AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2025, Vol. 154, No. 2, 420–434
https://doi.org/10.1037/xge0001659

Individual Differences Diminish the Pretest Effect Under Productive Memory Conditions

Lucy M. Cronin-Golomb, Julia T. Wilson, Alissa G. Miller, and Patricia J. Bauer
Department of Psychology, Emory University

Pretesting, or asking a test question prior to the onset of learning, is a well-established means of enhancing learning. Research on pretesting has focused primarily on direct factual learning outcomes. Yet building a coherent knowledge base also depends on productive memory processes that permit going beyond the information directly given. In the specific productive process of self-derivation through memory integration, individual differences are prominent; verbal comprehension is a consistent predictor. In the current work, we integrated these research trends by testing the extent to which pretesting enhances learning through productive memory processes and the role played by individual differences in verbal comprehension. Across four within-subjects experiments, we assessed the pretest effect after accounting for variability associated with verbal comprehension. In Experiments 1–3, we assessed the productive memory process of self-derivation through memory integration. Adults were more successful on pretest trials compared to control (i.e., no pretest) trials, but this effect was no longer significant after controlling for verbal comprehension. This pattern emerged when we used stem-fact pretests (Experiment 1) and integration-fact pretests (Experiment 2) to probe self-derivation across single-sentence stimuli and replicated when we used stimuli more akin to everyday learning materials (i.e., text passages and photographs; Experiment 3). In Experiment 4, we shifted the test target from productive processes to fact recall and found the pretest effect held even after controlling for verbal comprehension. This research bridges the pretest and productive process literature to provide novel insight into ways of maximizing learning.

Public Significance Statement

How can we best support the acquisition of new knowledge? One way is through pretesting or asking learners to answer test questions before learning. We know pretesting helps later recall of factual content. What we do not know is the extent to which pretesting helps learners build knowledge using productive memory processes, which require manipulation and a combination of two or more pieces of information. In our first three experiments, we tested the pretest effect under productive memory conditions. As predicted, pretests enhanced productive memory processes. However, in all three experiments, the effect diminished once we accounted for individual differences in verbal comprehension. We tested factual outcomes in our last experiment; the pretest effect was maintained even after controlling for verbal comprehension. Results speak to the importance of considering individual differences when designing ways of maximizing learning and the potential limitations of pretesting beyond learning from direct recall.

Keywords: pretest effect, self-derivation, productive memory, verbal comprehension, individual differences

This article was published Online First October 3, 2024.
Rachel Wu served as action editor.
Lucy M. Cronin-Golomb https://orcid.org/0000-0002-1442-1565
Data and analyses from this article have not been previously published nor are under consideration for publication elsewhere. Research materials are available upon request from the Bauer Lab Integration and Self-derivation Stimulus Bank at https://scholarblogs.emory.edu/bauerstimuli/. The code is available at https://osf.io/nz3ae/?view_only=2bf086fc517c4ced812c64eeeb873608.

The authors have no conflicts of interest to declare. This research was supported by funding from the Emory College of Arts and Sciences. Lucy M. Cronin-Golomb was supported by Institutional National Research Service Award Training, National Institutes of Health (Grant HD071845).

The authors thank the other members of the Memory at Emory laboratory for their invaluable feedback at each stage of this project, especially Hilary Miller-Goldwater for her help in curating the data cleaning code. The authors also extend a special note of appreciation to all the participants who took part in this research.

Lucy M. Cronin-Golomb played a lead role in data curation, formal analysis, investigation, project administration, visualization, and writing–original draft and an equal role in conceptualization, funding acquisition, methodology, validation, and writing–review and editing. Julia T. Wilson played a lead role in validation and writing–review and editing, a supporting role in formal analysis and investigation, and an equal role in conceptualization, data curation, methodology, project administration, and supervision. Alissa G. Miller played a supporting role in formal analysis, project administration, validation, and writing–review and editing and an equal role in conceptualization, data curation, investigation, and methodology. Patricia J. Bauer played a lead role in funding acquisition, resources, supervision, validation, and writing–review and editing and an equal role in conceptualization, formal analysis, and methodology.

Correspondence concerning this article should be addressed to Lucy M. Cronin-Golomb, Department of Psychology, Emory University, 36 Eagle Row, Emory University, Atlanta, GA 30322, United States. Email: lucycroningolomb@gmail.com

Learning inevitably occurs across time, context, and media. To continuously build a coherent knowledge base, learners must extend upon knowledge extracted from direct experience. They must engage *productive memory processes* to manipulate and combine related information across learning episodes (Bauer, 2021). For example, to promote the formation of new knowledge about the structure of an atom, a teacher might offer students the following comparison: An atom is structured similarly to a solar system, with the nucleus as the sun and the rotating protons and electrons the orbiting planets. In this example, students must rely on prior knowledge about the structure of the solar system to form a new conceptual understanding of the structure of an atom.

Identifying ways to support learners' acquisition of new knowledge is a major topic of research on learning. *Pretesting*,[1] or asking learners to answer test questions prior to the onset of new learning, is one particularly robust way to maximize learning outcomes (Grimaldi & Karpicke, 2012; Little & Bjork, 2016; Richland et al., 2009). Pretesting results in enhanced memory for subsequently learned material, a phenomenon known as the *pretesting effect* (Pan & Rivers, 2023). The pretesting effect has been shown to generalize across context and stimuli (Carpenter & Toftness, 2017; Pan & Carpenter, 2023; Richland et al., 2009; Soderstrom & Bjork, 2023). Yet, research to date has focused primarily on the effect of pretests on factual learning. The extent to which pretesting enhances learning from productive memory processes remains relatively understudied, with some research highlighting potential concerns regarding the efficacy of pretests beyond fact recall (Hausman & Rhodes, 2018; St. Hilaire et al., 2019). Further, though pretesting has been studied under an array of contexts, there is still much to learn about factors influencing variability in outcomes. Specifically, though individual differences are highlighted prominently in the productive process literature, their role in learning outcomes associated with pretesting is not yet clear.

In the current research, we aimed to bridge the gap across the pretesting and productive process literature. In doing so, we provide novel insight into open questions regarding the mechanisms of pretesting. We conducted four within-subjects experiments to provide crucial insight into (a) whether the well-established strategy of pretesting promotes learning outcomes beyond direct factual recall and (b) the effect of individual differences on learning from pretests.

Pretesting

There are numerous conceptual theories as to how pretests enhance learning outcomes. First, some researchers posit that pretests highlight knowledge gaps, which can then be operated on during learning to ultimately enhance outcomes. For example, learners might be more likely to seek out information that can be used to provide a previously unknown response to a question (Cronin-Golomb & Bauer, 2023; Pan & Carpenter, 2023; Pan et al., 2020; St. Hilaire & Carpenter, 2020). Others suggest that the pretest might provide an important episodic memory cue to which the learner can then anchor information encoded during their learning experience (Karpicke et al., 2014; Metcalfe & Huelser, 2020). Another conceptual viewpoint is that engaging with the pretest might help the learner activate information already available in relevant prior knowledge structures that can then be used to aid in new learning as well as retention (Carpenter, 2009). Yet another account suggests that pretesting may strengthen retrieval pathways to related content, which may be particularly relevant under productive memory conditions (Richland et al., 2009). Differences aside, there is consensus that pretesting falls under the umbrella of *test-potentiated new learning*: New learning is broadly enhanced when the learner attempts to generate a response to a test question (Chan et al., 2018). It does not matter if the attempt is successful (Kornell, 2014; Metcalfe, 2017). Indeed, the answer to a pretest question is often "I don't know," given the learner has not yet been exposed to the relevant educational materials.

Pretesting has been shown to support encoding and long-term retention of new knowledge under myriad conditions. For example, Richland et al. (2009) showed that pretests enhanced learning from text passages. Specifically, within-subjects performance increased by 50% when test questions were pretested compared to when they were not. Further, overall learning was significantly higher for participants who had access to pretest questions compared to those who were instead asked to spend additional time studying to-be-tested material. Carpenter and Toftness (2017) showed similar patterns when testing learning from videos. Within-subjects performance increased by 30% when posttest questions were pretested. Overall performance increased by 50% when participants had access to pretest questions, compared to those who only viewed videos (i.e., the control group). Finally, using a within-subjects design, Kliegl et al. (2024) found that pretesting paired associates not only enhanced learning on immediate posttests but also the benefit actually increased over a 1-week delay.

It is evident that pretesting enhances learning outcomes. Yet literature to date has focused primarily on the influence of pretests on factual learning outcomes. For example, Richland et al. (2009) identified 10 sentences within the to-be-studied material that could be tested for retention and designed factual open-ended and fill-in-the-blank pretests that specifically probed factual content available within those explicit sentences. There are few studies examining the influence of pretests on learning beyond factual recall. This represents an important gap in the literature, given that to continuously build a cohesive knowledge base, learners must move beyond facts gained from direct experience to productively extend knowledge (Bauer, 2021; Gentner & Holyoak, 1997; Han & Kim, 2019; Wilson & Bauer, 2021). Findings from the limited research examining pretests and learning beyond factual recall paint somewhat of an inconclusive picture. Hausman and Rhodes (2018) found no learning benefit of conceptual pretests designed to probe inference-making across two pieces of related information presented within text-based passages. Another study focusing on "integrative" pretests (i.e., pretests that probed combination across two pieces of related information directly embedded within the learning material) found a similar pattern as Hausman and Rhodes, but showed a pretest benefit when participants were instructed to look for answers during learning (St. Hilaire et al., 2019). Another study used factual pretests to promote conceptual

[1] The terms pretesting and prequestioning have often been used interchangeably in the literature. Some researchers make a distinction between these terms such that pretesting requires immediate feedback and prequestioning does not. However, in the works cited throughout this article, this distinction is not consistent, and these terms are not clearly demarcated in the literature. Thus, we use the term pretesting to refer broadly to asking test questions prior to learning, and we do not distinguish between the presence or lack thereof of feedback.

learning outcomes. Results showed that factual pretests enhanced conceptual learning outcomes, especially as a way to highlight key features of the deeper structures of statistical concepts (Sana et al., 2021).

This review illustrates that the potential benefits of pretesting on learning beyond fact recall remain unclear. It may be that pretesting effects arise from similar mechanisms for factual outcomes and inference-based outcomes, for example, by highlighting factual knowledge gaps and directing attention to relevant material and/or strengthening pathways between pretested and related material. On the other hand, pretesting alone might not be as effective under conditions requiring inference-making, necessitating additional support for learners (e.g., explicit instructions to look for answers) to fully capitalize on the benefits of pretesting.

Another relatively understudied domain in the pretest literature is whether individual differences contribute to learning outcomes—inclusion of individual difference measures is uncommon in the pretesting literature. Consideration of individual differences is an important direction for research given the substantial variability in learning outcomes and even in response to manipulations such as pretesting (Carpenter & Toftness, 2017; Kornell, 2014; Pan & Carpenter, 2023; Rickards & McCormick, 1988). One study that measured individual variability as a predictor of the effects of factual pretests included a suite of measures of individual variability: academic achievement as measured by GPA (grade point average, a numerical representation of a student's academic aptitude), confidence in test answers, accuracy on pretest questions, familiarity with topics probed by pretest questions, and how much of assigned reading students engaged with before class (Geller et al., 2017). The only significant predictor of performance was variability associated with accuracy on pretest questions (if a participant answered the pretest question correctly, they were more likely to answer the posttest question correctly), though the authors noted that Bayesian analyses showed weak evidence for the lack of effects of the other predictors. Another study examined the influence of achievement motivation and found that students with high achievement motivation benefitted more from pretests than those with low-achievement motivation (Yang et al., 2021). St. Hilaire et al. (2019) examined the influence of structure-building ability, or creating coherent mental representations of conversations, texts, pictorial stories, and other events (McDaniel et al., 2022), on the effect of integrative pretests. They found a positive relation between structure building and general posttest performance, though no statistical evidence of an interaction between the pretest effect and structure building. Further, they found that the benefit of integrative pretests, which only emerged once participants were instructed to look for answers within to-be-tested material, was maintained even after individual differences in structure building were accounted for.

These studies provide important insight into whether individual differences are associated with learning enhancement afforded by the pretest, but there is still much to be learned. In the current work, to address this gap, we turned to the productive process literature wherein individual differences are prominent. Moreover, we turn to the productive process literature to consider the pretesting effect on learning beyond direct recall. Learning through productive processes is essential to the formation of a coherent knowledge base, and thus, the effect of pretesting on productive outcomes is of critical interest. Specifically, in the current work, we used the *self-derivation through memory integration* paradigm to test the effect of pretesting on productive memory processes.

Self-Derivation Through Memory Integration

Self-derivation through memory integration is a productive memory process that requires the combination of two separate yet related pieces of new information to generate new knowledge (Varga & Bauer, 2017). For example, consider a learner who at one point learns the fact that "Talc is the softest known mineral." Later, they may learn "Ancient Egyptians were the first to use talc." When asked "Who were the first to use the softest mineral?" the learner can combine across the two related episodes of new learning to generate or self-derive the target answer "Ancient Egyptians." Importantly, the integration fact was never directly taught. The learner extracted and manipulated two separate pieces of information, to build new knowledge.

The self-derivation paradigm is an apt vehicle to address the question of whether pretesting influences learning from productive memory processes. For one, research shows that self-derivation is an ecologically valid productive memory process. Integration facts are quickly incorporated into the knowledge base, treated as "familiar" (Bauer & Jackson, 2015), and retained over time (Varga & Bauer, 2013). They are also used productively, to accumulate yet more new knowledge (Wilson & Bauer, 2021). Further, in adults, self-derivation relates to academic achievement as measured by SAT and GPA scores (Varga et al., 2019). The SAT is a standardized assessment used in college admissions in the United States, which evaluates high school students' skills in areas such as reading, writing, and math.

Another reason to use the self-derivation paradigm is because it allows for two different approaches to pretest design within the same paradigm: factual and integrative. For example, the pretest can probe one of the two integrable facts within a pair: "What is the softest known mineral?" or "Who were the first to use talc?" The pretest can also probe the productively derived fact itself: "Who were the first to use the softest mineral?" It is logical to hypothesize that a pretest question that features either one or the combination of two integrable facts may enhance learning, particularly since one mechanism by which pretesting is thought to work is through strengthening connections between pretested and related material.

Across time, context, and media, there is striking variability in self-derivation performance. Indeed, empirical findings show a success rate of 0%–100% correct under self-derivation task conditions (Cronin-Golomb & Bauer, 2022; Dugan & Bauer, 2022; Varga & Bauer, 2017; Varga et al., 2019). Research suggests that individual differences in learners is one reason for the vast variability seen in self-derivation performance. Particularly robust is the finding that in adults and children, *verbal comprehension* relates to self-derivation task performance (Bauer et al., 2019; Esposito & Bauer, 2022; Varga et al., 2019). Verbal comprehension, as featured in the self-derivation literature, is thought to be both a broad measure of a learner's semantic knowledge as well as a specific measure of lexical knowledge, vocabulary knowledge, and a learner's ability to reason using lexical knowledge. The extent to which verbal comprehension supports performance as a function of the pretest effect is not yet known. However, it is logical to hypothesize that the mechanisms by which pretesting is thought to enhance learning are related to verbal comprehension skills. For example, verbal comprehension supports positive change in reading comprehension or the ability to derive meaning from text (Reynolds & Turek, 2012). Deriving meaning from text is an essential means by which the pretest effect is thought to function, for example, seeking out information highlighted by a knowledge gap. Second, verbal comprehension is a proxy for semantic knowledge (Schrank & Wendling, 2018; Varga et al., 2019). One major theory by which pretesting is thought to "work" is that exposure to the pretest questions activates relevant prior knowledge structures, thus facilitating the encoding of new information during the subsequent learning phase (Carpenter, 2009). It is logical then to speculate that learners with higher semantic knowledge, as indexed by verbal comprehension, may benefit the most from a learning tool that relies on the activation of preexisting prior knowledge. Accordingly, we hypothesized that verbal comprehension would be a significant predictor of overall self-derivation performance and that there would be a significant interaction between verbal comprehension and the pretest effect across all experiments.

The Current Research

There are many theories as to the mechanisms by which pretesting enhances learning outcomes: highlighting knowledge gaps (Pan & Carpenter, 2023), activating relevant prior knowledge structures (Carpenter, 2009), and strengthening pathways between related pieces of information (Richland et al., 2009), to name a few. Research provides empirical support for many of these theories, yet factual recall is the primary outcome of focus to date, and individual differences are largely neglected. In the current work, to provide novel insight into the pretest effect, we tested productive memory outcomes and assessed individual differences, particularly in relation to verbal comprehension. Specifically, we conducted four within-subjects experiments. In Experiments 1, 2, and 3, we used the self-derivation through memory integration paradigm to test the influence of pretesting on productive memory processes. In Experiment 4, we used a subset of the stimuli in Experiment 1 to replicate prior work examining the effect of pretests on direct factual outcomes. Across all experiments, participants engaged in tests of verbal comprehension to assess individual differences.

In Experiments 1 and 2, learners were exposed to separate yet related episodes of new learning in a single-sentence format. In Experiment 1, pretests probed one member of an integrable fact pair (i.e., one of the directly taught facts). In Experiment 2, pretests probed the integration of both members of a fact pair (i.e., the self-derived outcome). In Experiment 3, to test the benefit of pretesting on materials more similar to everyday learning materials, and with the expectation that pretesting would benefit single-sentence self-derivation performance, we used text-based passages and photographs developed from preexisting art history museum exhibits. We had two central predictions for the current work: (1) pretest questions designed to probe either one or the combination of two integrable facts would enhance self-derivation through memory integration performance across stimulus types and (2) verbal comprehension would predict overall self-derivation performance, and there would be a significant interaction between verbal comprehension and the pretest effect across all experiments.

To foreshadow the results, in Experiments 1, 2, and 3, we found the expected effect of pretesting, such that performance was higher on trials that had been pretested relative to those that had not. However, in all three experiments, we also found that once we controlled for the variability associated with individual differences in verbal comprehension, we no longer observed a statistically significant effect of pretesting. To determine whether this pattern was specific to materials associated with productive processes, in Experiment 4, using a subset of the exact stimuli used in Experiments 1 and 2, we used a factual outcome paradigm: Both pretests and posttests probed factual outcomes. We found that the pretest effect was maintained even after accounting for individual variability associated with verbal comprehension. As a whole, this research provides novel insight into ways of maximizing learning from productive memory processes.

Transparency and Openness

For this and subsequent experiments, we adhere to the journal article reporting guidelines outlined by Kazak (2018). We provide comprehensive reports of how we determined sample sizes and exclusion criteria. We also provide detailed descriptions of all measures and manipulations. Research materials are available upon request from the Bauer Lab Integration and Self-derivation Stimulus (BLISS; https://scholarblogs.emory.edu/bauerstimuli/) bank (Bauer, 2020). We do not make materials available publicly as the fidelity of ongoing work requires novelty of stimulus items. The code is available at https://osf.io/nz3ae/?view_only=2bf086fc517c4ced812c64eeeb873608, and data are available upon request. Please note that researchers must sign data use agreements to access the data, as required by university's institutional review board guidelines. For this and subsequent experiments, all primary analyses are frequentist by nature, conducted using JASP Team, 2023) and R statistical software (V4.3.1; R Core Team, 2023). We tested and confirmed that all the data for this and subsequent experiments met the necessary statistical assumptions. None of the experiments in this article were preregistered.

Experiment 1

Method

Participants

Participants were 45 adults ($M_{\text{age}} = 20.42$; 30 females, 14 males, one nonbinary) recruited through Prolific, an online participant recruitment software. There is a relative paucity of research examining the pretesting effect under productive memory conditions. Thus, for this and subsequent experiments probing self-derivation through memory integration, we estimated a medium effect size. We used G*Power (Faul et al., 2009) with .8 power and a Cohen's *f* effect size of .25 to determine our sample size. Based on self-report, the sample was American Indian or Alaska Native (2.2%), Asian (8.8%), Black or African American (8.8%), Mixed or Other (8.8%), Native Hawaiian or Pacific Islander (2.2%), and White or Caucasian (64.4%). In total, 4.5% of participants did not report their race, and 17.8% of the sample identified as Hispanic or Latinx. Prolific policy allows researchers to set prescreening requirements to permit samples to be customized. All participants were prescreened such that their highest level of education completed was "high school diploma/A levels" (thus ensuring comparability with typical pretesting effect samples of college undergraduate students), and their first language was English, based on self-report. All participants engaged in an online consent process before beginning study protocols and were compensated $8.00 per hour of their time. For this and subsequent experiments, all study protocols and procedures were approved by the university's institutional review board prior to the onset of data collection.

Stimuli

Facts. In line with the self-derivation literature, we use the terminology "stem-fact pairs" here and henceforth to refer to two integrable facts that could be combined to generate novel, true information. Stimuli were 20 integrable stem-fact pairs that covered a range of topics, including science, history, and geography. They were developed based on pilot testing with adults (N = 200), to ensure exposure to both facts was necessary to provide an accurate response to self-derivation questions. Stem facts within integrable pairs were 8.07 words on average (range = 4–16 words) and were matched to have approximately the same number of words ($M_{\text{difference}}$ = .55 words). The fact pairs are available from the BLISS bank: stimulus numbers S002, S069, S084, S093, S118, S121, S126, S128–129, S133, S138, S142–150.

Stimuli also included 14 nonintegrable "filler" facts, averaging 11.36 words each (range = 8–16). Filler facts represented unique semantic content compared to fact pairs: They could not be integrated with the stem facts or with each other. Filler facts were included to provide for engagement checks (e.g., "What was the last fact about?"), thus avoiding potential differences in the depth of processing of to-be-tested stem facts.

Test of Verbal Comprehension. We used Test 1: Verbal Comprehension from Woodcock–Johnson Test of Cognitive Abilities, Third Edition (Woodcock et al., 2001) adapted for online administration. It consisted of four subtests: Picture Vocabulary, Synonyms, Antonyms, and Verbal Analogies. This test has a median reliability of .95 in the adult samples. We adhered to Woodcock–Johnson protocols for scoring as well as basal and ceiling criteria.

Procedure

Participants were sent a link to an online survey built using QualtricsXM software. They were instructed, "We are interested in how people learn. For this study, you will be asked to read through a variety of facts and answer some questions. You will also complete some other language and memory tasks." Participants completed an online consent procedure and demographic form. The general procedure is illustrated in Figure 1.

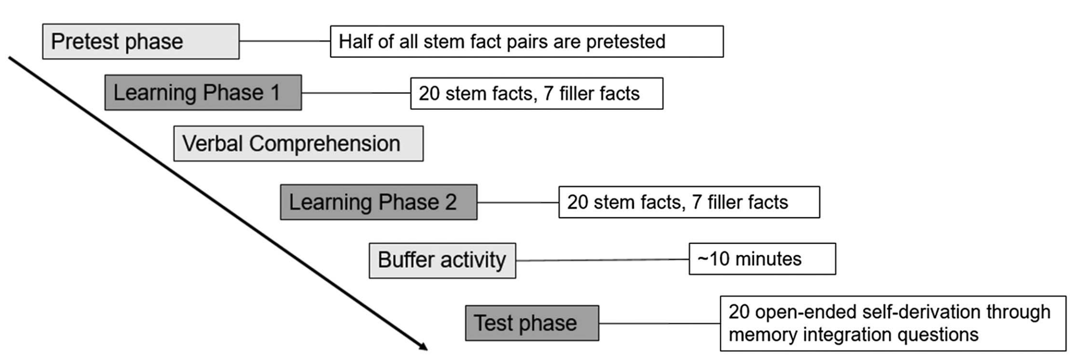
Pretest Phase. Participants were instructed, first, you will be asked to answer a set of questions. You will be required to respond before moving to the next page. If you do not know the answer, please take your best guess. If you have no guess, type "idk."

Importantly, there was no indication as to the relevance of these questions. That is, participants were not made explicitly aware that they would be learning material related to these questions. The pretest manipulation was within subjects such that half of all to-be-learned fact pairs were pretested and half were not.

Participants were exposed to 10 total open-ended pretest questions, one at a time. The survey was programmed such that participants could not proceed to the next question until they typed a response to the current one. Each pretest question probed one of two *stem facts* of 10 of the 20 to-be-learned stem-fact pairs. For example, the pretest "What is the softest known mineral?" probed the specific stem fact "Talc is the softest known mineral." The second, related stem fact "Ancient Egyptians were the first to use talc" was not probed. Pretest questions were counterbalanced such that stem facts were probed equally often across the sample. Additionally, pretest questions were programmed such that five probed stem facts were available in Learning Phase 1, and five probed stem facts were available in Learning Phase 2 (see below).

Learning Phase 1. After completing the pretest portion, participants were told, "You will now be learning some facts. Please read each fact carefully before moving on to the next." Participants then read 27 facts, one at a time. Participants had a maximum of 30 s to read each fact (timing based on pilot testing). Twenty of the facts were the first member of pairs of stem facts, the corresponding member of which would be presented in Learning Phase 2. The other seven facts were filler facts and did not have a counterpart in Learning Phase 2. After reading each filler fact, participants were asked to type a response to the question "What was the last fact about?" These questions permitted assessment of whether participants were engaged with the learning phase. We did not ask engagement questions for facts belonging to a stem-fact pair, to avoid variability in the depth of processing for to-be-tested content. Facts were counterbalanced such that they appeared

Figure 1
General Session Procedure



equally often in Learning Phase 1 and Learning Phase 2. Filler facts were dispersed roughly evenly throughout the learning phase (e.g., appearing once every two to three stem facts). The fact order was randomized for each participant.

Verbal Comprehension Test Phase. Upon completion of the first learning phase, participants engaged in all four subtests of Woodcock–Johnson Tests of Cognitive Abilities, Third Edition, Test 1: Verbal Comprehension (Woodcock et al., 2001). Due to the online nature of the current protocols, the presentation of the items occurred via QualtricsXM software. This task took approximately 10 min to complete.

Learning Phase 2. Participants read through 27 more facts in the second learning phase, seven of which were new filler facts and 20 of which were counterparts to facts presented in Learning Phase 1. As in the first learning phase, filler facts were followed by engagement questions. As in the first learning phase, filler facts were dispersed roughly evenly throughout. The fact order was randomized for each participant.

Self-Derivation Test Phase. Following a 10-min buffer activity, participants read, "You will now be asked some questions about the facts you learned earlier. You will be required to respond. If you do not know the answer, please take your best guess. If you have no guess, type 'idk'." Participants then had as much time as needed to answer 20 open-ended integration-fact questions. As in the pretest phase, the survey was programmed such that participants were required to type a response before advancing to the next question. All 20 questions were entirely novel. They were programmed to appear in a randomized order for each participant.

Scoring

Pretest Phase. Participants were assigned a score of 1 for each correct response and 0 for each incorrect response. Proportion correct was used in analyses.

Verbal Comprehension Test Phase. In line with Woodcock–Johnson III guidelines, a sum score was calculated for each participant by adding the number of correct answers from each of the four subtests. We then standardized the summed scores using the proportion of the maximum scaling method (T. D. Little, 2013). The standardized scores were used in all analyses.

Test Phase. Participants were assigned a score of 1 (correct) or 0 (incorrect) for each response, and the proportion correct was used in analyses.

Results and Discussion

There were two main goals of the present experiment: to assess the influence of stem-fact pretest questions on self-derivation through memory integration outcomes and to analyze the influence of verbal comprehension on postlearning performance. For this and subsequent online experiments, we determined a priori that participants would be excluded if they scored 85% or lower on engagement check questions. In Experiment 1, all participants scored 85% or higher on the engagement check questions (M = 99.5% correct, range = 92%–100% correct).

Pretest Phase Performance

On average, participants provided successful responses to 1.5 of the 10 pretest probes (.15 proportion correct). We examined the overall performance and identified one outlier (one participant who answered five out of 10 pretest probes correctly). We excluded this participant from future analysis.

Self-Derivation Test Performance

Descriptive statistics of task performance are provided in Table 1. Test questions that probed the integration of stem-fact pairs that were not pretested are referred to as "control" trials. Those that were pretested are "pretest" trials. To examine the effect of the pretest manipulation on self-derivation performance, we conducted a Type III sum of squares repeated-measures analysis of variance (ANOVA). We found a main effect of pretesting on memory integration outcomes, $F(1, 43) = 7.755$, $p = .008$, $\eta_p^2 = .153$.

Verbal Comprehension Performance

We first calculated verbal comprehension performance, available in Table 1. We then added verbal comprehension as a covariate to the above model. The main effect of verbal comprehension was not statistically significant, $F(1, 42) = 2.387$, $p = .130$. When verbal comprehension was included in the model, the main effect of pretest was diminished, and no longer statistically significant, $F(1, 42) = 2.843$, $p = .099$, $\eta_p^2 = .063$. Further, the interaction between pretest and verbal comprehension was not statistically significant, $F(1, 42) = .329$, $p = .569$, $\eta_p^2 = .008$. Results are illustrated in Figure 2.

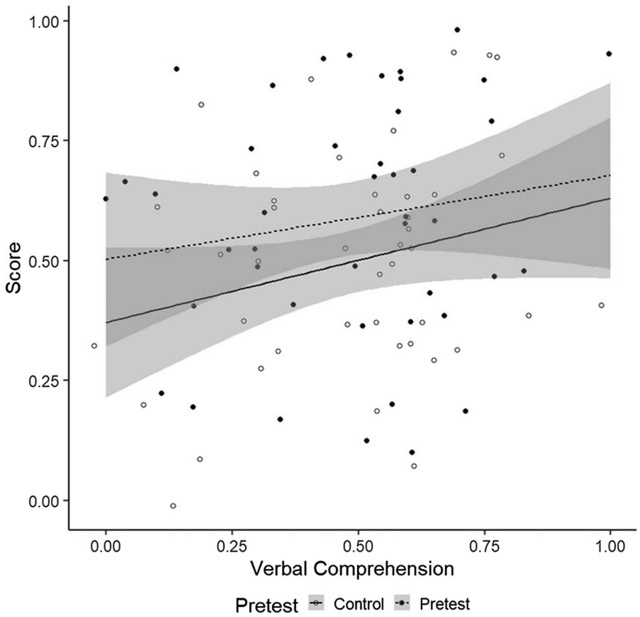
In Experiment 1, we used factual pretest questions to probe self-derivation outcomes. In line with our initial hypothesis and with prior research showing increases in learning from factual pretests (Sana et al., 2021), we observed an effect of pretesting. However, once verbal comprehension was included in the model, the effect of pretesting no longer was statistically significant. This suggests that the difference in levels of success on pretest and control trials diminished after controlling for individual variability associated with verbal comprehension.

Table 1
Task Performance (Proportion of the Maximum Correct) on Control Trials, Pretested Trials, and Verbal Comprehension

			Experiment 3		
Variable	Experiment 1	Experiment 2	Text-only	Text + photo	Experiment 4
Control, M (SD)	.50 (.22)	.44 (.27)	.29 (.25)	.36 (.26)	.52 (.23)
Pretest, M (SD)	.57 (.25)	.50 (.29)	.39 (.26)	.51 (.29)	.68 (.24)
Verbal comprehension, M (SD)	.50 (.23)	.44 (.27)	.50 (.21)		.47 (.18)

Figure 2

Self-Derivation Through Memory Integration and Verbal Comprehension Task Performance (Score) on Control and Pretest Trials in Experiment 1



Note. Bars represent standard error.

Contrary to our initial hypotheses, we did not see an overall effect of verbal comprehension or an interaction between verbal comprehension and the pretest effect. One reason for this might be the type of pretest used. That is, we used factual pretest probes that targeted only one member of the fact pair. Given the relation between verbal comprehension and self-derivation through memory integration performance, we hypothesized that the effect of verbal comprehension might emerge if integration-fact pretest probes were used instead of stem-fact pretests. We tested this hypothesis in Experiment 2.

Experiment 2

Method

Participants

Participants were 38 adults (M_{age} = 20.32; 30 females, five males, three nonbinary) recruited through Prolific. None had participated in Experiment 1. Based on self-report, the sample was Asian (7.89%), Black or African American (2.63%), Mixed or Other (15.79%), and White or Caucasian (71.05%). In total, 2.63% of participants did not report their race. 15.79% of the sample identified as Hispanic or Latinx. The prolific prescreening requirement, consent, and compensation were identical to Experiment 1. An additional 12 participants engaged in study protocols but were excluded due to failure to complete full protocols (N = 11) or evidence of repeat study participation (i.e., they took the survey twice; N = 1).

Stimuli

The stimuli were the same as those used in Experiment 1.

Procedure

The procedure of Experiment 2 was the same as that of Experiment 1 (Figure 1), except for the type of pretest probe used. Specifically, each of the 10 pretest questions probed the *integration fact* of 10 of the 20 to-be-learned trials. For example, the pretest "Who were the first to use the softest known mineral?" probes the fact generated from the integration of the following fact pair: "Talc is the softest known mineral" and "Ancient Egyptians were the first to use talc." During the open-ended self-derivation test phase, participants answered 20 total self-derivation questions, 10 of which were novel and 10 of which had been previously available in the pretest phase.

Scoring

We used the same scoring procedures as those described in Experiment 1.

Results and Discussion

The main goal of Experiment 2 was to assess the influence of integration-fact pretest questions and the role of verbal comprehension on self-derivation outcomes.

Pretest Phase Performance

On average, participants provided successful responses to .8 of the 10 pretest probes (.08 proportion correct). In Experiment 2, all participants scored 85% or higher on the engagement check questions (M = 99.5% correct, range = 92%–100% correct).

Self-Derivation Test Performance

We first calculated performance on pretested and control trials (Table 1). To test for the effect of the pretest manipulation, we conducted a Type III sum of squares repeated-measures ANOVA. We found a main effect of pretesting on test performance, $F(1, 37) = 7.494$, $p = .009$, $\eta_p^2 = .168$.

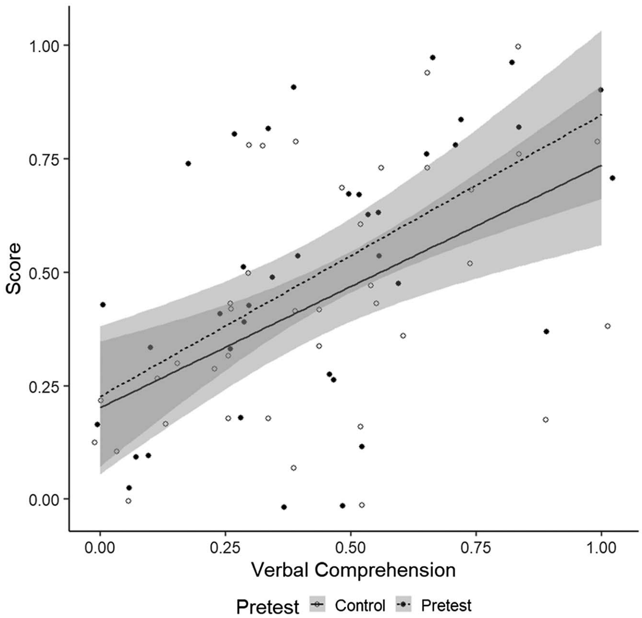
Verbal Comprehension Performance

We next calculated verbal comprehension performance (Table 1) and added it as a covariate to the above model. Unlike in Experiment 1, there was a main effect of verbal comprehension, such that participants with higher verbal comprehension scores performed at higher levels relative to participants with lower verbal comprehension scores, $F(1, 36) = 17.693$, $p < .001$. As we observed in Experiment 1, with verbal comprehension added to the model, there was no longer a statistically significant effect of pretest, $F(1, 36) = .318$, $p = .576$, $\eta_p^2 = .009$. Neither was the interaction of pretest effect and verbal comprehension statistically significant, $F(1, 36) = .994$, $p = .325$, $\eta_p^2 = .027$. Results are illustrated in Figure 3.

As in Experiment 1, in Experiment 2, once the variance associated with verbal comprehension was considered, the effect of pretest no longer was statistically significant. Further, as hypothesized, under integration-fact pretest conditions, there was a main effect of verbal comprehension.

Figure 3

Self-Derivation Through Memory Integration and Verbal Comprehension Task Performance (Score) on Control and Pretest Trials in Experiment 2



Note. Bars represent standard error.

Cross-Study Comparison of Experiments 1 and 2

To provide a stronger test of the effects afforded in each of the experiments separately, we conducted a cross-study comparison of Experiments 1 and 2. A cross-study comparison is appropriate as identical stimuli were used, with the exception of the type of pretest. Additionally, we used the same Prolific screening requirements to source the samples across the two experiments. Finally, to detect possible differences in performance across the two studies, we conducted independent measures t tests. There were no statistically significant differences in control, $t(80) = 1.082$, $p = .283$, $d = .240$, or pretest, $t(80) = 1.439$, $p = .154$, $d = .319$, performance from Experiments 1 and 2. Additionally, there was no difference in verbal comprehension across the two groups, $t(80) = .705$, $p = .483$, $d = .156$.

On the combined samples, we conducted a Type III sum of squares repeated-measures ANOVA and found an effect of pretest, $F(1, 81) = 14.552$, $p < .001$, $\eta_p^2 = .152$. On the combined sample, the main effect of verbal comprehension was significant, $F(1, 80) = 18.393$, $p < .001$. Participants with higher levels of verbal comprehension had higher test scores relative to participants with lower levels of verbal comprehension. As in Experiments 1 and 2 considered separately, once verbal comprehension was added to the model as a covariate, the effect of pretest no longer was statistically significant, $F(1, 81) = 2.655$, $p = .107$, $\eta_p^2 = .032$. Overall, results suggest that once the individual variability in verbal comprehension is accounted for, there is no difference in performance across pretest and control trials.

Experiment 3

In Experiments 1 and 2, the pretest stimuli were individual sentences. We used single-sentence stimuli because they are common in the productive memory paradigm of self-derivation through integration. However, much of the pretest literature features “real-world” stimuli like text passages and videos (Carpenter & Toftness, 2017; Hausman & Rhodes, 2018). To determine whether the patterns observed in Experiments 1 and 2 were specific to the types of stimuli used, in Experiment 3, we tested the influence of pretesting on productive learning outcomes from more “naturalistic” stimuli, specifically text passages, some of which were accompanied by photographs. Research shows that adding a photograph, or other types of visual aids, may influence the way learners interact with educational materials (Mayer, 2002). Thus, in Experiment 3, we also directly compared learning episodes that included a photograph to those that did not.

Method

Participants

Participants were 54 undergraduate students enrolled in introductory psychology courses at a midsize private research university ($M_{\text{age}} = 19.44$; 40 females, 14 males). All participants were recruited via SONA research participation software. None had participated in Experiments 1 or 2. Participants self-reported as Asian (20.4%), Black or African American (12.96%), Mixed Race (11.11%), and White or Caucasian (55.5%). Nine percent of participants self-reported as Hispanic or Latinx. All participants met native English speaker criteria, based on self-report. There were no exclusions from this sample. All participants were compensated with introductory psychology course credit. Written informed consent was obtained from all participants prior to the start of their session.

Stimuli

Passages. Twenty integrable stem-fact passage pairs were developed based on exhibits in an on-campus art history museum. An example stem-fact passage pair is available in Figure 3. Pilot testing ($N = 25$) showed that, for all stimulus pairs, both members of the passage pair were necessary for successful self-derivation. Each passage pair was designed in conjunction with a photograph of a relevant museum exhibit. Although the photograph was always relevant to the overall theme of the passage pair, it was most supportive of one member of the stem-fact passage pair and consequently was only ever presented with that particular passage. For instance, the photograph of the priestess figurehead in Figure 4 directly supported the passage indicating what a priestess might wear. The stem-fact passage pairs are available from the BLISS bank: stimulus numbers P011–P014, P016–P018, P020–P026, P028–P031, and P033–P034.

Ease of readability of each passage pair was scored based on the Flesch–Kincaid readability tests. All passage pairs were adjusted to be at most at a ninth-grade reading level. Length of passages varied (range = 22–40 words, M = 32 words). However, each passage had approximately the same number of words as its separate yet related counterpart ($M_{\text{difference}}$ = 4.2 words). Unlike Experiments 1 and 2, there were no “filler” passages, due to time

Figure 4
Example Stem Fact Passage Pair in Experiment 3

Theme	Learning Phase 1	Learning Phase 2	Test
Priestess	In ancient Rome, **priestesses wore a thickly rolled hair band.** It was used to keep their elaborate curled hairstyle in place. More importantly, it signified their integral role in society. 	**Priestesses watched over the sacred fire in ancient Rome.** The fire was thought to represent the life and soul of the city itself. Priestesses had the integral role of making sure it never went out.	What did the people who watched over the sacred fire in ancient Rome wear? *A: A thickly rolled hairband*

Note. The presence of a photograph depends on the condition (text-only or text + photograph).

constraints. The test of verbal comprehension was identical to that used in Experiments 1 and 2.

We used integration-fact pretest questions to ensure pretest questions would remain the same regardless of whether the trial was presented under text-only or text + photograph conditions. That is, stem-fact pretests differed depending on what type of material the relevant passage pair featured (text-only, text + photograph). Because each passage pair appeared equally often under both text-only and text + photograph conditions, using integration-fact pretests allowed for the most consistent experimental manipulation.

Procedure

Participants were tested in groups of two to four ($M = 3$), by one female experimenter (L.C.G). Group testing was for convenience, and group size was not a variable of interest. Participant groups were randomly assigned to one of four experiment versions. All participants were exposed to 10 text-only stem-fact passage pairs and 10 text + photograph passage pairs. Stem fact passage pairs appeared equally often in text-only and text + photograph formats across versions. Each participant was asked to answer 10 integration-fact pretest questions before the first learning phase. Half of the pretest integration questions probed text-only passage pairs and half probed text + photograph passage pairs. Passages were pretested equally often within format across versions. The design was within subjects such that each participant was exposed to text-only and text + photograph learning episodes and both pretest and no pretest manipulations. The general structure of the session (Figure 1) was the same as Experiments 1 and 2.

Pretest Phase. All participants were given a blank response sheet and a pen. They were spaced in the room so as to prevent onlooking of other students' responses. Participants were instructed,

> You will now be asked to answer some questions. Please try to make a best guess. If you do not know the answer you can leave the space blank or write "idk." You have as much time as you need to answer each one.

As in Experiments 1 and 2, there was no indication made to the explicit relevance of these questions. Participants saw 10 open-ended integration-fact pretest questions projected one at a time on a large screen at the front of the room. The experimenter advanced to the next question once all participants in the group had placed their pen down, indicating they finished writing their response. Pretest questions were counterbalanced such that all stem-fact passage pairs were probed equally often across the sample.

Learning Phase 1. Participants saw 20 passages (five of which were paired with a photograph) projected one at a time on a large screen. They had 15 s to read text-only passages and 20 s to read/view text passages that were presented with a photograph. Pilot participant data (N = 25) established timing constraints. Passages (including those paired with a photograph) were counterbalanced to appear equally often in the first and second learning phases. The order of passages was randomized based on the version of the procedure to which the group was assigned.

Verbal Comprehension Test Phase. Due to the in-person, group nature of the current protocols, the presentation of the items occurred via a PowerPoint projected on a large screen but was otherwise the same as for Experiments 1 and 2.

Learning Phase 2. Participants saw 20 more passages, five paired with a photograph, and all of which were counterparts to passages previously seen in Learning Phase 1. Passages were counterbalanced such that they appeared equally often in the first learning phase as the second across versions. The order of passages was randomized based on the version of the procedure to which the group was assigned.

Self-Derivation Test Phase. Following a 10-min buffer activity, the experimenter presented 20 open-ended self-derivation questions one at a time on the large screen. Ten of the prompts were the same as those experienced in the pretest phase and 10 novel. The order of questions was counterbalanced across versions to control for order effects. Participants wrote their responses on a worksheet. The experimenter advanced to the next question once all group members had put their pens down. The order of questions was randomized based on the version of the procedure to which the group was assigned.

Results and Discussion

There were two main goals of Experiment 3. The first was to assess the influence of integration-fact pretest probes on text passages as opposed to single-sentence stimuli such as those used in the previous experiments. We compared the influence of pretesting on text-only-based stem passages to text and photograph-based stem-fact passages. The second goal was to examine the influence of verbal comprehension on task performance.

Pretest Phase Performance

On average, participants provided successful responses to .3 of the 10 pretest probes (.03 proportion correct).

Self-Derivation Test Performance

Means and standard deviations as a function of pretest and condition are provided in Table 1. A 2 (pretest: pretest, control) × 2 (condition: text-only, text + photograph) Type III sum of squares repeated-measures ANOVA revealed statistically significant effects of pretest, $F(1, 53) = 24.396$, $p < .001$, $\eta_p^2 = .315$, and condition, $F(1, 53) = 11.73$, $p = .001$, $\eta_p^2 = .181$. Participants were more successful on pretested trials and on trials that included photographs. The interaction of the variables was not statistically significant, $F(1, 53) = .738$, $p = .394$, $\eta_p^2 = .014$.

We note that participants were exposed to text passages paired with a photograph for a longer time than those without a photograph (additional 5 s), a purposeful design element, based on pilot performance, to ensure participants had enough time to encode the additional information provided by the photograph. It is possible, however, that this difference in timing contributed to the effect of the condition. Future research should examine this effect while controlling for the timing of the item presentation.

Verbal Comprehension Performance

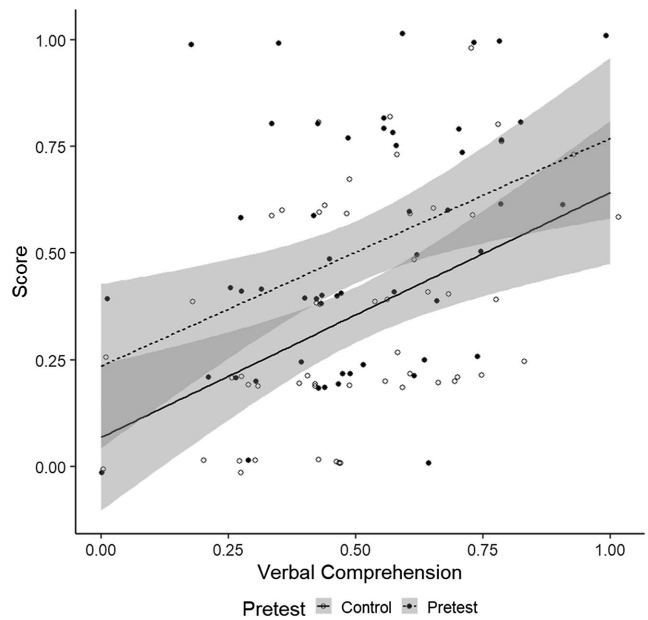
We calculated verbal comprehension task performance (Table 1) and added it as a covariate in the above model. As in Experiment 2, we found an effect of verbal comprehension, $F(1, 52) = 14.13$, $p < .001$. In contrast, the main effects of pretesting and condition were no longer statistically significant, $F(1, 52) = 1.923$, $p = .171$, $\eta_p^2 = .036$, and $F(1, 52) = .080$, $p = .778$, $\eta_p^2 = .002$, respectively. None of the interactions were statistically significant (ps > .088). Results are illustrated in Figures 5 and 6.

Overall, in the present experiment, the results followed a similar pattern as in Experiments 1 and 2. From this, we may conclude that the effects of verbal comprehension were not restricted to the use of single-sentence stimuli. On the contrary, the results suggest that in the context of tests of productive memory processes, the effect of the pretest is diminished once individual variability associated with verbal comprehension is accounted for.

To determine whether verbal comprehension plays a similar role in more conventional pretest paradigms wherein factual pretests appear before learning and are used to probe factual outcomes, in Experiment 4, we used a subset of the single-sentence stimuli used in Experiments 1 and 2 (i.e., one fact of each stem-fact pair) and provided factual pretest questions to probe factual outcomes.

Figure 5

Self-Derivation Through Memory Integration and Verbal Comprehension Task Performance (Score) on Control and Pretest Trials in Experiment 3, Text–Photograph Condition



Note. Bars represent standard error.

Experiment 4

Method

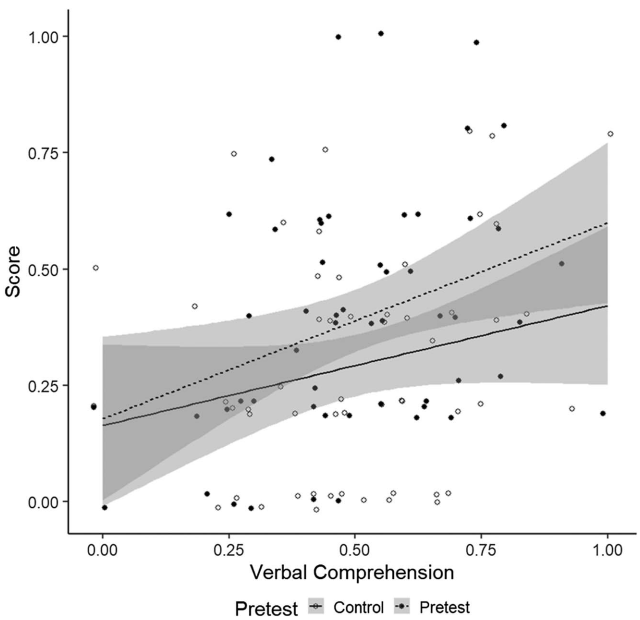
Participants

Participants were 70 adults ($M_{age} = 24.86$; 31 females, 36 males, two nonbinary, one did not report) recruited through Prolific. None had participated in any of the previous experiments. Based on prior research examining factual outcomes, we estimated a small-to-medium effect size. We used G*Power, with .8 power and a Cohen's f of .20 to determine our sample size. Based on self-report, the sample was American Indian or Alaskan Native (1.4%), Asian (14.3%), Black or African American (7.1%), Mixed or Other (4.3%), and White or Caucasian (68.6%). 4.3% of participants did not report their race. 4.3% of the sample identified as Hispanic or Latinx. The Prolific prescreening requirements were the same as for Experiments 1 and 2. All participants engaged in an online consent process before beginning study protocols and were compensated $10.00 per hour of their time.

Stimuli

Stimuli were 20 facts. All 20 facts represented unique information. Facts were nonintegrable and were sourced directly from Experiment 1 stimuli. Specifically, we took one member of integrable stem-fact pairs in Experiment 1 for use in Experiment 4. Ten additional facts were used as "filler" facts, to which engagement questions were

Figure 6
Self-Derivation Through Memory Integration and Verbal Comprehension Task Performance (Score) on Control and Pretest Trials in Experiment 3, Text-Only Condition



Note. Bars represent standard error.

attached. As in Experiments 1 and 2, filler facts were not probed at the pretest or posttest phases.

Procedure

The procedure was the same as in Experiments 1 and 2, with the following adjustments. As in those experiments, participants first answered 10 factual pretest questions. Five pretests probed factual recall of facts to be presented in Learning Phase 1 and five probed facts to be presented in Learning Phase 2. Participants read through 15 facts in each learning phase, five of which were filler. Filler facts and their paired engagement questions were dispersed evenly throughout the learning phases. Otherwise, the presentation of facts was randomized by participants. As in Experiments 1, 2, and 3, the verbal comprehension task occurred between Learning Phases 1 and 2. For the posttest phase, participants answered a total of 20 factual recall questions, 10 of which were novel and 10 of which had been seen in the pretest phase. Across the sample, facts appeared equally often under pretest and control conditions.

Scoring

The scoring procedures were the same as in previous experiments.

Results and Discussion

The major goal of Experiment 4 was to assess the influence of factual pretests on factual learning outcomes while accounting for verbal comprehension. In Experiment 4, all participants scored 85% or higher on the engagement check questions (M_{score} = 99.5% correct, range = 90%–100% correct).

Pretest Phase Performance

On average, participants were successful on 1.1 of the 10 pretest probes (.11 proportion correct).

Factual Test Performance

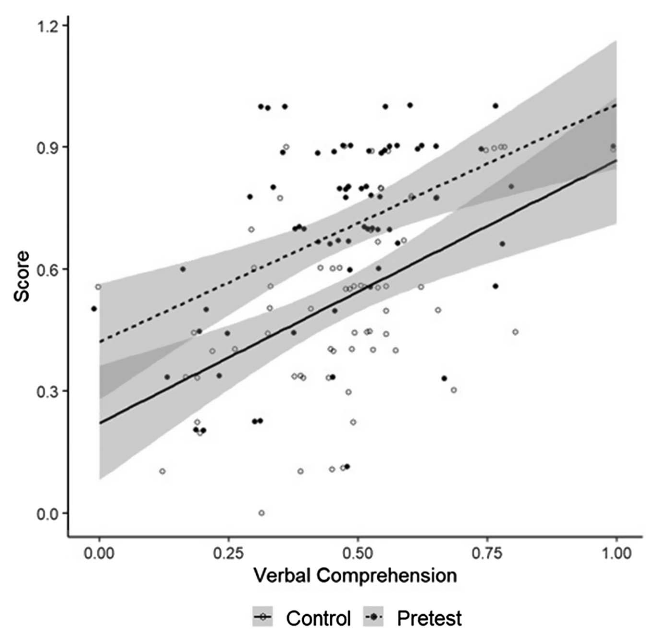
Descriptive statistics on task performance are provided in Table 1. A Type III sum of squares repeated-measures ANOVA revealed a main effect of pretest, $F(1, 69) = 39.199$, $p < .001$, $\eta_p^2 = .362$.

Verbal Comprehension Performance

Two participants were excluded from verbal comprehension analysis due to failure to comply with directions (i.e., they provided synonym answers on the antonyms task). We calculated verbal comprehension performance (Table 1) and added it as a covariate to the above model. There was a significant main effect of verbal comprehension, $F(1, 66) = 27.058$, $p < .001$. The interaction of pretest and verbal comprehension was not statistically significant, $F(1, 66) = .159$, $p = .691$, $\eta_p^2 = .002$. Unlike in Experiments 1, 2, and 3, the main effect of pretest remained statistically significant, $F(1, 66) = 6.60$, $p = .012$, $\eta_p^2 = .091$. Results are illustrated in Figure 7.

In Experiment 4, we used stem-fact stimuli from the self-derivation literature and replicated prior research that found an effect of pretest on factual outcomes. Unlike in previous experiments in this series, the pretest effect was apparent even after controlling for verbal comprehension.

Figure 7
Factual Recall and Verbal Comprehension Task Performance (Score) on Control and Pretest Trials



Note. Bars represent standard error.

General Discussion

A major goal of the present research was to extend the literature on the pretest effect to provide novel insight into the mechanisms by which it enhances learning outcomes. Specifically, we built upon prior work examining the pretest effect on direct factual outcomes to examine its influence on the productive memory process of self-derivation through integration. Additionally, we included measures of individual differences, as indexed by verbal comprehension. We tested the pretest effect under productive memory conditions in Experiments 1–3. In Experiment 4, we tested the effect on factual outcomes. In all four experiments, we considered the role of verbal comprehension, with the expectation that there would be an interaction between verbal comprehension and the pretest effect. We also expected a pretest effect in all four experiments given findings from previous research examining the pretest effect and inference-making (Sana et al., 2021; St. Hilaire et al., 2019).

Discussion of Major Findings

The main effects of pretesting were observed in Experiments 1–3. However, in contrast to our initial hypotheses, they no longer were statistically significant once the individual variability associated with verbal comprehension was taken into account, and we did not observe an interaction between verbal comprehension and the pretest effect.

In the final experiment, rather than the productive memory process of self-derivation through integration, the memory process tested was factual recall; factual recall is the outcome of interest in most studies of pretest effects (e.g., Carpenter & Toftness, 2017; James & Storm, 2019; Kliegl et al., 2024; Pan & Carpenter, 2023; Richland et al., 2009). In the case of factual recall, in the present research, we once again observed a pretesting effect. This time, however, the effect remained statistically significant even once the individual variability associated with verbal comprehension was accounted for. As in Experiments 1–3, in Experiment 4, we did not observe an interaction between verbal comprehension and the pretest effect.

In addition to extension of the literature testing the efficacy of pretesting on the productive memory process of self-derivation through integration, the present research extended the literature to feature consideration of individual differences. Individual differences are relatively understudied in relation to the pretest effect, and research to date presents somewhat of a mixed picture of findings. Geller et al. (2017) showed that, though pretest confidence, academic achievement as measured by GPA, pretest familiarity, and out-of-class topic reading accounted for a significant amount of variance in posttest accuracy, pretest accuracy (whether or not the participant answered the pretest question correctly) was the only significant predictor of test performance. Research by Yang et al. (2021) found that pretest effects were more pronounced among students with high relative to low-achievement motivation. St. Hilaire et al. (2019) showed that, generally, individual differences in structure building positively predicted posttest performance, but there was no interaction between structure building and the pretest effect. St. Hilaire et al. also found that the benefits associated with integrative pretest questions were maintained even after accounting for individual differences in structure building. However, this finding was constrained to experimental design wherein participants were instructed to look for answers within to-be-tested material.

The present research adds to this small body of research on individual differences as they relate to the effects of pretesting. It provides evidence that when learning outcomes based on productive memory processes are the target, the benefit of pretesting is diminished once the variability associated with verbal comprehension is controlled. In contrast, when the target is factual recall, the pretesting manipulation drives a significant difference in performance on pretest and control trials even after variability associated with verbal comprehension is accounted for. We targeted the individual difference of verbal comprehension for test of its relation with pretesting. We did so in part because verbal comprehension has been shown to relate to performance in the productive process of self-derivation through integration (Varga et al., 2019). As well, verbal comprehension is a measure both of semantic knowledge and of the ability to derive meaning from text (Schrank & Wendling, 2018) and is directly applicable to potential mechanisms of pretest questions (e.g., targeted activation of preexisting semantic knowledge to facilitate encoding of new information).

There are numerous features of the present research that give us confidence in the findings. First, the effects of pretesting such as reported in the broader literature were apparent in each of the four experiments. In the present research, the findings were generalized across different populations (SONA and Prolific) and different testing formats (in-person and online). Second, the finding that verbal comprehension is related to self-derivation performance (Experiments 2 and 3) is a replication of patterns reported in the literature (Varga et al., 2019). The results of Experiment 4 suggest that verbal comprehension also is related to the learning outcome of factual recall. Third, the finding that pretesting effects were diminished once verbal comprehension was controlled for was observed in each of the experiments in which the learning outcome was based on a productive process. In Experiment 1, the pretested items were facts that later would be explicitly taught to the participants, whereas in Experiments 2 and 3, the pretested items were facts that participants were expected to self-derive. As well, in Experiments 1 and 2, the tests were over single sentences, whereas in Experiment 3, the test was over brief text passages, some of which featured photographs. Thus, the impact of verbal comprehension was apparent across probe and stimulus types.

Implications

Our findings have novel implications for open questions in theory on mechanisms and limitations of pretesting. Pretesting is thought to work through myriad mechanisms, including highlighting knowledge gaps, activating prior knowledge structures to facilitate encoding of new information, and potentially strengthening connections between pretested and related material. The majority of research to date has focused on the pretest effect and factual test outcomes. For instance, one meta-analysis consisting of 76 records (journal articles, conference presentations, theses/dissertations, and unpublished data sets) revealed a moderate specific effect of pretesting on factual outcomes (St. Hilaire et al., 2023). Another review examined effect sizes across six published empirical articles and showed Cohen's d measures of pretest effect sizes ranging from .40 to 2.68 (Pan & Carpenter, 2023). Yet, not all learning is restricted to facts and recall thereof. Some research shows that pretesting may not be as efficacious for inference-making tasks (Hausman & Rhodes, 2018). The present research corroborates this finding and provides insight into why this might be. When posttest questions required learners to combine two facts, individual differences in verbal comprehension (e.g., semantic knowledge) were more salient to test success than pretesting. Indeed, verbal comprehension was critical for test success regardless of whether the test question was pretested prior to the onset of learning (i.e., lack of an interaction effect). Yet, when posttest questions required recall of only one fact (Experiment 4), both pretesting and verbal comprehension were salient to test success.

Overall, results accentuate the importance of the test outcome to the pretest effect. This conclusion can be seen most directly in comparison to the results of Experiments 1 and 4. In Experiment 4, we used a subset of the pretests that we used in Experiment 1. We also used the same learning materials. As such, it is reasonable to speculate that the participants in the two experiments engaged in the pretest and learning phases in similar ways. The primary difference between the experiments was the test outcome of interest—whether it involved combination of two facts (Experiment 1) or recall of one fact (Experiment 4)—a difference that ultimately impacted overall pretest efficacy. We conclude that because productive memory processes, such as self-derivation, require increased recruitment of semantic knowledge compared to processes of direct factual recall, the individual difference in verbal comprehension was prominent relative to the experimental manipulation of pretesting.

Our findings imply that an important qualifier needs to be added to the conclusion that pretesting improves performance. It does so for factual recall (Experiment 4). However, in the case of more productive processes (Experiments 1–3), its impact may be overshadowed by individual variability in verbal comprehension. We suggest then that interventions to improve test performance beyond factual recall may be more effective if they also focus on enhancing student's verbal comprehension.

Future Directions

In the present research, we focused on the self-derivation of new knowledge through integration and pretesting. It will be left to future research to test whether similar patterns are observed for other productive learning processes such as inferential reasoning—a process that requires integrating a new fact into existing knowledge rather than combining two new facts. Additionally, there is some research examining how other means of test-potentiated learning, such as retrieval practice, facilitate learning beyond factual recall. As is the case for the pretest literature, results are mixed. Some researchers found that the benefits of retrieval practice do not extend to deductive inferences (Tran et al., 2015), and others found the opposite, once educational materials were optimized for relational reasoning (Eglington & Kang, 2018). Future work should examine not only the format of educational material but also the role of individual differences learners bring to the table, to consider the extent to which other common learning strategies, such as retrieval practice support learning beyond factual recall.

Constraints on Generalizability

The present research has numerous strengths, yet it is not without limitations. The participants in the research were comparable across experiments, yet in all experiments, they were of college age. This limits the extent to which the findings of the research can be generalized to other age groups. We note that this limitation is not specific to the present research: Much of the research on pretesting has been conducted with college students. Second, the extension of the findings to productive processes beyond self-derivation through memory integration (e.g., inference) is not clear and remains an open line for future research. Third, the current protocols occurred online or in the laboratory, and thus, the potential implications of findings in more applied educational settings such as classrooms remain an important direction for future work. Last, to provide a stronger test of relations, we conducted a cross-study comparison collapsing data across Experiments 1 and 2; we did not include Experiment 3 because of differences in stimulus type (single sentence vs. text passages and photographs) and method of administration (e.g., online vs. in-person). The results of the cross-study comparison were consistent with those of each experiment individually, providing confidence in the findings. However, to provide an even stronger test for the generalizability of findings, future work might include larger sample sizes.

Conclusions

In conclusion, in the present research, we took the important step of extending the literature on pretesting beyond the learning outcome of factual recall to that of performance based on productive processes. Productive processes permit the expansion of knowledge beyond that which is directly given, thus serving as a "force multiplier" to facilitate learning. We also extended the literature to consider the individual difference variable of the level of verbal comprehension. Learners of all ages and in all settings can be expected to vary in this skill, making it an important target for consideration. Consistent with this suggestion, in three of the four experiments, effects of verbal comprehension were observed such that individuals with higher verbal comprehension scores demonstrated greater overall learning. Importantly, consideration of variability in verbal comprehension provided an important caveat to the seemingly pervasive finding that learning is facilitated by pretesting per se. That is, in the case of learning outcomes based on productive processes, the difference in levels of success on control compared to pretest trials diminished once the variance associated with verbal comprehension was accounted for. The findings of the present research should serve as motivation for further research on the interaction of individual difference variables and the effects of learning manipulations such as pretesting. Such research will permit more impactful interventions targeted to those who may benefit from them the most.

References

Bauer, P. J. (2020). *Bauer Lab Integration and Self-derivation Stimulus (BLISS) bank*. https://doi.org/10.31234/osf.io/rv9n7

Bauer, P. J. (2021). We know more than we ever learned: Processes involved in accumulation of world knowledge. *Child Development Perspectives*, *15*(4), 220–227. https://doi.org/10.1111/cdep.12430

Bauer, P. J., Dugan, J. A., Varga, N. L., & Riggins, T. (2019). Relations between neural structures and children's self-derivation of new knowledge through memory integration. *Developmental Cognitive Neuroscience*, *36*, Article 100611. https://doi.org/10.1016/j.dcn.2018.12.009

Bauer, P. J., & Jackson, F. L. (2015). Semantic elaboration: ERPs reveal rapid transition from novel to known. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 271–282. https://doi.org/10.1037/a0037405

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563–1569. https://doi.org/10.1037/a0017021

Carpenter, S. K., & Toftness, A. R. (2017). The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory and Cognition*, *6*(1), 104–109. https://doi.org/10.1016/j.jarmac.2016.07.014

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*(11), 1111–1146. https://doi.org/10.1037/bul0000166

Cronin-Golomb, L. M. (2024, July 11). *Individual differences diminish the pretest effect under productive memory conditions*. https://osf.io/nz3ae

Cronin-Golomb, L. M., & Bauer, P. J. (2022). Support for learning under naturalistic conditions. *Cognitive Research: Principles and Implications*, *7*(1), Article 86. https://doi.org/10.1186/s41235-022-00435-0

Cronin-Golomb, L. M., & Bauer, P. J. (2023). Self-motivated and directed learning across the lifespan. *Acta Psychologica*, *232*, Article 103816. https://doi.org/10.1016/j.actpsy.2022.103816

Dugan, J. A., & Bauer, P. J. (2022). This should help with that: A behavioral investigation into self-derivation of knowledge about prescription medications. *Applied Cognitive Psychology*, *36*(2), 378–390. https://doi.org/10.1002/acp.3926

Eglington, L. G., & Kang, S. H. K. (2018). Retrieval practice benefits deductive inference. *Educational Psychology Review*, *30*(1), 215–228. https://doi.org/10.1007/s10648-016-9386-y

Esposito, A. G., & Bauer, P. J. (2022). Determinants of elementary-school academic achievement: Component cognitive abilities and memory integration. *Child Development*, *93*(6), 1777–1792. https://doi.org/10.1111/cdev.13819

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Geller, J., Carpenter, S. K., Lamm, M. H., Rahman, S., Armstrong, P. I., & Coffman, C. R. (2017). Prequestions do not enhance the benefits of retrieval in a STEM classroom. *Cognitive Research: Principles and Implications*, *2*(1), Article 42. https://doi.org/10.1186/s41235-017-0078-z

Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy. *American Psychologist*, *52*(1), 32–34. https://doi.org/10.1037/0003-066X.52.1.32

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*(4), 505–513. https://doi.org/10.3758/s13421-011-0174-0

Han, M., & Kim, H. (2019). Elementary students' modeling using analogy models to reveal the hidden mechanism of the human respiratory system. *International Journal of Science and Mathematics Education*, *17*(5), 923–942. https://doi.org/10.1007/s10763-018-9895-x

Hausman, H., & Rhodes, M. G. (2018). When pretesting fails to enhance learning concepts from reading texts. *Journal of Experimental Psychology: Applied*, *24*(3), 331–346. https://doi.org/10.1037/xap0000160

James, K. K., & Storm, B. C. (2019). Beyond the pretesting effect: What happens to the information that is not pretested? *Journal of Experimental Psychology: Applied*, *25*(4), 576–587. https://doi.org/10.1037/xap0000231

JASP Team. (2023). *JASP* (Version 0.17.2.1) [Computer software]. https://jasp-stats.org/

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Chapter seven—Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 61, pp. 237–284). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kliegl, O., Bartl, J., & Bäuml, K.-H. T. (2024). The pretesting effect comes to full fruition after prolonged retention interval. *Journal of Applied Research in Memory and Cognition*, *13*(1), 63–70. https://doi.org/10.1037/mac0000085

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 106–114. https://doi.org/10.1037/a0033699

Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition*, *44*(7), 1085–1101. https://doi.org/10.3758/s13421-016-0621-z

Little, T. D. (2013). *Longitudinal structural equation modeling* (Methodology in the social sciences). The Guilford Press.

Mayer, R. E. (2002). Multimedia learning. *Psychology of learning and motivation*. Academic Press. https://doi.org/10.1016/S0079-7421(02)80005-6

McDaniel, M. A., Marsh, E. J., & Gouravajhala, R. (2022). Individual differences in structure building: Impacts on comprehension and learning, theoretical underpinnings, and support for less able structure builders. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *17*(2), 385–406. https://doi.org/10.1177/17456916211000716

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*(1), 465–489. https://doi.org/10.1146/annurev-psych-010416-044022

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia*, *138*, Article 107296. https://doi.org/10.1016/j.neuropsychologia.2019.107296

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and pretesting effects: A review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review*, *35*(4), Article 97. https://doi.org/10.1007/s10648-023-09814-5

Pan, S. C., & Rivers, M. L. (2023). Metacognitive awareness of the pretesting effect improves with self-regulation support. *Memory & Cognition*, *51*(6), 1461–1480. https://doi.org/10.3758/s13421-022-01392-1

Pan, S. C., Schmitt, A. G., Bjork, E. L., & Sana, F. (2020). Pretesting reduces mind wandering and enhances learning during online lectures. *Journal of Applied Research in Memory and Cognition*, *9*(4), 542–554. https://doi.org/10.1016/j.jarmac.2020.07.004

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reynolds, M. R., & Turek, J. J. (2012). A dynamic developmental link between verbal comprehension-knowledge (Gc) and reading comprehension: Verbal comprehension-knowledge drives positive change in reading comprehension. *Journal of School Psychology*, *50*(6), 841–863. https://doi.org/10.1016/j.jsp.2012.07.002

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15*(3), 243–257. https://doi.org/10.1037/a0016496

Rickards, J. P., & McCormick, C. B. (1988). Effect of interspersed conceptual prequestions on note-taking in listening comprehension. *Journal of Educational Psychology*, *80*(4), 592–594. https://doi.org/10.1037/0022-0663.80.4.592

Sana, F., Yan, V. X., Clark, C. M., Bjork, E. L., & Bjork, R. A. (2021). Improving conceptual learning via pretests. *Journal of Experimental Psychology: Applied*, *27*(2), 228–236. https://doi.org/10.1037/xap0000322

Schrank, F. A., & Wendling, B. J. (2018). The Woodcock–Johnson IV: Tests of cognitive abilities, tests of oral language, tests of achievement. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 383–451). Guilford Press.

Soderstrom, N. C., & Bjork, E. L. (2023). Pretesting enhances learning in the classroom. *Educational Psychology Review*, *35*(3), Article 88. https://doi.org/10.1007/s10648-023-09805-6

St. Hilaire, K. J., & Carpenter, S. K. (2020). Prequestions enhance learning, but only when they are remembered. *Journal of Experimental Psychology: Applied*, *26*(4), 705–716. https://doi.org/10.1037/xap0000296

St. Hilaire, K. J., Carpenter, S. K., & Jennings, J. M. (2019). Using prequestions to enhance learning from reading passages: The roles of question type and structure building ability. *Memory*, *27*(9), 1204–1213. https://doi.org/10.1080/09658211.2019.1641209

St. Hilaire, K. J., Chan, J. C. K., & Ahn, D. (2023). Guessing as a learning intervention: A meta-analytic review of the prequestion effect. *Psychonomic Bulletin & Review*, *31*, 411–441. https://doi.org/10.3758/s13423-023-02353-8

Tran, R., Rohrer, D., & Pashler, H. (2015). Retrieval practice: The lack of transfer to deductive inferences. *Psychonomic Bulletin & Review*, *22*(1), 135–140. https://doi.org/10.3758/s13423-014-0646-x

Varga, N. L., & Bauer, P. J. (2013). Effects of delays on 6-year-old children's self-generation and retention of knowledge through integration. *Journal of Experimental Child Psychology*, *115*(2), 326–341. https://doi.org/10.1016/j.jecp.2013.01.008

Varga, N. L., & Bauer, P. J. (2017). Young adults self-derive and retain new factual knowledge through memory integration. *Memory & Cognition*, *45*(6), 1014–1027. https://doi.org/10.3758/s13421-017-0711-6

Varga, N. L., Esposito, A. G., & Bauer, P. J. (2019). Cognitive correlates of memory integration across development: Explaining variability in an educationally relevant phenomenon. *Journal of Experimental Psychology: General*, *148*(4), 739–762. https://doi.org/10.1037/xge0000581

Wilson, J. T., & Bauer, P. J. (2021). Prompt-facilitated learning: The development of unprompted memory integration and subsequent self-derivation. *Memory & Cognition*, *49*(7), 1473–1487. https://doi.org/10.3758/s13421-021-01155-4

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Riverside Publishing.

Yang, J., Zhang, Y., Pi, Z., & Xie, Y. (2021). Students' achievement motivation moderates the effects of interpolated pre-questions on attention and learning from video lectures. *Learning and Individual Differences*, *91*, Article 102055. https://doi.org/10.1016/j.lindif.2021.102055

Received November 7, 2023
Revision received July 17, 2024
Accepted July 24, 2024 ■