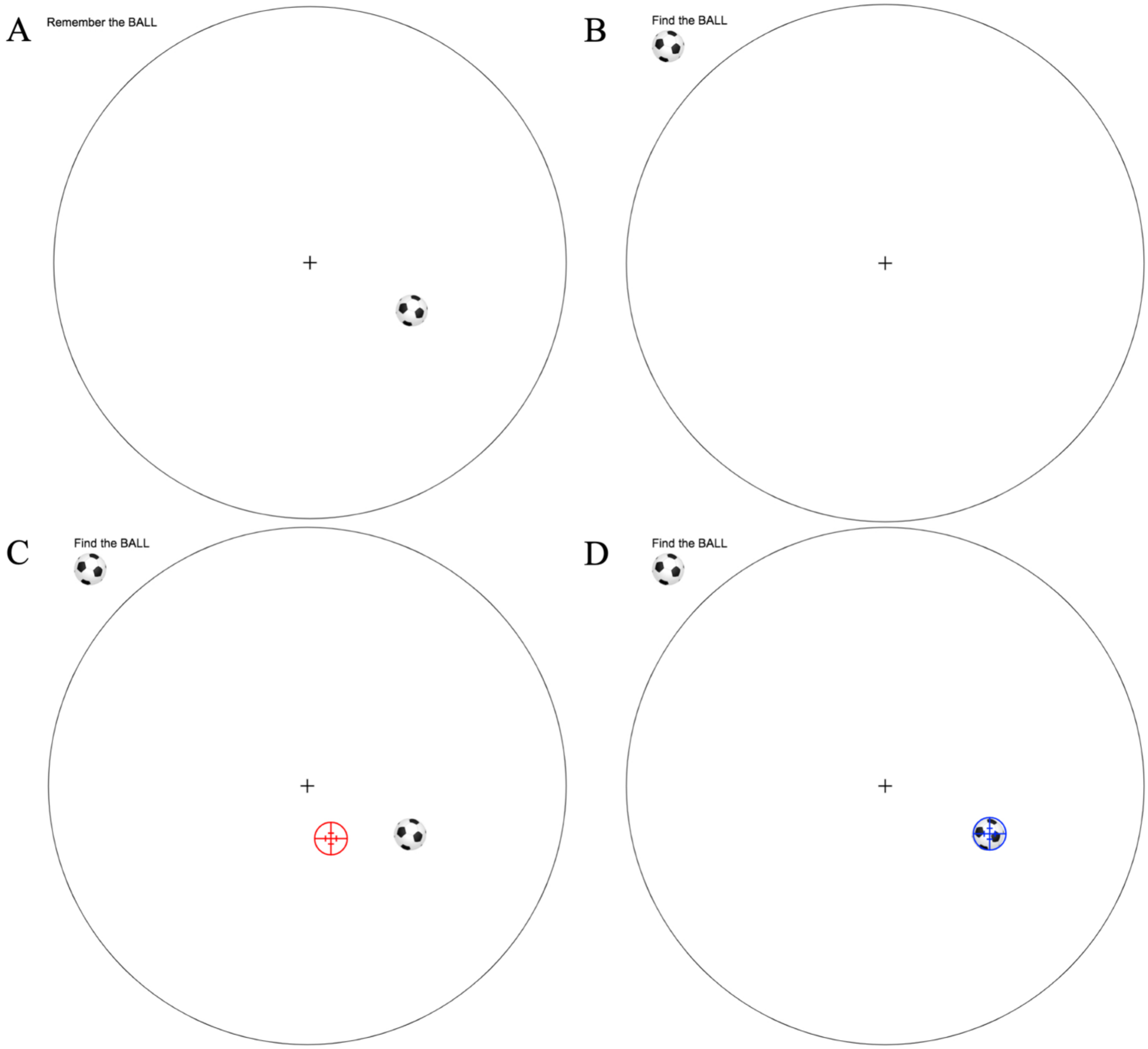

Figure 2. Trial sequence for the object locations task. (A) Participants are instructed to remember the locations of objects in the training phase. (B) During testing, participants are cued to recall each object. (C) Feedback for incorrect responses was provided in the form of a red crosshair at the clicked location. (D) A correct response was designated with a blue crosshair.

As in the LET, objects that were correctly recalled once were dropped from subsequent testing blocks, and testing proceeded until all object locations were dropped. Thirty seconds of addition and subtraction problems were interleaved between test blocks to prevent maintenance of object locations in working memory. Once criterion was reached (correct recall of each object location precisely one time), participants restudied all object locations as in the initial study phase and then played Tetris for 60 s. A final cued-recall test of all 28 object locations was administered; this test was identical to the first test block. After completing the OLT, participants answered a post-task questionnaire, distinct from the one administered after the LET, that collected basic demographic information and probed subjective task difficulty, subjective performance, effort, focus, and strategy use.

Because participants completed the study within their own web browser rather than in a lab setting, display size, display resolution, and viewing distance were not controlled. However, participants were barred from using smartphones or tablets, and they were instructed to maximize their browser window to ensure they could see the totality of the circle and all stimuli.

2.2 Results

In the present analysis, efficient learning is defined as learning that is both fast and durable. Mirroring the approach taken by Nelson et al. (2016) and Zerr et al. (2018), it is operationalized as a composite of three measures: the number of items correctly recalled on Test 1, the number of tests required to reach criterion performance, and the number of items correctly recalled on the Final Test. These three subcomponents of learning efficiency were *z*-score standardized and averaged together to yield a Learning Efficiency Score (LES) for each task. Tests to Criterion was reverse scored in this calculation as higher scores indicate slower, and therefore less efficient, learning.

Descriptive statistics for both the Lithuanian-English and Object-Locations tasks (LET and OLT) are presented in Table 1. Consistent with pilot data collected in the lab, learning object locations proved more difficult than learning Lithuanian words. Comparing the LET to the OLT, participants recalled more words than objects in the initial test, $M_D = 4.0$, $t(111) = 6.84$, $p < .001$, $CI_{95} = [2.8, 5.2]$, took fewer tests to reach criterion performance, $M_D = -6.6$, $t(111) = -15.65$, $p <$

.001, CI_{95} = [-7.4, -5.8], and exhibited greater recall on the final test, $M_D = 7.6$, $t(111) = 13.95$, $p < .001$, CI_{95} = [6.5, 8.7].

From the beginning of the initial study phase to the end of the final recall test, including instructions presented between phases and the 60 s Tetris game, participants spent an average of 23.8 minutes on the LET ($SD = 6.0$, range = 14.6-41.9) and 22.9 minutes on the OLT ($SD = 5.2$, range = 13.4-36.3). Despite taking nearly twice as many tests to reach criterion on the OLT compared to the LET, a Wilcoxon Signed-ranks test revealed that time on task did not significantly differ between the tasks, $M_D = 0.9$, $Z = -0.51$, $p = .30$. This discrepancy is likely attributable to trial timing differences. Whereas LET test trials advanced every 5 s and were separated by a 1 s ISI, OLT trials advanced as soon as a response was made or after 5 s if no response was given, with an average trial time of 2080 ms ($SD = 974$), and had no ISI. Total time on task highly correlated with tests to criterion on both the LET, $r = .87$, $p < .001$, CI_{95} = [.81, .91], and the OLT, $r = .81$, $p < .001$, CI_{95} = [.74, .87].

Table 1 Descriptive statistics of the learning efficiency metrics for the Lithuanian-English and object locations tasks.

Task	Measure	Mean	Median	*SD*	Min	Max
Lithuanian-English	Test 1 Recall	9.0	8.0	6.0	1	24
	Tests to Criterion	6.6	6.0	2.8	2	16
	Items to Criterion	83.5	75	36.7	34	212
	Final Test Recall	19.6	20	5.6	2	28
Object Locations	Test 1 Recall	5.0	4	3.2	0	15
	Tests to Criterion	13.2	12	4.2	6	25
	Items to Criterion	152.8	144.5	54.2	55	360
	Final Test Recall	12.0	12.0	4.1	3	22
	Final Test Error (Pixels)	119.7	110.2	45	50.0	228.5

No effects of task order were found on any of the learning efficiency measures. Participants who completed the Lithuanian-English task first did not perform significantly differently than those who completed it second. Specifically, scores did not differ on Lithuanian Test 1, M_D = -1.71, M-W U = 1280, p = .095, Lithuanian Tests to Criterion, M_D = 0.56, M-W U = 1778, p = .21, Lithuanian Final Test, M_D = -1.37, M-W U = 1335, p = .18, objects Test 1, M_D = -0.21, M-W U = 1556, p = .96, objects Tests to Criterion, M_D = 0.29, M-W U = 1581, p = .93, or objects Final Test, M_D = -0.79, $t(111)$ = -1.02 , p = .31.

2.2.1 Participant Characteristics

LE Scores did not relate to either participant age or years of education on either the Lithuanian or objects tasks (ps > .05). Welch two sample t-tests indicated that learning efficiency did not differ between males and females on the Lithuanian, $t(108.49)$ = 0.08, p = .94, and objects tasks, $t(94.93)$ = -0.47, p = .64.

2.2.2 Avoidance of Ceiling Effects

Memory measures designed to assess individual differences are threatened by ceiling effects, which attenuate variability and reduce the reliability and validity of a test (Uttl, 2005). A key advantage of the LET over widely used standardized memory measures such as the Weschler Memory Scales, Rey Auditory Verbal Learning Test, and the California Verbal Learning Test is that it is calibrated for healthy, young adults to sidestep this source of measurement error (Zerr et al., 2018). Accordingly, a precondition to examining the generalizability of learning efficiency is to verify that the LET and OLT data are not compromised by a restricted range.

Uttl (2005) advises that a test is not overly burdened by ceiling effects when the mean of a measure is at least 1.5 standard deviations from the maximum score. The maximum or optimum possible scores for the learning efficiency subcomponents are 28 items recalled on Test

1, 1 Test to Criterion, and 28 items recalled on the Final Test. On the LET, the Test 1 and Tests to Criterion measures satisfied the $\geq$1.5 *SD* heuristic, with mean scores 3.2 and 2.0 *SD* from the optimum. The Final Test mean narrowly met this standard at 1.5 *SD* from ceiling. Five participants (4.4%) recalled all 28 words on the Final Test. The OLT, being a comparatively more challenging task, was completely devoid of ceiling effects. Test 1, Tests to Criterion, and Final Test means were 7.2, 2.9, and 3.9 *SD*, respectively, off ceiling.

2.2.3 Relation Between Learning Rate and Retention

All learning efficiency submeasures were intercorrelated in the LET, replicating past findings (Becker, 2018; Nelson et al., 2016; Zerr et al., 2018). Participants who recalled more on the initial test learned the Lithuanian-English pairs more quickly as indexed by Tests to Criterion, $r = -.65$, $p < .001$, $CI_{95} = [-.75, -.53]$ (Fig. 3B). Performance on the initial test tracked retention on the final test, $r = .44$, $p < .001$, $CI_{95} = [.28, .58]$ (Fig. 3C). Critically, faster learners retained more on the final test, $r = -.69$, $p < .001$, $CI_{95} = [-.77, -.57]$ (Fig. 3D).

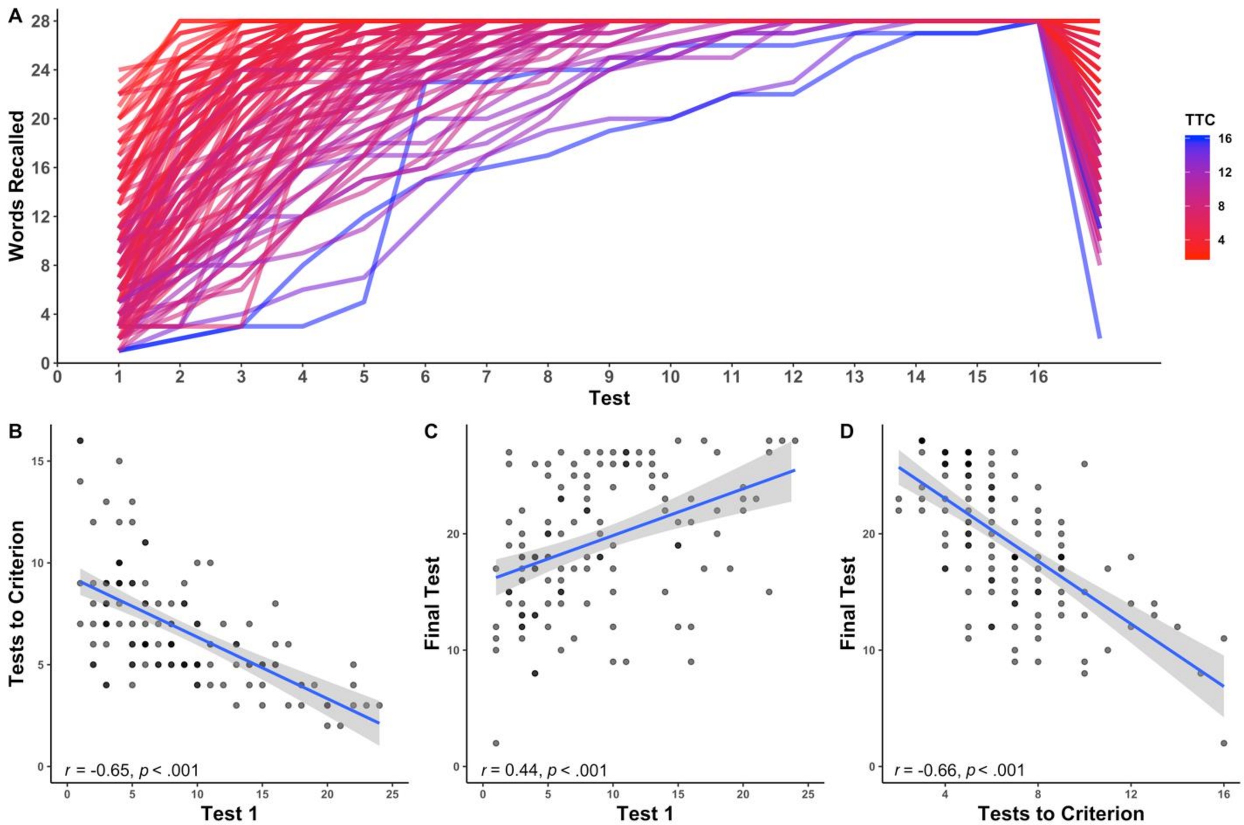

Figure 3. Faster learners retained more in the Lithuanian-English task. (A) Learning curves are plotted for each participant. The fastest learners, represented by the red traces, tended to recall the most at final test. TTC = Tests to Criterion. (B-D) The three submeasures of learning efficiency, Test 1 recall, Tests to Criterion, and Final Test recall, robustly intercorrelate with each other. Shaded regions indicate 95% confidence intervals.

The same overall pattern of associations was found for the object locations task. Participants who recalled more objects on the initial test reached criterion more quickly, $r = -.36$, $p < .001$, $CI_{95} = [-.51, -.19]$ (Fig. 4B) and had better retention on the final test, $r = .52$, $p < .001$, $CI_{95} = [.37, .65]$ (Fig. 4C). As in the LET, faster learners remembered more on the final test, $r = -.53$, $p < .001$, $CI_{95} = [-.65, -.38]$ (Fig. 4D). The complete correlation matrix of the learning efficiency measures is presented in Table 2.

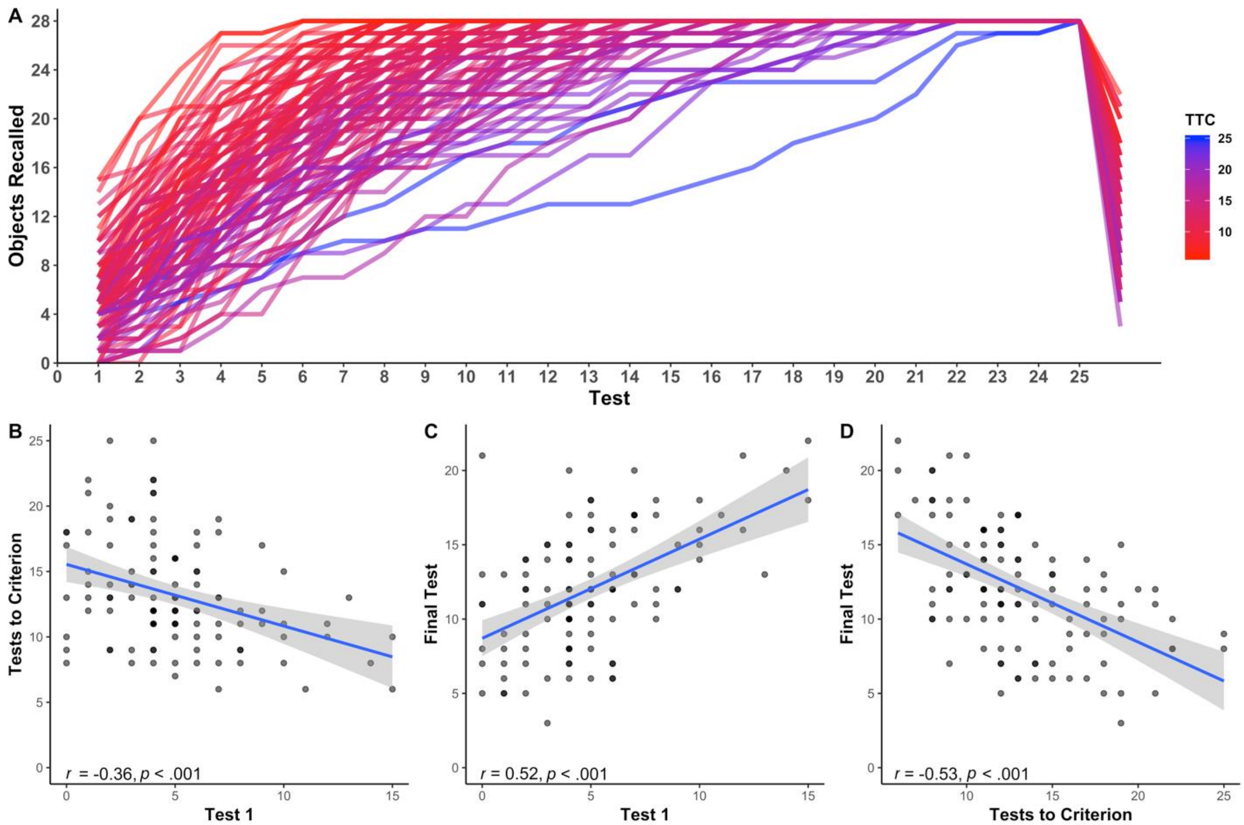

Figure 4. Faster learners retained more in the object locations task. (A) Learning curves are plotted for each participant. The fastest learners, represented by the red traces, tended to recall the most at the final test. TTC = Tests to Criterion. (B-D) The three submeasures of learning efficiency, Test 1 recall, Tests to Criterion, and Final Test recall, robustly intercorrelate with each other. Shaded regions indicate 95% confidence intervals.

Table 2 Correlation matrix of the learning efficiency measures for verbal and visuospatial tasks.

Variable	1	2	3	4	5	6	7	8
Lithuanian-English								
1 Test 1 Recall								
2 Tests to Criterion	-.65**							
3 Final Test Recall	.44**	-.69**						
4 LES	.82**	-.91**	.83**					
Object Locations								
5 Test 1 Recall	.19*	-.24*	.30**	.29**				
6 Tests to Criterion	-.14	.23*	-.25**	-.24*	-.36**			
7 Final Test Recall	.18	-.24*	.33**	.29**	.52**	-.53**		
8 LES	.21*	-.29**	.37**	.34**	.78**	-.78**	.85**	
9 Final Test Error	-.21*	.26**	-.35**	-.32**	-.50**	.43**	-.82**	-.73**

Note. * indicates $p < .05$. ** indicates $p < .01$. LES = Learning Efficiency Score. Correlations are reported as Pearson's correlation coefficients. Correlations of the same metrics between the LET and OLT, representing the generalizability of learning efficiency, are highlighted in blue.

2.2.4 Generalizability of Learning Efficiency

A central question of the present study is whether learning efficiency generalizes across learning domains. To what extent are fast and retentive verbal learners also fast and retentive visuospatial learners? Learning performance as indexed by the learning efficiency submeasures correlated across tasks (Fig. 5A-C), including Test 1 recall, $r = .19$, $p = .04$, $CI_{95} = [.009, .37]$, Tests to Criterion, $r = .23$, $p = .017$, $CI_{95} = [.04, .39]$, and Final Test recall, $r = .33$, $p < .001$, $CI_{95} = [.16, .49]$. The overall Learning Efficiency Scores (LES), the average of the three *z*-score standardized submeasures, also correlated across tasks, $r = .34$, $p < .001$, $CI_{95} = [.17, .49]$ (Fig. 5D).

Although observed between-task correlations are small to medium effects according to conventional interpretations of effect sizes in the social sciences (Cohen, 1992; Ferguson, 2009), that there is a consistent relationship in learning performance across tasks with divergent demands is itself informative.

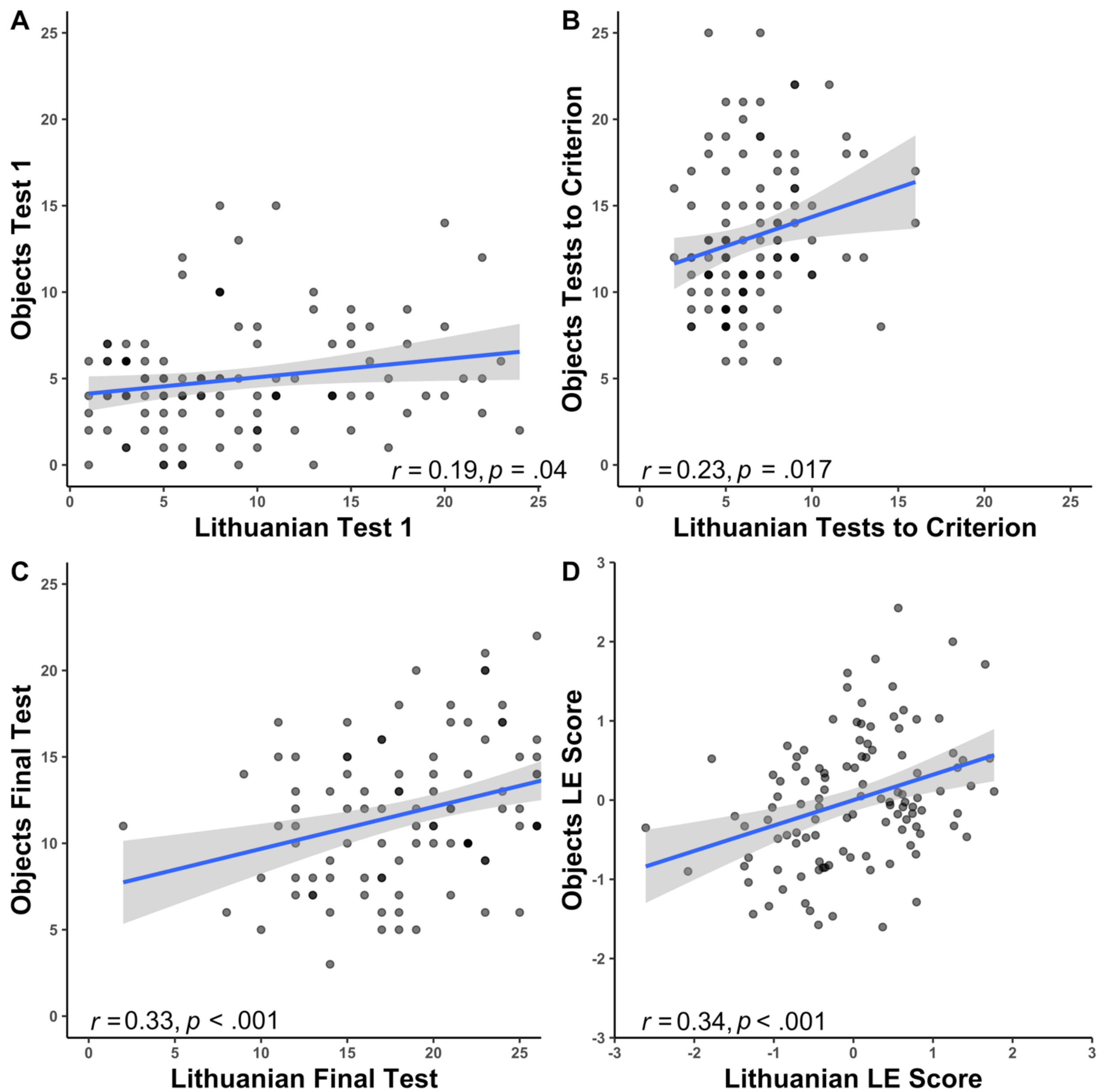

Figure 5. Learning efficiency generalizes across the Lithuanian-English and object locations tasks. Each of the learning efficiency submeasures, along with Learning Efficiency Scores themselves, correlate between the two tasks. Shaded regions indicate 95% confidence intervals.

2.2.5 Reliability of Learning Efficiency

To reliably detect effects in experimental research, cognitive tasks should exhibit low between-participant variability. This is because in experimental paradigms, between-participant variability is a nuisance factor that masks group differences. Antithetically, in correlational research it is necessary to have a high ratio of between-participant to within-participant (or between-measure) variability to reliably measure individual differences. The Intraclass Correlation Coefficient (ICC) can be used to quantify the degree to which different measures reliably rank-order people (Hedge, Powell, & Sumner, 2017). Another interpretation of the ICC is the proportion of total variance accounted for by between-participant variability, expressed as a value ranging from 0 to 1. An ICC of 1 reflects complete between-participant variability and no variability between measures, while an ICC of 0 reflects no between-participant variability but high variability between measures.

To assess the reliability between the Lithuanian-English and object locations learning efficiency estimates, a two-way random ICC (corresponding to model type ICC2k from Shrout and Fleiss, 1979) was calculated for participants' LE Scores from the two tasks. Refer to Field (2005) and Koo and Li (2016) for an in-depth discussion of ICC models. The ICC function from the R package *psych* (Version 1.8.12; Revelle, 2018) was used. The computed ICC between the two tasks was .51, $F(2, 111) = 2.05$, $p < .001$, $CI_{95} = [.29, .66]$, indicating that approximately half of the variance in learning efficiency across the two measures is attributable to between-participant variability. Thus, participants' learning efficiency generalizes to a large degree even across two highly disparate tasks.

2.2.6 Trials to Criterion

Tests to Criterion is a somewhat coarse measure of learning rate. In principle, two participants could reach criterion after the same number of tests yet complete a substantially different number of individual trials. To illustrate this point, imagine two learners, A and B, who both reach criterion after five tests (Table 3). Learner A recalls 24 items on Test 1 but only a single item on subsequent tests; conversely, learner B recalls only 8 items on Test 1 but 5 items on Tests 2-5. Both learners reach criterion after the same number of tests, but learner B takes over twice as many trials to do so and views more instances of individual items. Thus, Trials to Criterion is a more granular measure than Tests to Criterion because it better captures variability in learning rate across tests within a participant.

In the Lithuanian-English task, learners who reached criterion after fewer trials had greater Test 1 performance, $r = -.78$, $p < .001$, $CI_{95} = [-.84, -.69]$ (Fig. 6A), and retained more on the Final Test, $r = -0.62$, $p < .001$, $CI_{95} = [-.72, -.49]$ (Fig. 6B). Similarly, in the object locations task, Trials to Criterion correlated with Test 1, $r = -.64$, $p < .001$, $CI_{95} = [-.74, -.52]$ (Fig. 7A), and with Final Test, $r = -.58$, $p < .001$, $CI_{95} = [-.69, -.45]$ (Fig. 7B). Although the correlation values differ compared to those computed using Tests to Criterion, the direction of the associations remain consistent, reinforcing the finding that faster learners retain more. This result is not surprising given that Trials to Criterion is highly correlated with Tests to Criterion on both the LET, $r = .92$, $p < .001$, $CI_{95} = [.89, .95]$, and the OLT, $r = .86$, $p < .001$, $CI_{95} = [.80, .90]$. Although these two measures of learning speed can diverge in principle, in practice they correspond closely.

Table 3 Items recalled and trials completed per test for two hypothetical participants.

	Learner A		Learner B	
	Items Recalled	Trials Completed	Items Recalled	Trials Completed
Test 1	24	28	8	28
Test 2	1	4	5	20
Test 3	1	3	5	15
Test 4	1	2	5	10
Test 5	1	1	5	5
Total	**28**	**38**	**28**	**78**

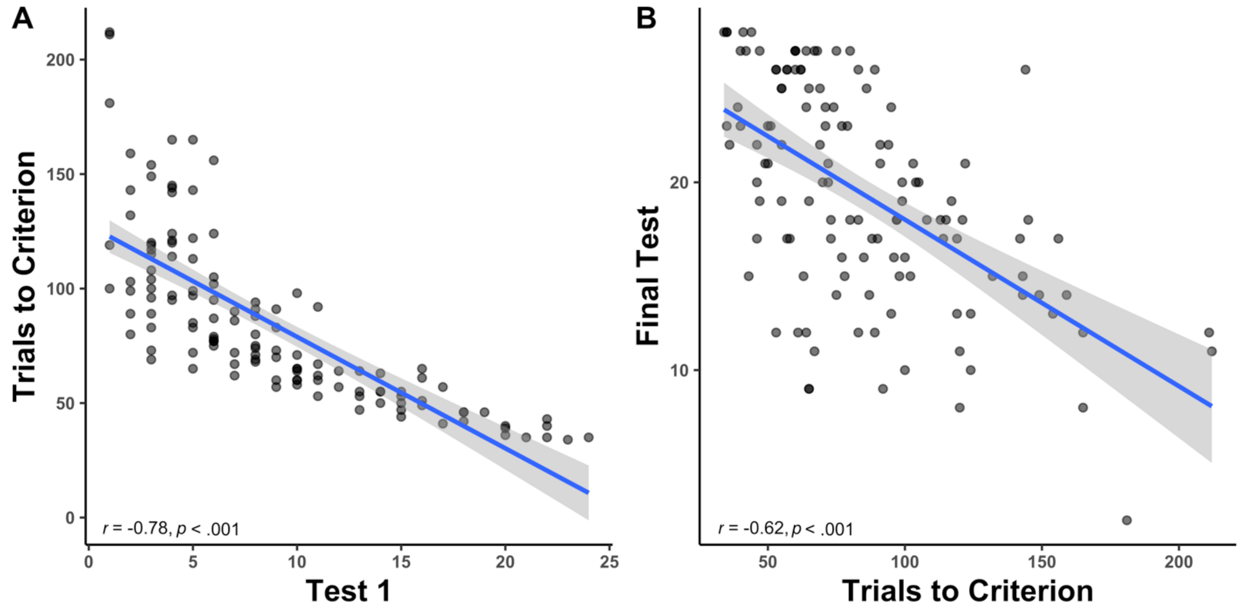

Figure 6. The Trials to Criterion measure correlates with the other learning efficiency submeasures in the Lithuanian-English task. Shaded regions indicate 95% confidence intervals.

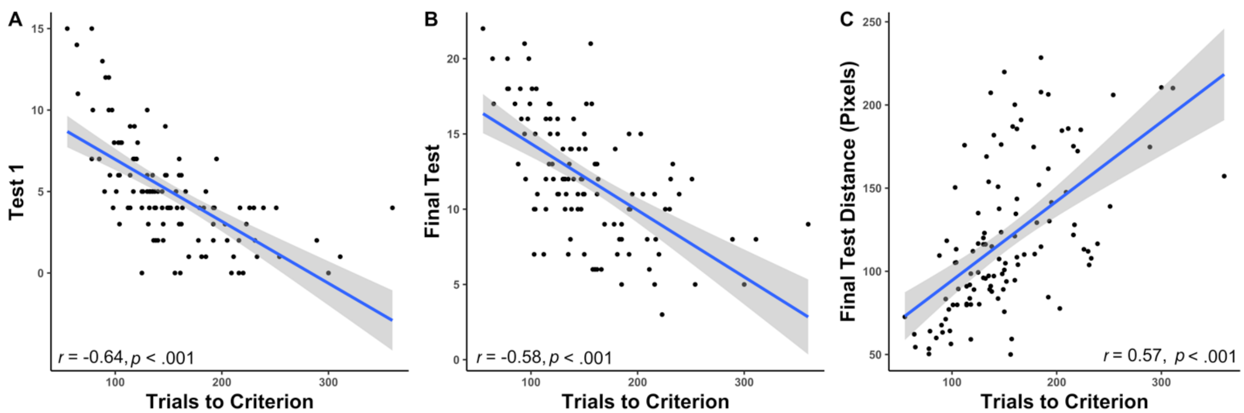

Figure 7. The Trials to Criterion measure correlates with the other learning efficiency submeasures in the object locations task. Shaded regions indicate 95% confidence intervals.

2.2.7 Spatial Precision

To succeed on the object locations task, participants needed to associate objects with precise spatial coordinates. Spatial precision, operationalized as the Euclidean distance in pixels between selected and target coordinates for each object, is a more fine-grained measure of learning and retention than a binary correct/incorrect classification. A participant may have repeatedly missed an object's location yet nevertheless progressively become more precise across blocks as they refined their spatial representation. This subthreshold learning can only be captured by precision data and not accuracy (correct/incorrect).

Averaging across participants, responses were more precise on the Final Test (*Mdn* = 255.0) compared to Test 1 (*Mdn* = 110.2), $Z = -9.2$, $p < .001$ (Fig 8.). See Figure A2 for the trial distributions of precision on Test 1 and the Final Test and Figure A3 for across-participant block means of precision. The improvement in precision from the first to final test varied considerably between participants ($M = 129.9$, $SD = 64.7$) (Fig. 9).

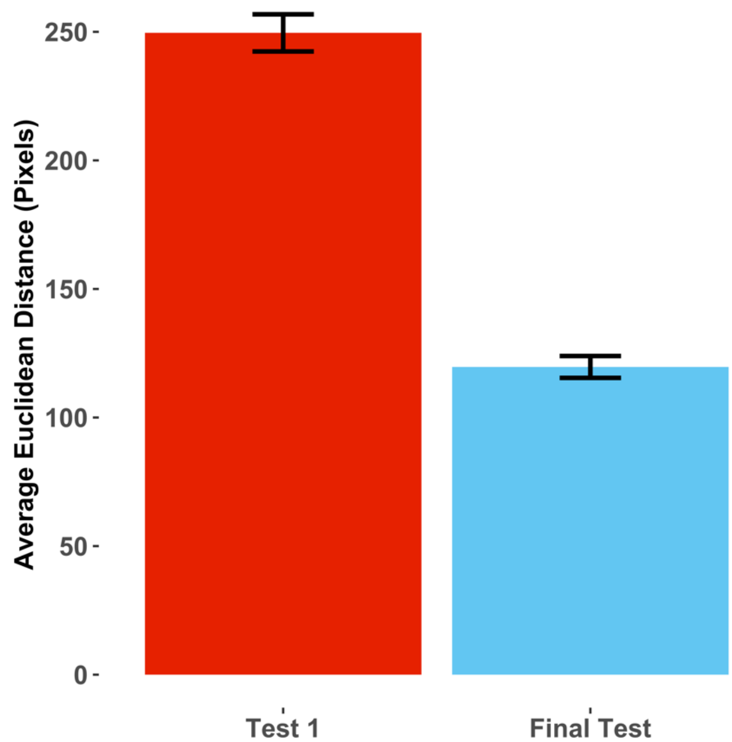

Figure 8. Responses to object locations become more precise on the Final Test relative to Test 1 on average. Error bars represent the standard error.

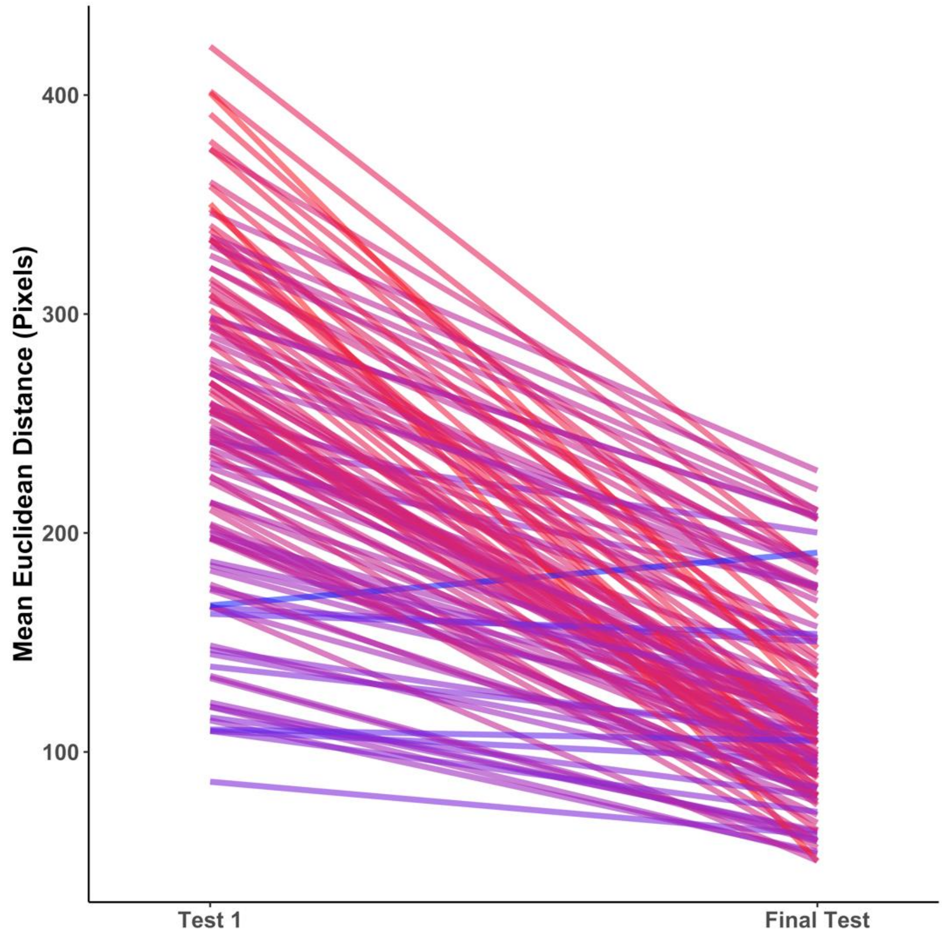

Figure 9. Spaghetti plot showing change in the precision of object location responses between Test 1 and the Final Test across all participants. Lines are colored such that red represents a greater improvement and blue a lesser improvement.

Final Test error, in pixels, was found to correlate with Test 1 recall, $r = -.50$, $p < .001$, $CI_{95} = [-.63, -.35]$ (Fig. 10A), Tests to Criterion, $r = .43$, $p < .001$, $CI_{95} = [.27, .57]$ (Fig. 10B), and Final Test recall, $r = -.82$, $p < .001$, $CI_{95} = [-.88, -.75]$, suggesting that it may be another viable measure to characterize learning efficiency. Additionally, Final Test error weakly to moderately correlated with the Lithuanian learning efficiency metrics, further buttressing the generalizability of learning efficiency across domains (Table 2).

One potential limitation of the precision measure is that, because trials automatically advanced to the next object after five seconds, participants could selectively not respond to objects whose locations they were unsure of. Such selective responding would artificially inflate precision scores. However, an inspection of the data suggest that this concern is unwarranted. Across all participants, the mean non-response rate for all trials was 2.4% (*SD* = 4.9%). On the final test specifically, responses were provided for an average of 27.8 (*SD* = 0.94) out of 28 objects. When the number of objects responded to on the final test was included as a covariate to the previously reported correlations between Final Test error and the other learning efficiency submeasures, the magnitude of the correlations did not decrease.

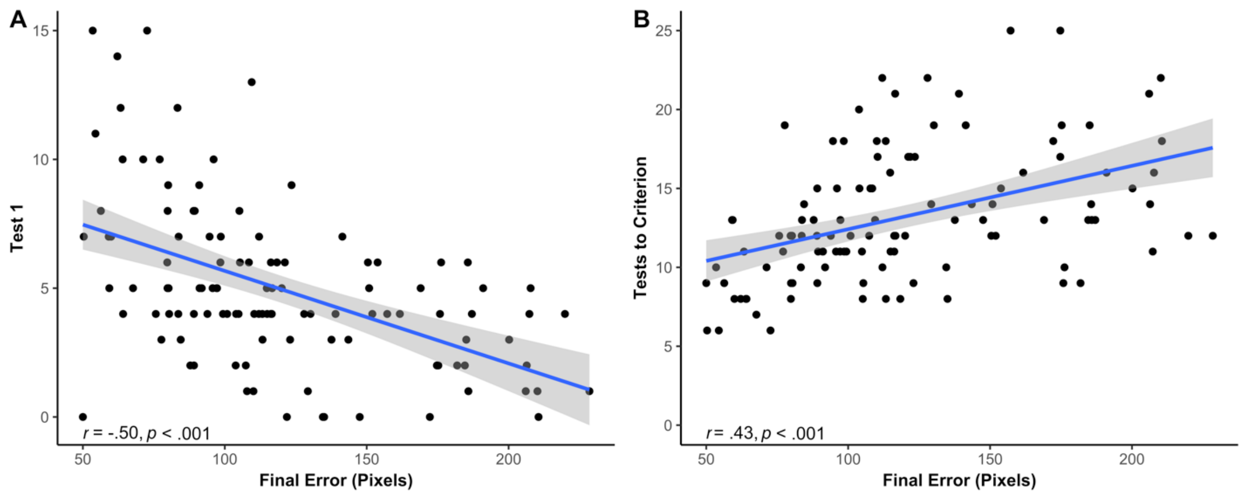

Figure 10. Precision on the final test of the object locations task correlates with the number of objects recalled on Test 1 and Tests to Criterion. Shaded regions indicate 95% confidence intervals.

2.2.8 Learning Strategies

Differences in learning strategy selection and application across participants may be one of the factors that accounts for the association between learning rate and retention. Do efficient learners systematically rely on learning strategies more than less efficient ones? Are there particular strategies that high performers gravitate towards? To shed light on these and related questions, participants responded to a Likert-type questionnaire interrogating their strategy use after the Lithuanian-English task. The learning strategy questions were taken from Zerr (2017) and were originally adapted from McDaniel and Kearney (1984). The complete question list is presented in Table 4.

Table 4 Lithuanian-English learning strategy questions.

Strategy	Question	*M (SD)*
Keyword	How often did you 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?	2.8 (1.1)
Other Language	How often did you think of a word in a different language to link to the Lithuanian and English word?	1.7 (1.0)
Physical	How often did you construct sentences to associate the word pairs that described what you physically saw?	2.4 (1.4)
Repetition	How often did you repeat the two words in a pair together over and over (either in your head or out loud) to commit them to memory?	3.6 (1.2)
Failure	How often did your various strategies not work for helping you learn the word pairs?	2.9 (0.8)
None	How often did you struggle or have difficulty trying to come up with a strategy for learning the word pairs?	3.1 (1.0)
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?	2.9 (1.1)
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?	2.9 (1.1)

Note. Strategy questions are from Zerr (2017) and were originally adapted from McDaniel and Kearney (1984). 1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Usually; 5 = Always.

Participants reporting that their strategies did not work more frequently performed worse on the Lithuanian-English task, $r_S = -.51$, $p < .001$, $CI_{95} = [-.64, -.36]$ (Fig. 11A). Similarly, those who struggled to come up with a strategy more often had lower scores, $r_S = -.59$, $p < .001$, $CI_{95} = [-.70, -.46]$ (Fig. 11B). Additionally, answers to the strategy *Failure* and *None* questions correlated, $r_S = .66$, $p < .001$, $CI_{95} = [.53, .77]$, implying that participants who struggled to come up with strategies tended to use less effective ones and/or implemented them less effectively.

Curiously, the *Physical* strategy (constructing sentences that described what was physically seen) was the only strategy that related to overall task performance, $r_S = .20$, $p = .035$, $CI_{95} = [.006, .38]$ (Fig. 11C). Reliance on the keyword method, other languages as mediators, or repetition was not related to learning efficiency (*p*s > .05). Contrary to expectations, those who claimed to persevere with ineffective strategies did not do worse, and participants reporting frequent strategy switching exhibited no advantage. *Perseverance* scores were negatively correlated with *Switch* scores, $r_S = -.69$, $p < .001$, $CI_{95} = [-.82, -.54]$, indicating that participants were attending to the questionnaire sufficiently to not provide identical answers to opposite questions.

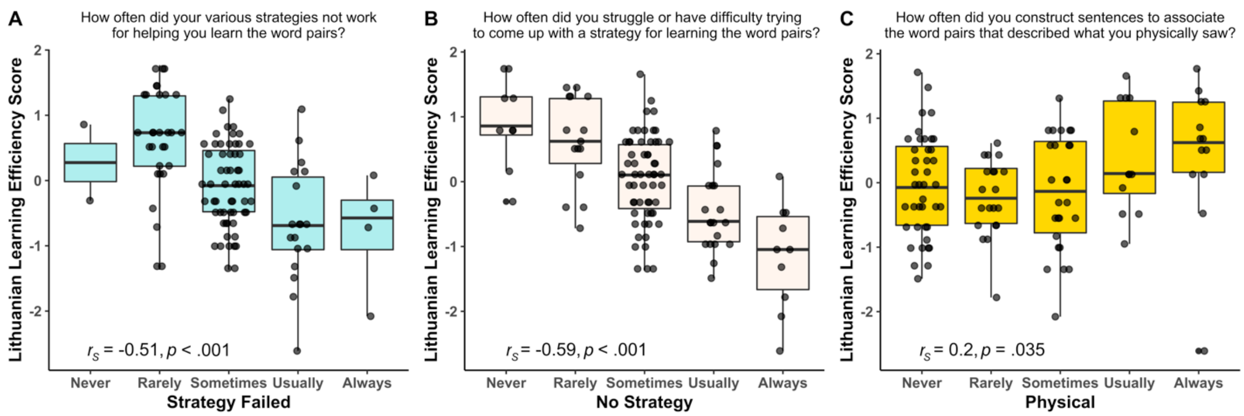

Figure 11. (A and B) Participants who had difficulty finding or implementing effective strategies had lower LE Scores on average. (C) The only strategy associated with task performance was the *Physical* strategy.

Strategy differences were also assessed for the object locations task. Participants were asked to describe any strategies or techniques they used to learn the object locations. Because responses were unstructured and open-ended, a different set of analyses were used than with the Lithuanian-English strategy data. As an initial exploratory procedure to identify common strategies, and in order to examine whether high efficiency learners reported using different approaches than their low efficiency counterparts, a unigram and bigram (i.e., single and double word) frequency analysis was conducted, which was adapted from the *n*-gram analysis reported in Selmeczy and Dobbins (2014).

From the 112 participant sample, 108 supplied a typed description of the strategies they employed. The average response length was 23.8 words (SD = 24.4), and response length was not significantly associated with object Learning Efficiency Scores, $r_S = .17$, p = .07. To prepare the text for analysis, spelling errors were manually corrected and contractions were expanded (e.g., “didn’t” was changed to “did not”). All subsequent text processing was done using the *R* package *tidytext* (Version 0.2; Silge & Robinson, 2016). Each participant’s typed response was tokenized into its constituent unigrams and bigrams. Punctuation was stripped and words were converted to lowercase to facilitate aggregation. In the unigram analysis, stop words (e.g., “I”, “the”, “of”, etc.) were then removed using the *SMART*, *snowball*, and *onix* lexicons. Additionally, inflectional endings were truncated so that only the base form of words were included, a process known as lemmatization. For example, “strategies” was simplified to the singular “strategy,” and “remembered” and “remembering” were reduced to the simple present tense form “remember.” Morphological derivations (e.g., “location” and “locate”) were not combined to conserve lexical category. Stop words and inflected forms were preserved in the bigram analysis because they might contain important contextual information and could be distributed differently across high and low efficiency learners. To limit the influence of individual differences in verbosity, duplicated instances of unigrams and bigrams were counted only once per participant.

The most frequently used words are displayed in Figure 12. Unsurprisingly, terms germane to the task (“object”, “location”, “item”, “circle”, “position”) and its objective (“remember”, “memorize”) feature prominently in participants’ responses. More informative is the occurrence of words associated with specific strategies, including “clock” and “repeat”. Does usage of any of these words discriminate between high and low efficiency learners? A one proportion Z-test was used to determine whether the distribution of each *n*-gram across efficiency levels reliably differed. Specifically, word frequency was compared between learners scoring in the top and bottom quartiles on the objects task. Positive *z*-scores indicate that the word occurred more often in top learners, whereas negative *z*-scores reflect greater usage among bottom learners. The results of this analysis for both unigrams and bigrams are reported in Table 5. Because the word frequency analysis is exploratory, uncorrected *p*s are reported. To limit the number of comparisons, *n*-grams occurring fewer than five times across top and bottom learners were omitted.

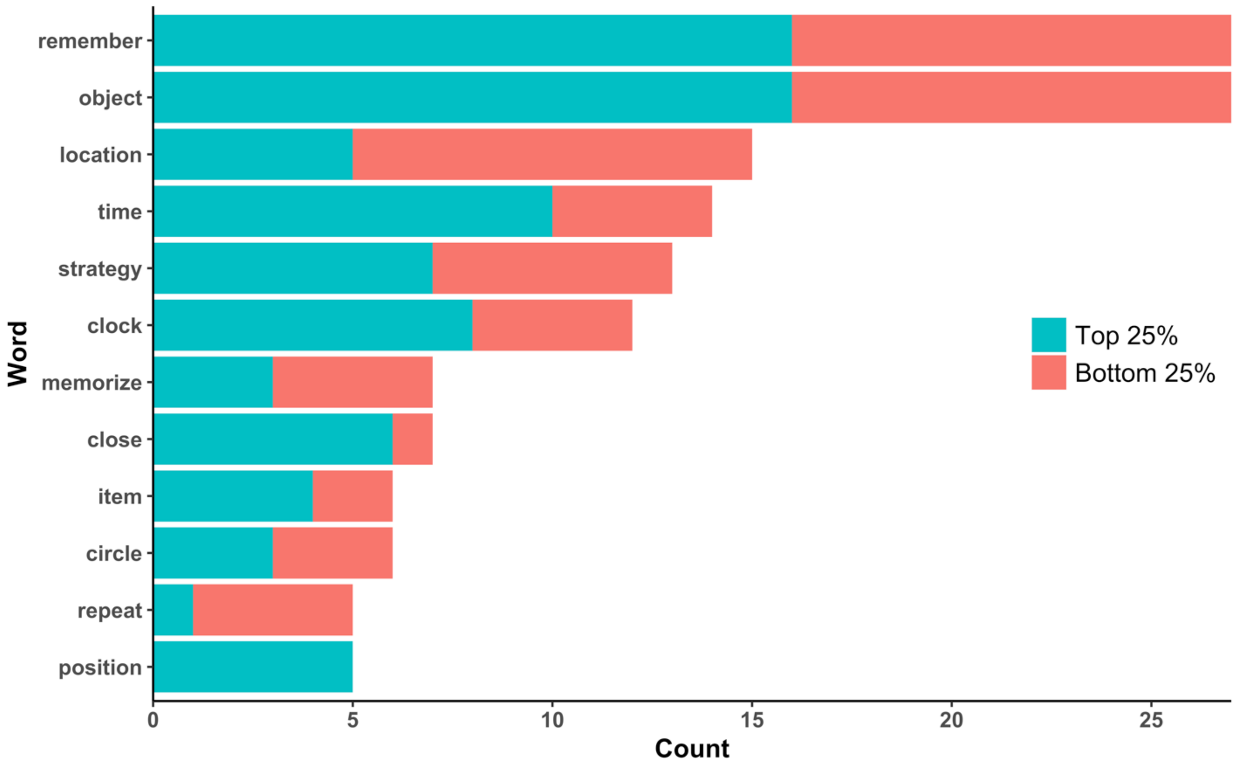

Figure 12. Unigram frequency counts of the most common words for high and low efficiency learners.

Table 5 Unigram and bigram analysis for top and bottom learners.

n-gram	Top 25%	Bottom 25%	*z*	*p*
		Unigrams		
position	5	0	2.24	.025
close	6	1	1.89	.059
time	10	4	1.60	.109
clock	8	4	1.15	.248
object	16	11	0.96	.336
remember	16	11	0.96	.336
item	4	2	0.82	.414
strategy	7	6	0.28	.782
circle	3	3	0.00	1.000

Table 5 (*cont.)*

n-gram	Top 25%	Bottom 25%	*z*	*p*
memorize	3	4	-0.38	.705
location	5	10	-1.29	.197
repeat	1	4	-1.34	.180
		Bigrams		
they were	7	0	2.65	0.008
did not	7	1	2.12	0.034
so i	6	1	1.89	0.059
to remember	15	7	1.71	0.088
i did	4	1	1.34	0.180
i really	4	1	1.34	0.180
remember where	4	1	1.34	0.180
where they	4	1	1.34	0.180
remember the	5	2	1.13	0.257
the objects	5	2	1.13	0.257
just tried	7	4	0.9	0.366
a clock	4	2	0.82	0.414
the items	4	2	0.82	0.414
i tried	11	9	0.45	0.655
i was	3	2	0.45	0.655
to memorize	3	2	0.45	0.655
try to	3	2	0.45	0.655
tried to	17	15	0.35	0.724
to the	5	4	0.33	0.739
of the	5	5	0.00	1.000
the object	4	4	0.00	1.000
the circle	3	3	0.00	1.000
object was	2	3	-0.45	0.655
i just	5	7	-0.58	0.564
in my	1	4	-1.34	0.180
trying to	1	5	-1.63	0.102

Thresholding at an alpha level of .05, the only *n*-grams that occurred significantly more frequently in top relative to bottom learners were the unigram “position” and the bigrams “they were” and “did not.” No *n*-grams were used significantly more frequently by low efficiency learners. After adjusting *p*-values for multiple comparisons using the false discovery rate procedure (Benjamini & Hochberg, 1995), no significant differences remain.

The *n*-gram analysis may have failed to detect differences between the responses of high- and low efficiency learners due to the relatively low frequency counts of the *n*-grams. Additionally, a simplistic *n*-gram comparison is not sensitive to semantics that may be conveyed across complex word strings or lengthy, idiosyncratic descriptions of a particular strategy. In order to capture the holistic meaning of the provided answers, the responses of all participants were read, from which a set of nine ad hoc strategy categories were created. Descriptions of and exemplar responses from each strategy category are listed in Table 6.

Table 6 Ad hoc learning strategy categories.

Strategy	Description	Example Response	Count
Track with Cursor	Following and hovering over object locations with the mouse cursor	"I followed the items with my pointer so I could get a good feel for where the objects were."	7
Complex Associations	Creating complex spatial or semantic associations with objects	"The apple is high on a tree and it was at the top of the circle."	5
Clock	Relating object locations to the positions of analogue clock numerals	"If an object were at the bottom, it would be near 6 o'clock."	19
Spatial Grouping	Grouping objects that cluster close to one another	"I tried to group items together that were in the same area in order to have a rough estimate of where items were located."	8
Relate to Other Object Locations	Relating object locations in reference to other object locations	"I tried to pinpoint the objects according to where they were from one another."	4
Coordinate System or Cardinal Points	Using a non-clock based geometric coordinate system, cardinal points, or other directional markers to remember the object locations	"I tried to just remember if an object was close to the middle or close to the outer ring. Or I would say to myself, 'clock center top.'"	8
Repetition	Repeatedly visualizing object locations	"I kept repeating the locations in my head."	7
Other	Miscellaneous strategies	"I tried to associate close objects to a letter."	21
None	No strategy	"Just tried to remember where they were."	45

Each response was classified as belonging to one or more categories, with the exception of *Other* and *None* responses which were assigned those labels exclusively. Fifty-six participants reported using a single strategy, ten used two strategies, one used three, and forty-five used no strategy or did not provide a response. Participants using the *Track with Cursor* strategy outperformed those using no strategy, M_D = 0.92, $t(7.12) = 2.54$, $p = .038$, CI_{95} = [0.07, 1.77], as did those relying on *Complex Associations*, M_D = 0.86, $t(5.00) = 2.67$, $p = .044$, CI_{95} = [0.03, 1.68]. Mean Learning Efficiency Scores on the objects task for users of each strategy are presented in Figure 13. Additionally, Wilcoxon rank sum tests indicated that participants relying on the *Follow with Cursor*, *Complex Associations*, and *Grouping* strategies recalled the object locations more precisely on the final test relative to those who used no strategy, ps < .05 (see Figure A5).

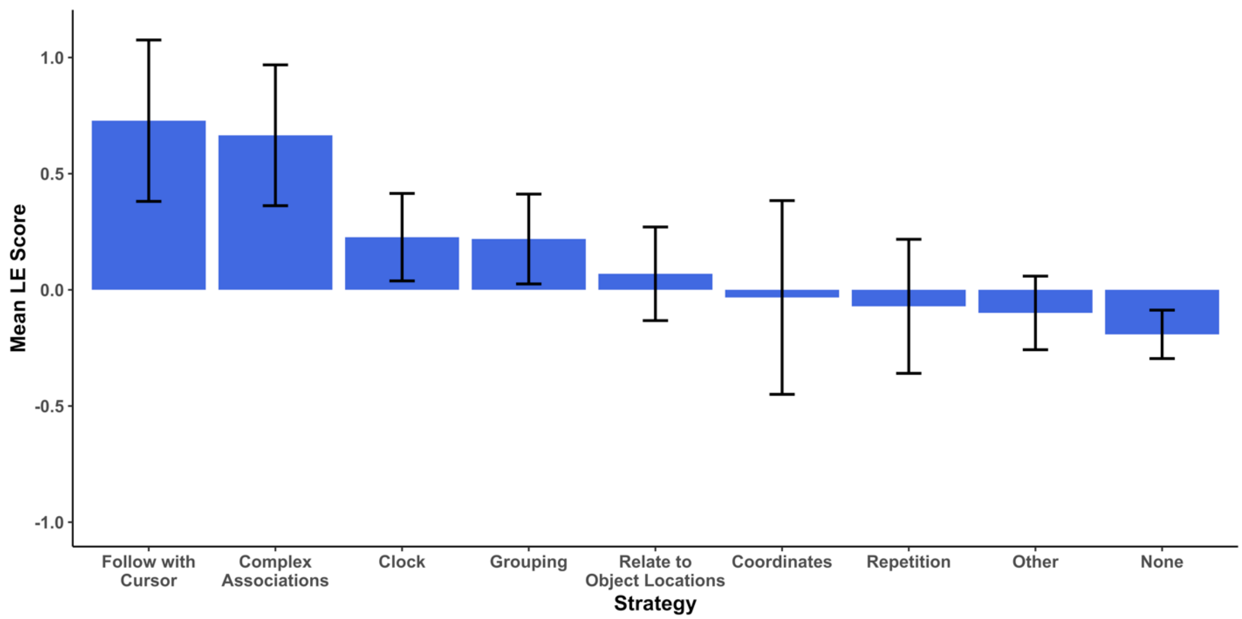

Figure 13. Participants relying on the *Follow with Cursor* and *Complex Associations* strategies outperformed those reporting no strategies. Error bars represent the standard error.

2.2.9 Response Latencies

One factor that is hypothesized to underlie some of the individual variability in learning efficiency is working memory capacity (WMC), and especially attentional control components of working memory (Becker, 2018; Nelson et al., 2016; Zerr et al., 2018). On both free and cued recall tasks, high WMC individuals typically exhibit shorter response times on correct recall trials relative to their low WMC counterparts (Unsworth & Engle, 2007). It is believed that higher WMC enables more contextually-irrelevant information to be discarded during long-term memory searches, reducing the size of search sets and therefore speeding up retrieval times (Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth & Engle, 2007). If WMC variability accounts for a portion of learning efficiency differences, this may be reflected in faster response latencies on correct trials for highly efficient learners.

For the Lithuanian-English task, response times (RTs) were operationalized as the interval, in milliseconds, between the presentation of a Lithuanian cue word and the initial keystroke of the English target. Object locations task RTs were defined as the interval between presentation of the object image and the selection of the object location. A potential concern with collecting behavioral data online is that response times may not be recorded accurately or reliably. However, although it has been found that Javascript-based programs add approximately 25 ms to response time measurements relative to conventional, offline experimental software (e.g., MATLAB's Psychophysics Toolbox), Javascript does not affect the variability of response time distributions (de Leeuw & Motz, 2016). To preclude outliers from unduly influencing analyses, the following removal procedure was carried out: first, responses with latencies below 200 ms were filtered out for being probable anticipatory responses. Following this, RTs were z-score standardized within participants. Finally, trials with standardized RTs more than three standard deviations from a participant's mean were removed, after which RTs were restandardized.

For correct responses on the Final Test of the Lithuanian task, the mean of mean participant RTs was 2090 ms (SD = 536). A Wilcoxon signed rank test revealed that this was significantly faster than the RTs for incorrect responses, M_D = -992 ms, Z = -8.05, $p < .001$. Learning Efficiency Scores negatively correlated with correct trial response times such that for every one standard deviation increase in LE Score, RTs decreased by 127 ms on average, r = -.11, $p < .001$. By contrast, Learning Efficiency Scores were positively correlated with incorrect trial response times such that for every one standard deviation increase in LE Score, RTs increased by 138 ms on average, r = .11, p = .018. Efficient learners' greater error latencies may reflect their propensity to continue searching memory longer in the absence of retrieval success (MacLeod & Nelson, 1984).

On the Final Test of the object locations task, the mean of mean participant RTs was lower for correct than incorrect trials, M_D = -78, Z = -4.15, $p < .001$. Learning Efficiency Scores were weakly positively correlated with RTs for both correct, r = .06, p = .042, and incorrect trials, r = .08, $p < .001$. In contrast with the Lithuanian task, on the objects task the association between learning efficiency and correct RTs is at best equivocal and at worst contradicts the hypothesized result of more efficient learners having reduced latencies. A potential explanation for the unexpected direction of this association is that, in accordance with Fitts' law (Fitts, 1954), it takes longer to make a controlled motor movement to a smaller target area. Participants may have taken slightly longer to position their cursors when they more precisely recalled a location. Indeed, across all test trials precision weakly but significantly correlated with participant-standardized RTs such that more precise responses had longer latencies, $r = -.04$, $p < .001$. The extra time required to make a precise response may have washed out retrieval speed differences.

2.2.10 Self-Assessments

Subjective focus ratings were not associated with LE Scores on the Lithuanian, $r_S = .08$, $p = .39$, or objects tasks, $r_S = .05$, $p = .62$. Similarly, subjective effort was not related to overall Lithuanian performance, $r_S = .13$, $p = .18$, or objects performance, $r_S = -.02$, $p = .82$. However, subjective difficulty negatively correlated with Lithuanian LE Scores, $r_S = -.54$, $p < .001$, and objects LE Scores, $r_S = -.21$, $p = .026$, with participants rating a task as more difficult doing worse. To probe metacognitive awareness, participants were asked to rate their performance on the objects task on a 1-5 rating scale that ranged from “significantly below average” to “significantly above average”; subjective performance ratings were not collected for the Lithuanian task. Subjective performance correlated positively with actual performance, $r_S = .49$, $p < .001$, indicating that participants’ self-assessments were reasonably well calibrated.

Chapter 3: Discussion

The main aim of this project was to test whether learning efficiency generalizes across verbal and visuospatial learning. In a 112 person sample, learning efficiency measures correlated between Lithuanian-English and object locations paired associates tasks, consistent with the hypothesis that learning efficiency is a domain-general ability. As in prior work (Nelson et al., 2016; Zerr et al., 2018), measures of initial learning, tests-to-criterion, and final retention were robustly related within tasks. Critically, these variables also positively correlated across tasks, as did Learning Efficiency Scores, a standardized average of those measures.

3.1 Why Does Learning Efficiency Generalize?

A natural follow-up question to ask is what underlying mechanisms account for the domain-generality of learning efficiency. Zerr and colleagues (2018, 2019) have proposed that attentional control, usage of learning strategies, and prior knowledge may explain variation in learning efficiency. Let us consider each of these in turn and whether their relation to efficient learning is supported by the present findings.

Even when partialing out related factors such as working memory capacity, multiple studies have found that long-term memory abilities are related to, albeit not completely subsumed by, attentional control (Shipstead et al., 2014; Unsworth, 2019; Unsworth & Engle, 2007; Unsworth & Spillers, 2010). It is believed that during encoding, heightened attentional control is required to attend to the to-be-learned information and inhibit external or internally generated distractors. Meanwhile, during retrieval attentional control modulates the specificity of search processes. Individuals with greater attentional control capabilities are thought to be better at filtering out irrelevant contextual cues (e.g., associations, timing and spatial context) and are therefore better able to hone in on cues that promote retrieval success. On the other hand, people with lesser attentional control resources fail to adequately focus on target items, diminishing the efficacy of encoding, and retrieve more irrelevant contextual cues. Less refined retrieval of cues in turn generates proactive interference that reduces recall success. Indeed, Kyllonen and Tirre (1988) found that slow learners were especially susceptible to interference. The variance shared between attentional control, encoding, and retrieval processes may partially explain the correlation between speed of learning and retention as well as the domain generalizability of learning efficiency.

The current study did not include attentional control tests, and so the question of whether attentional control underlies differences in learning efficiency cannot be definitively answered.

However, response time latencies offer partial support for this hypothesis. Participants with greater Learning Efficiency Scores recalled correct Lithuanian translations more quickly on average. That said, on the objects task the correlation between retrieval speed and learning efficiency was equivocal, although that may be an artifact of the motor response requirements of the task.

Another factor that may explain the generalizability of learning efficiency is strategy use. Usage of effective strategies at encoding and retrieval is strongly related to recall on a range of memory tasks (Dunlosky, Hertzog, & Powell-Moman, 2005; McDaniel & Kearney, 1984; Unsworth, 2019; Zerr, 2018). In paired associates recall tasks, mediators linking cues and targets are particularly effective (McDaniel & Kearney, 1984). As discussed by Dunlosky et al. (2005), people can differ in whether they spontaneously generate mediators, the quality of the mediators that they generate, whether they are able to recall the right mediators, and whether they appropriately decode mediators. Deficiencies in any of these steps could lead to poor learning and retention across various memory tasks. Thus, it may be that more efficient learners generate mediators more consistently at encoding, use higher quality mediators, and later recall and decode these mediators more successfully during retrieval.

The importance of strategy use for efficient learning is supported by the presence and sophistication of responses to the strategy questionnaire in this study. On average, participants who struggled to come up with strategies or who reported that their strategies did not work performed worse on the Lithuanian-English task. Additionally, participants that failed to generate strategies more frequently also tended to report less success when they did generate strategies. However, aside from the *Physical* strategy, no strategies correlated with overall task performance. One possibility is that, because strategy use was queried at the end of the task using a self-report questionnaire rather than on every trial, participants failed to accurately report which strategies they used and to what extent they relied on them. Alternatively, perhaps participants used strategies that were not reflected in the questionnaire.

On the objects task, the only strategies that were related to higher Learning Efficiency Scores were the *Follow with Cursor* and *Complex Associations* strategies. It is surprising that the *Clock* and *Coordinates* strategies were not correlated with higher performance as these were anticipated to be the most effective on this task. It may be that using these strategies effectively requires extensive practice or that they only provide a benefit when used appropriately. For example, remembering that the Apple was positioned at 7 o’clock is not a sufficiently precise description of the location to recall it accurately. Instead, participants would need to remember that the Apple was located at 7:30 o’clock and three-quarters of the way to the circle’s circumference.

A third potential mechanism of learning efficiency variability is differences in general knowledge. Kyllonen and Tirre (1988) found that general knowledge predicted unique variance on a battery of long-term memory tasks. In a follow-up experiment, Kyllonen et al. (1991) found that general knowledge predicted paired associates recall, and that the magnitude of this correlation increased with longer study times, presumably because high-knowledge individuals were afforded enough time to use their knowledge to generate effective associations. Reinforcing the importance of knowledge, Hundal and Horn (1977) found that crystalized intelligence correlated with paired associates learning. Both the Lithuanian-English and the object locations tasks were explicitly designed to minimize the influence of prior knowledge. Nonetheless, well informed participants may have used their knowledge repositories to generate better associations or to generate associations more quickly, facilitating encoding and retrieval alike.

Other variables that may underlie learning efficiency include processing speed (Kyllonen et al., 1991; Zerr et al., 2018), interest, and motivation (Unsworth, 2019). Unexpectedly, and in contrast to Nelson et al. (2016), demographic characteristics such as age did not relate to learning efficiency.

3.2 Limitations and Future Directions

A logical extension to the current study would be to test whether learning efficiency extends to other types of memory tasks such as free recall or recognition tests. Using the logic of the multitrait-multimethod matrix (Campbell & Fiske, 1959), if learning efficiency is a unique construct, it should not merely reflect shared method variance but rather an underlying trait that is dissociable from the tasks used to measure it. If performance still correlates across task type, this would rule out the possibility that learning efficiency solely reflects a general paired associates factor.

A limitation of the present study is that both the Lithuanian-English and the object locations tasks contained words in the cues. Future research should use non-verbalizable cues and targets to minimize the influence of prior vocabulary knowledge or language ability. To test whether learning efficiency is modality independent, future studies should also use stimuli from other sensory domains (e.g., sounds, haptic stimuli).

A novel contribution of this project is that spatial precision, a continuous index of spatial learning, was found to be associated with both visuospatial and verbal learning efficiency measures. Such continuous measures of memory fidelity have the advantage of tracking subthreshold learning that is not captured by binary recollection accuracy scores. The additional granularity provided by these tasks could afford greater sensitivity in detecting individual differences in memory ability, which could be valuable for studying populations with mild deficits or that are in the incipient stages of cognitive decline. In addition to the continuous visual measures reported here, continuous verbal measures have been developed that could be used in future studies (Lew et al., 2016). Future work should seek to determine and develop the clinical utility of these methods.

Precision in the present study was defined as the Euclidean distance between selected and target object locations. It should be noted that this definition differs from prior work where precision was statistically decomposed into three sources using mixture models: error arising from random guessing, misassociations, and imprecision (Lew, et al., 2016). Thus, future research could use measures of precision that account for guessing and misassociations.

To more systematically investigate the underlying mechanisms of learning efficiency, an individual-by-treatment interaction experimental approach could be employed (Kane & Miyake, 2008). For example, to assess the role that strategy use plays in learning efficiency, high and low efficiency learners could receive strategy training. If effective strategy usage boosts learning efficiency, there should be a main effect of strategy training that raises Learning Efficiency Scores. More interestingly, we might also expect an individual-by-treatment interaction in which low efficiency learners benefit more than high efficiency learners who may use strategies more skillfully by default.

Ultimately, a better understanding of individual differences in how quickly people learn and how long they remember may enable the creation of new assessments and interventions to aid learners. To this end, future work should determine whether and how learning efficiency measures relate to real-world learning outcomes such as classroom grades, and whether targeted interventions can improve performance both in the lab and in applied settings.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*, 445-459.

Becker, H. C. (2018). Individual differences in efficient learning: The relative contributions of attentional control and working memory capacity (Honor's thesis). Washington University in St. Louis, St. Louis, MO.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Bors, D. A., & MacLeod, C.M. (1996). Individual differences in memory (PDF). In E. L. Bjork and R. A. Bjork (Eds.), *Handbook of perception and cognition: Memory* (Vol. 10, pp. 411-441). San Diego, Ca: Academic Press.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*, 14325-14329.

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*, 981-990.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*, 155.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33*, 497-505.

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.

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*, 1-12.

Dennis, S. A., Goodson, B. M., & Pearson, C. (2018). Mturk Workers' Use of Low-Cost "Virtual Private Servers" to Circumvent Screening Methods: A Research Note.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*, 615.

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, *41*, 389.

Ebbinghaus, H. (2013). Memory: A contribution to experimental psychology. *Annals of neurosciences*, *20*, 155.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*, 532.

Field, A. P. (2014). Intraclass correlation. *Wiley StatsRef: Statistics Reference Online*.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, *47*, 381.

Gillette, A. L. (1936). Learning and retention: A comparison of three experimental procedures. *Archives de Psychologie, 28*, 198.

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.

Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin & Review*, *20*, 318-325.

Harlow, I. M., & Yonelinas, A. P. (2016). Distinguishing between the success and precision of recollection. *Memory*, *24*, 114-127.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166-1186.

Hundal, P. S., & Horn, J. L. (1977). On the relationship between short-term learning and fluid and crystallized intelligence. *Applied Psychological Measurement*, *1*, 11-21.

Kane, M. J., & Miyake, T. M. (2008). Individual differences in episodic memory. In H.L. Roediger, III (Ed.), *Cognitive psychology of memory* (Vol. 2, pp 773-785). Oxford, UK: Elsevier.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*, 155-163

Kyllonen, P. C., & Tirre, W. C. (1988). Individual differences in associative learning and forgetting. *Intelligence*, *12*, 393-421.

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*, 57-59. http://dx.doi.org/10.1037/0096-3445.120.1.57

Lew, T. F., Pashler, H. E., & Vul, E. (2016). Fragile associations coexist with robust memories for precise details in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 379.

MacLeod, C. M., & Nelson, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica*, *57*, 215-235.

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

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.

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.12.

Richter, F. R., Cooper, R. A., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *Elife*, *5*, e18260.

Selmeczy, D., & Dobbins, I. G. (2014). Relating the content and confidence of recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 66.

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116-141.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*, 420.

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in r. *The Journal of Open Source Software*, *1*, 37.
Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology*, *25*, 207-222.
Underwood, B. J. (1954). Speed of learning and amount retained: A consideration of methodology. *Psychological Bulletin*, *51*, 276-282. http://dx.doi.org/10.1037/h0056741
Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*, 79-139. http://dx.doi.org/10.1037/bul0000176
Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological review*, *114*, 104.
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.
Uttl, B. (2005). Measurement of individual differences: lessons from memory assessment in research and clinical practice. *Psychological Science*, *16*, 460-467.
Woodworth, R. S. (1914). A contribution to the question of "quick learning, quick forgetting." *Psychological Bulletin*, *11*, 58-59.
Zerr, C. L. (2017). The domain-generality and durability of efficient learning (Master's thesis). Washington University in St. Louis, St. Louis, MO.
Zerr, C. L., Berg, J. J., Nelson, S. M., Fishell, A. K., Savalia, N. K., & McDermott, K. B. (2018). Learning Efficiency: Identifying Individual Differences in Learning Rate and Retention in Healthy Adults. *Psychological science*, *29*, 1436-1450.
Zerr, C. L., & McDermott, K. B. (2019). *Are efficient learners of verbal stimuli also efficient learners of visuospatial stimuli*? Manuscript submitted for publication.
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233.

Appendix

Table A1. Lithuanian-English word pairs and concreteness ratings.

Lithuanian	English	Concreteness
Obuolys	Apple	620
Tvartas	Barn	614
Vonia	Bath	600
Tiltas	Bridge	623
Pastatas	Building	589
Pyragas	Cake	624
Puodelis	Cup	539
Durys	Door	606
Bugnas	Drum	602
Akis	Eye	634
Zuvis	Fish	597
Plaukas	Hair	583
Raktas	Key	612
Riteris	Knight	579
Koja	Leg	626
Turgus	Market	551
Pienas	Milk	670
Burna	Mouth	568
Nafta	Oil	581
Augalas	Plant	594
Lietus	Rain	600
Ziedas	Ring	593
Kambarys	Room	566
Muilas	Soap	598
Laiptelis	Stair	558
Gatve	Street	579
Stalas	Table	604
Vanduo	Water	616

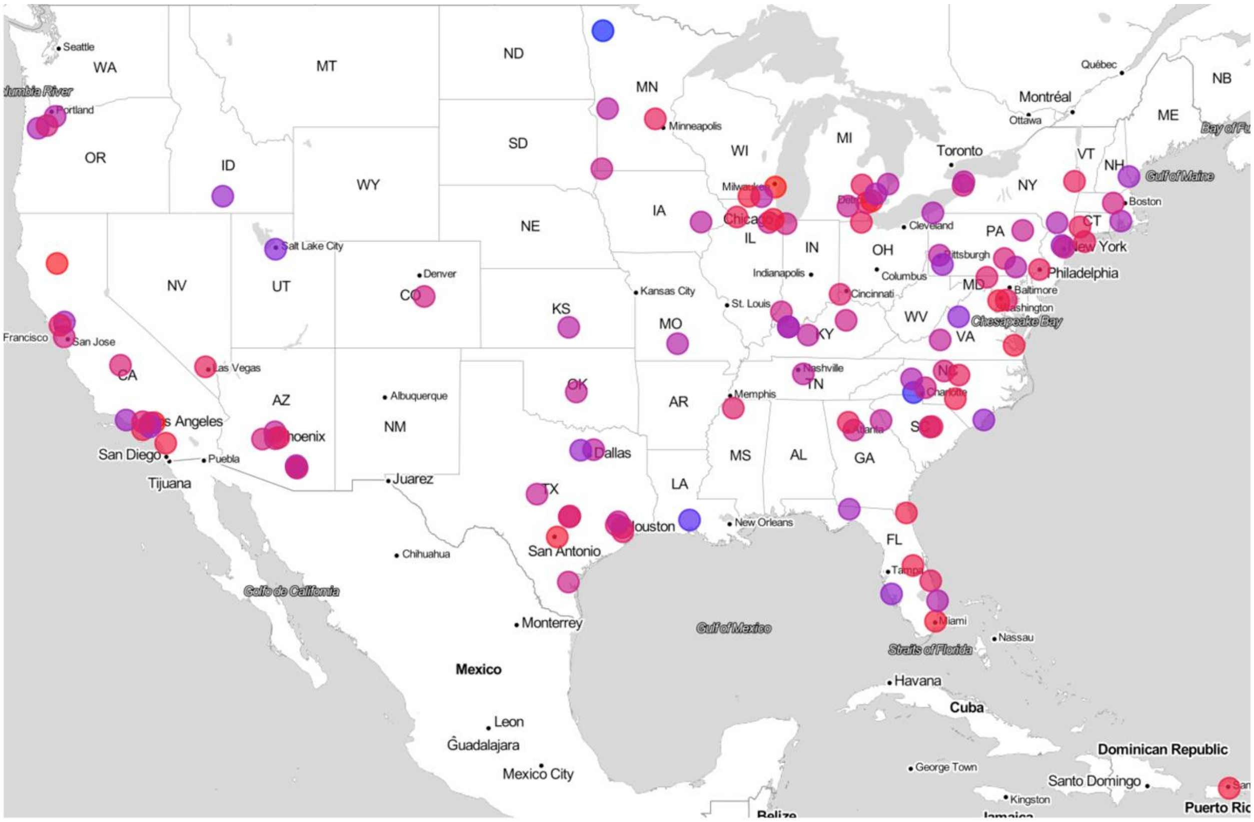

Figure A1. Map of participant locations. All participants resided within the continental U.S. or a U.S. territory. Locations are plotted using latitude and longitude coordinates extracted from the Qualtrics survey data and are approximations derived by comparing IP addresses to a location database.

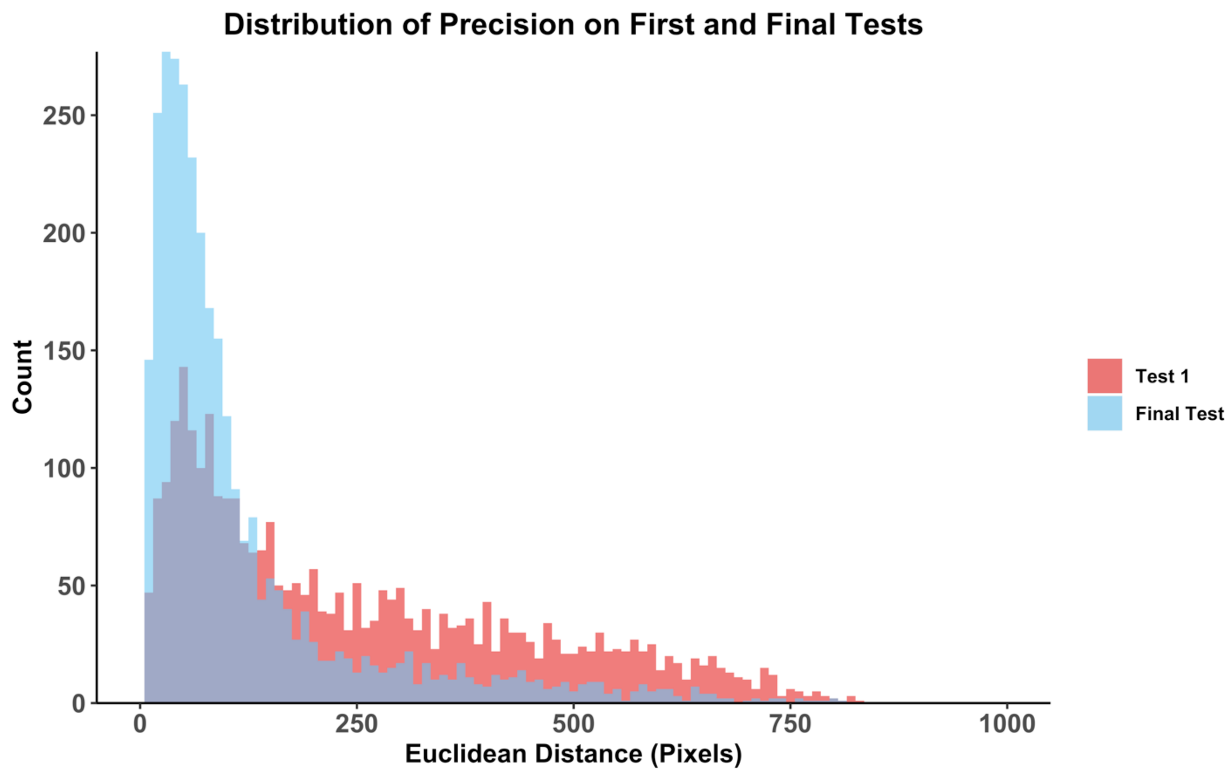

Figure A2. Histograms comparing the distribution of spatial precision on the object locations task on Test 1 and the Final Test across all trials. Precision improves by the Final Test.

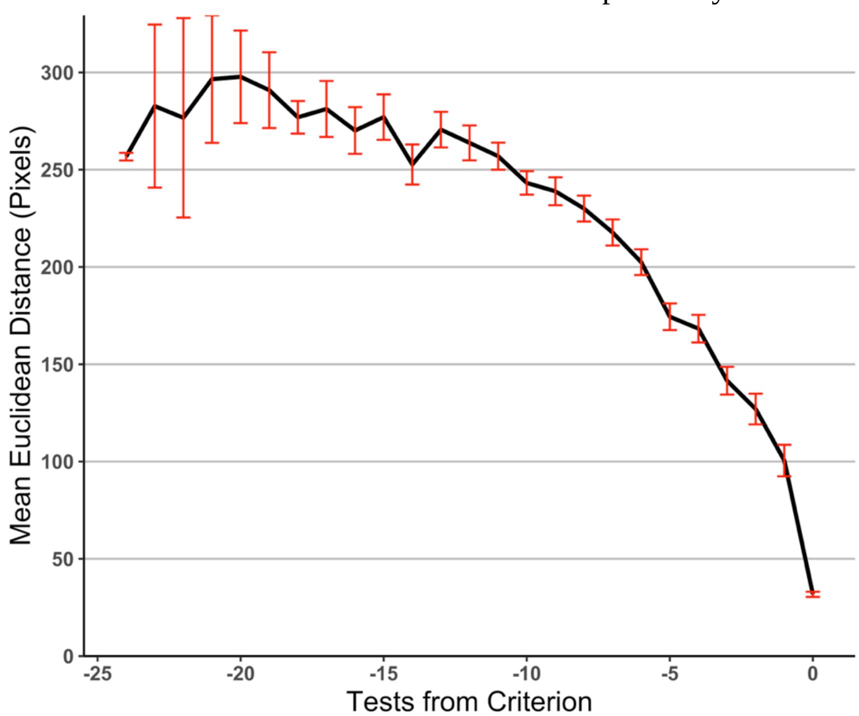

Figure A3. The learning curve of spatial precision indicates that participants progressively learned the object locations. The curve displays the across-participant mean precision of each test block relative to the final block in the Tests to Criterion phase. Error bars are standard errors of the mean.

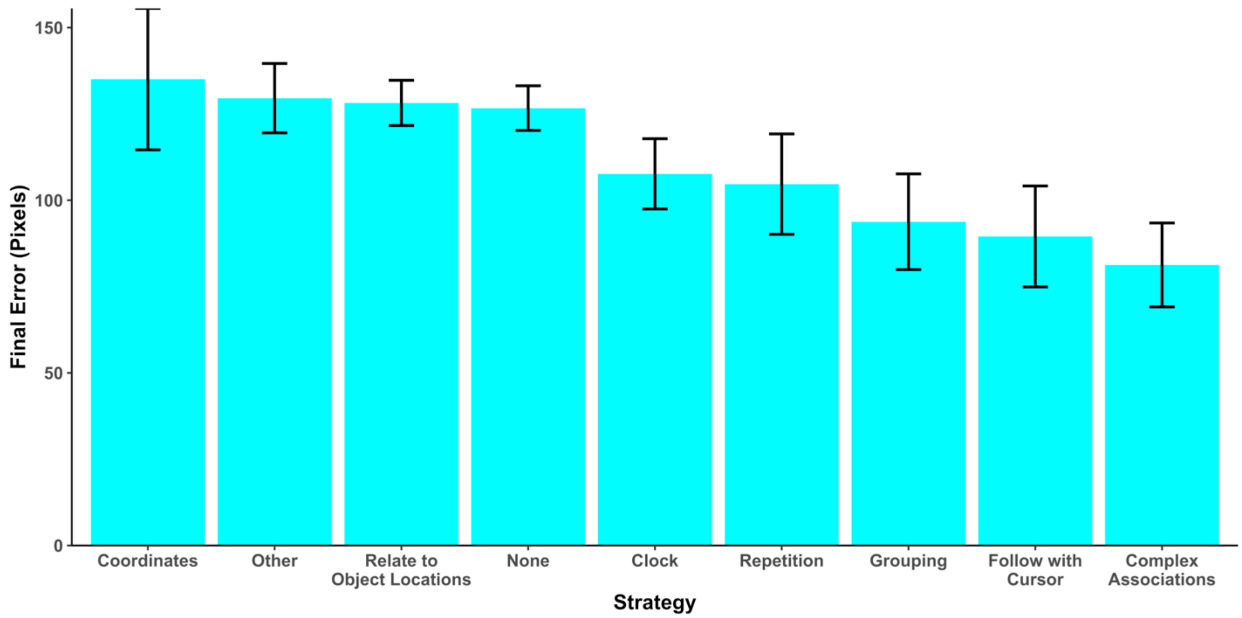

Figure A4. Recall precision of object locations on the final test binned by reported strategy. Participants relying on the *Complex Associations*, *Follow with Cursor*, and *Grouping* strategies recalled objects more precisely than participants without a strategy. Error bars represent standard errors.