Distribution Agreement

Tulasi Kadiyala March 13, 2020

Self-Derivation as a Tool to Enhance Learning

by

Tulasi Kadiyala

Dr. Patricia Bauer
Adviser

Psychology

Dr. Patricia Bauer

Adviser

Dr. Jessica Barber

Committee Member

Dr. Patrick Cafferty

Committee Member

2020

Self-Derivation as a Tool to Enhance Learning

By

Tulasi Kadiyala

Dr. Patricia Bauer

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Psychology

2020

Abstract

Self-Derivation as a Tool to Enhance Learning
By Tulasi Kadiyala

Retrieval practice, also known as the testing effect, is an established tool used to enhance learning by asking questions in order to actively recall learned information. Memory integration means combining facts learned across separate learning episodes, and self-derivation through memory integration refers to the process of deriving a novel fact by merging these facts. There has been little research comparing the roles of retrieval practice and self-derivation through integration in promoting long-term retention and knowledge extension. Because retrieval practice requires the recall of information, it represents a shallower level of processing. On the other hand, self-derivation represents a deeper level of processing because it requires memory integration. In the current study, we aimed to measure the effectiveness of different levels processing in promoting long-term retention and knowledge extension. We hypothesized that a deeper level of processing will promote knowledge extension while a shallower level of processing will promote long-term retention of information. Before we investigated this question, we conducted Experiment 1A to validate stimuli that are used in Experiment 1B (which investigates this question). Stimulus sets that met certain criteria (see methods section) continued to be used in Experiment 1B, in addition to stimulus sets that had been validated in prior studies. Experiment 1B (like Experiment 1A) used a recursive integration paradigm (see Appendix Fig. 2) to compare a control condition (no questions), retrieval practice condition (fact questions), and self-derivation condition (integration questions) in performance on 2-day or 7-day delayed stem fact questions (to measure retention) and integration questions (to measure knowledge extension). There were three major findings from Experiment 1B – initial performance in the retrieval practice condition was better than initial performance in the self-derivation condition, those who performed well initially also tend to perform well in delayed testing, and the self-derivation condition yielded significantly better performance on stem fact questions than the retrieval practice condition or control condition. The last finding in particular was interesting because it illustrates the utility of self-derivation in promoting learning. These results should be interpreted cautiously and more studies are needed to validate these results.

Self-Derivation as a Tool to Enhance Learning

By

Tulasi Kadiyala

Dr. Patricia Bauer

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Psychology

2020

Acknowledgements

My thanks to Dr. Patricia Bauer, my advisor, who took time out of her busy schedule to help me examine the connection between self-derivation, retrieval practice, and learning. In addition to my other committee members, Dr. Jessica Barber and Dr. Patrick Cafferty, Julia Wilson, a graduate student at the Bauer Lab, has been very helpful and supportive with this project, for which I am grateful.

Table of Contents

Introduction

A major goal of education is learning. To successfully learn and apply knowledge, it is necessary to build an effective knowledge base. The generation of this knowledge base often relies on a process called memory integration, which involves the coalescence of facts learned across separate learning episodes to create an integrated representation. Then, self-derivation through integration is a process applied over this integrated representation to derive new knowledge (Varga & Bauer, 2017; Varga, Esposito, Bauer, 2019). As students progress through the educational system, they are expected to not only memorize facts, but also self-derive new information from facts learned in the classroom. There has been debate among the educational community regarding the most effective methods of facilitating learning to enhance the retention and application of knowledge. Therefore, it is important to compare and investigate methods of learning to find the most efficient ones.

One such method of facilitating learning is retrieval practice, the practice of actively recalling learned material through strategies such as testing or free recall. There has been research suggesting that retrieval practice enhances the memorization of facts (Agarwal et al., 2017; Agarwal, 2019; Tran, Rohrer, Pashler, 2015; Whiffen & Karpicke, 2017). However, there has been limited research testing how the depth of processing affects long-term retention of material and the ability to self-derive new knowledge. The levels of processing model was proposed by Craik and Lockhart (1972) and in this model depth referred to the meaning that was obtained from a stimulus rather than knowledge of the stimulus itself (Craik & Lockhart, 1972). For example, we hypothesize that being asked to recall previously learned information is a shallow level of processing, whereas being asked to self-derive new information from previously learned facts is a deeper level of processing. Self-derivation is separate from retrieval practice and is likely a deeper form of processing because it requires the application of information and is linked to cognitive skills (Varga, Esposito, & Bauer, 2019). The effects that these different levels of processing –retrieval practice and self-derivation through memory integration – have on future test performance (both tests that require integration as well as recall of presented information) has yet to be explored. To address this gap in literature, the current study aims to investigate the effects of different levels of processing on test performance by altering question type – retrieval practice questions, self-derivation questions, or no questions. Any differences between these conditions in the future ability to recall information or self-derive information through memory integration will then be assessed.

Previous research suggests that retrieval practice may be of substantial use in enhancing long-term memory for presented information (Agarwal, Finley, Rose, & Roediger III, 2017; Coane, 2013; Hanawal, 1937; Roediger & Karpicke, 2016). Retrieval practice, also known as the testing effect, has been shown to be more effective in facilitating learning than restudy of material (Agarwal et al., 2017; Agarwal, 2019; Brunyé, Smith, Hendel, Gardony, Martis, & Taylor, 2019; Carpenter, Pashler, & Cepeda, 2009; Coane, 2013; Smith, Blunt, Karpicke, Blunt, & Smith, 2016; Whiffen, & Karpicke, 2016; Tran, Rohrer, Pashler, 2015; Whiffen & Karpicke, 2017). For example, participants in Whiffen & Karpicke (2017) studied a list of words during a learning phase. After this learning phase, participants either restudied the word list or were asked to make list discrimination judgements – meaning they were asked to choose the words that had appeared in the original list from a list of words. List discrimination judgement questions were used as a form of retrieval practice. Then, after a buffer task, participants were given a free recall test in order to measure the number of words they could remember from the original list. The results showed that those in the list discrimination judgement condition performed significantly better on the free recall test than those in the restudy condition with an effect size of $d = 0.62$, which fell within the range of Cohen's (1988) convention for a medium effect ($d = 0.60$).

Other experiments have also found a medium effect size for the benefit of retrieval practice over restudy (e.g., Brunyé et al., 2019). However, some studies comparing the benefit of retrieval practice over restudy in terms of final test performance have found large effect sizes (Agarwal, 2019; Smith et al., 2016; Karpicke, Blunt, & Smith, 2016; Tran, Rohrer, & Pashler, 2015) as well as small effect sizes (Agarwal et al., 2017; Carpenter, Pashler, & Cepeda, 2009; Coane, 2013). This variability in effect sizes may be due to the different forms of testing employed for retrieval practice as well as different measures for final test performance between experiments. Therefore, retrieval practice appears to be better than a restudy condition in enhancing long-term memory, but the extent to which it enhances long-term memory is not clear.

As early as 1937, Hanawalt (1937) demonstrated the usefulness of retrieval practice in remembering geometric information. More recent studies have also reproduced this result of retrieval practice promoting recall in a variety of populations. For example, in Agarwal (2019), Carpenter et al. (2006), and Chan et al. (2006), researchers compared testing and restudy conditions in students from middle school through college and found that the participants in the testing condition consistently recalled information better than participants in the restudy condition. In addition, Agarwal et al. (2017) found that the longer the recall delay (ex: 2 days vs. 10 minutes), the stronger this effect seems to be. A regression analysis revealed that for both restudy and retrieval practice conditions, the average amount of information recalled with the 10-minute delay was higher than the average amount of information recalled with the 2-day delay. This means as more time went by, participants tended to forget more information. When separated by condition, researchers found that the difference in performance between the restudy and retrieval practice conditions was not significant at the 10-minute delay period, but the difference in performance between the restudy and retrieval practice conditions was significant at the 2-day delay period (Agarwal et al., 2017). This means that in the absence of retrieval practice, the longer the delay, the greater the forgetting of information. These findings imply that the longer the delay period, the stronger the effect of retrieval practice (Agarwal et al., 2017).

Other studies have specifically demonstrated that the type of question asked during retrieval practice affects delayed test performance. For example, Agarwal et al. (2017) examined differences in effectiveness of fact questions, higher order, or a mix of fact and higher order questions in promoting higher order learning. In the context of this study, fact questions and higher order questions refer to Bloom's Taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Bloom et al. (1956) claimed that lower-order processes (like memorization) must occur before being able to reach higher-order processes (like analysis, synthesis, and evaluation). Agarwal et al. (2019) developed questions that require the use of lower-order processing and questions that require the use of higher-order processing. In this study, participants were instructed to read passages and then engaged in retrieval practice in the form of fact questions, higher order questions, or a mix of both types of questions about the passage. There were also groups that restudied the passage as a control. After a 2-day delay, they retested all participants on both fact questions and higher order questions. For example, a fact question was, "Which is the primary reason the "yes" author is against welfare programs?" because the answer to this question was stated directly in the passage. A higher order question was, "How do you predict the "yes" author would react if he or she became unemployed and needed welfare assistance?" because this question required readers to predict an outcome that was not stated in the passage.

Agarwal (2019) showed that participants engaging in any form of retrieval practice performed better than participants restudying the material on both final tests (Agarwal, 2019). However, results also provided evidence that the type of question asked during retrieval practice – fact questions or higher order questions – affected delayed test performance (Agarwal, 2019). That is, higher-order learning questions seemed to enhance later performance on test questions that go beyond the text, and fact questions seemed to enhance later performance on fact questions (e.g., memory for the text itself; Agarwal, 2019). Further, fact questions and restudy of material yielded equally low performance on delayed higher-order processing questions whereas higher-order processing questions and restudy of material yielded equally low performance on delayed fact questions (Agarwal, 2019). In other words, retrieval practice using fact questions appears to enhance delayed performance on fact questions but not higher-order questions, and retrieval practice using higher-order questions appears to enhance delayed performance on higher-order questions but not fact questions (Agarwal, 2019).

Tran et al. (2015) also showed a similar trend of a fact-based retrieval practice condition having no benefit over a restudy condition on future performance on higher order questions. In this study, participants were instructed to read scenarios with a number of premises. Premises refer to facts related to a specific topic area. Then, they were either told to reread the scenario in the restudy condition, or told to recall and write down a missing key word in the retrieval practice condition. After these learning phases, they were given multiple-choice questions that could only be answered by creating inferences between the presented premises (Tran et al. 2015). There was no significant difference on inference question performance between the restudy and retrieval practice conditions, further illustrating that fact-based retrieval practice may not be very useful for promoting higher order learning (Tran et al. 2015).

The literature seems to suggest that deeper processing of material may facilitate retention above and beyond retrieval. One such form of processing that may enhance later retention—as well as the ability to build effectively on existing knowledge—is self-derivation through memory integration. Self-derivation through integration is the process through which novel information is derived through facts learned in separate episodes. Thus, processes such as self-derivation may be instrumental in promoting retention and use of learned knowledge by promoting knowledge extension and adding novel information to a knowledge base. Self-derivation is both a model for productive knowledge extension and reliant upon memory integration. For example, if two facts such as, “pumice is a type of volcanic rock” and “volcanic rock floats in water” are presented across separate learning episodes, integration would require the incorporation of both of these facts into an integrated knowledge base about volcanic rock, and self-derivation comprises of self-deriving the fact, “pumice floats in water.”

Self-derivation is a useful process to investigate because it has been linked to academic outcomes and cognitive skills (Varga, Esposito, & Bauer, 2019). In a study by Varga, Esposito, and Bauer (2019), researchers investigated whether there was any correlation between cognitive factors and self-derivation levels in young adults and children. They found that verbal comprehension and verbal working memory span accounted for a large percentage of variance in self-derivation levels for adults, and they also found that self-derivation levels had a large association with future verbal SAT scores and a small to moderate association with future GPA (Varga, Esposito, Bauer, 2019). This shows that self-derivation is ecologically valid because it is linked to academic outcomes and cognitive skills (Varga, Esposito, Bauer, 2019). Therefore, using self-derivation as a way to measure higher-order skills can give rise to real-world implications. Children as young as 4 years of age can self-derive (Bauer & Souci, 2010) and remember self-derived information after a 1-week delay (Varga, Stewart, & Bauer, 2016). Varga & Bauer (2017) also measured self-derivation levels in young adults by using related fact pairs. After a 1-week delay, they asked questions that required the use of self-derivation through memory integration to be answered correctly. They found that newly derived information tended to remain accessible for young adults even after this 1-week delay period (Varga & Bauer, 2017). These studies show that self-derivation through integration has the potential to facilitate learning.

In the current study, integration questions are used to measure self-derivation and deeper processing of information. Integration questions prompt participants to synthesize information across multiple sources to self-derive a novel fact, thus representing deeper processing. Fact questions, on the other hand, require recall of the information already learned and thus represent shallow processing. There have been many studies investigating the effects of different types of retrieval practice on later performance on recall and higher order learning questions. However, there have not been any studies comparing the effects of retrieval practice and self-derivation through memory integration on subsequent retention and extension of learned knowledge. This comparison is especially interesting because it has the potential to provide insights into how humans learn and integrate information to establish a knowledge base across different levels of processing. Whereas retrieval practice represents a shallow level of processing, self-derivation represents a deeper level of processing. Thus, the aim of the current study was to investigate the effects of retrieval practice and self-derivation on future performance on fact questions and integration questions. This was done by comparing delayed memory performance on an integration test and fact test between items for which we asked integration questions (self-derivation condition), fact questions (retrieval practice condition), and no questions (control).

To perform this study, we used a novel self-derivation behavioral paradigm called recursive integration. Recursive integration is the sequential integration of four novel, related semantic facts (see methods section). Two integration questions are used in this paradigm – the first question prompts the integration of two facts to self-derive a novel fact. The second integration question requires the integration of the newly self-derived fact and a third, related fact to self-derive a second novel fact. The first integration question, prompting integration of the first two facts, is only presented during the self-derivation condition. A fact question and no question is asked in retrieval practice and control conditions, respectively. The second integration question is presented after a delay for all conditions. The benefit of this paradigm is that a second, novel integration question (rather than the previously used integration question) can be used to assess knowledge extension following the first test phase. This feature is important because it allows researchers to measure knowledge extension without using integration questions that participants were exposed to during the first test phase.

In the current study, we first undertook a control experiment (1A) to examine whether our stimuli required memory integration for high performance. Next, we performed an experiment (1B) to examine the effects of three testing conditions – retrieval practice, self-derivation, and no testing (control) – on future performance on delayed integration questions and delayed stem questions. Delayed performance on both integration questions and stem questions were examined to test knowledge extension (through integration questions) and retention (through stem questions); these are both important aspects of learning.

We expected that for the items in the self-derivation condition, performance would be high on the final integration test. This result was expected because integration questions, which represent a deeper level of processing, are asked during the self-derivation condition as well as the final integration test. Further, for items in the retrieval practice condition, we expected performance to be high on the final fact test. This result was expected because stem questions, which represent a shallow level of processing, are asking during the retrieval practice condition as well as the final fact test. Consistent with prior retrieval practice literature, items presented in either the retrieval practice or the self-derivation conditions were expected to yield higher performance on the final tests than fact sets in the control condition (Agarwal et al., 2017; Agarwal, 2019; Brunyé, Smith, Hendel, Gardony, Martis, & Taylor, 2019; Carpenter, Pashler, & Cepeda, 2009; Coane, 2013; Smith, Blunt, Karpicke, Blunt, & Smith, 2016; Whiffen, & Karpicke, 2016; Tran, Rohrer, Pashler, 2015; Whiffen & Karpicke, 2017). Participants were also expected to perform better during the first session on questions asked in the retrieval practice condition than in the self-derivation condition because retrieval practice requires shallower processing (and therefore requires less cognitive effort) whereas self-derivation requires deeper processing (and therefore more cognitive effort). Finally, prior studies have shown that the effects of the testing effect appear to be stronger after a longer delay period (Agarwal et al., 2017). So, we expected that both retrieval practice and self-derivation will be enhanced for participants in the longer 7-day delay condition compared to the participants in the shorter 2-day delay condition.

Experiment 1A

Method

Participants

The final sample for Experiment 1A consisted of 29 undergraduate students (21 female) at a mid-sized university recruited online through the university psychology department's SONA system. Five additional participants were tested but excluded due to falling asleep during the learning phases (*n* = 1), and experimental error (*n* = 4). All participants were enrolled in introductory psychology courses at the time of participation. Participants ranged in age from 18 to 20 years (*M* = 18.88, *SD* = 0.67). Two participants did not enter their birth date on the demographic form, so their data has been excluded from this calculation. 24.14% identified as Asian/Pacific Islander, 10.34% identified as Black/African American, 44.83% identified as White/Caucasian, and 3.45% did not report their race. 6.90% of participants were Hispanic or Latinx, 89.66% were not Hispanic or Latinx, and 3.45% did not report their ethnicity. All participants were consented before the study. Participants were compensated for their participation with course credit. Protocols and procedures were approved by the university Institutional Review Board.

Materials and Measures

Stimuli

Stimuli were 30 fact sets comprised of 4 novel, related facts (“stem facts”). There were 2 stem-fact pairs per set (A1, A2; A/B1, B2; see Fig. 1). Stem-fact pairs are 2 facts that can be integrated to self-derive a novel “integration” fact. A1 and A2 is the stem-fact pair that can be integrated to self-derive the A/B1 integration fact, and A/B1 can be subsequently paired with the B2 stem fact to self-derive the B integration fact (see Fig. 1).

Buffers

Two buffer activities were used during the experiment to allow filler periods between learning phases as well as filler periods between learning and test phases. The working memory task, a group-administered lag task adapted from Shelton, Metzger, and Elliot (2007), was the first buffer task. The cloze sentence probability task, with stimuli adapted from Block and Baldwin (2010), was the second buffer task.

Procedure

There were 2 conditions presented in each version of the test: the 1-stem condition and the 2-stem condition. Both stem facts were presented for 15 of the 30 stem-fact pairs in each presentation order for the 2-stem condition, and only one of the two stem facts were presented for 15 of the 30 stem-fact pairs in each presentation order in the 1-stem condition. The purpose of comparing the 1-stem condition to the 2-stem condition was to determine which stimulus sets to use for Experiment 1B. Stimuli were included in Experiment 1B if participants performed at 15% or lower in the 1-stem condition and at least 40% correct in the 2-stem condition, and/or if participants performed twice as well in the 2-stem condition than in the 1-stem condition. These criteria were used because they are the general conventions when using this paradigm. A higher performance in the 2-stem condition compared to the 1-stem condition implies that integration of both stem facts is required to self-derive the integration fact.

Participants were tested in small groups in an in-person session (maximum of 4 participants per group). All participants were tested in both 1-stem and 2-stem conditions. For half of the stimuli, only one of the stem facts was presented (1-stem condition) and for half of the stimuli, participants were presented with both members of a stimulus pair (2-stem condition). The order was counterbalanced such that all stimuli were tested in both conditions equally often.

A total of 45 stem facts were presented during the learning phases for each presentation order. The order of facts was randomized across all versions in order to minimize primacy and recency effects. The stem facts (stimuli) in each version were split into two phases – Learning Phase I and Learning Phase II. Either 22 or 23 stem facts were presented during each learning phase (since 45 is an odd number and 22 or 23 is roughly half of the total stem facts). If 22 facts were presented during Learning Phase I, then 23 facts were presented during Learning Phase II and vice versa. The order of each session consisted of: Learning Phase I, Buffer I, Learning Phase II, Buffer II, Test Phase I, and Test Phase II. Each buffer task took approximately 7-10 minutes to complete.

There were 8 versions of the test. The order of presentation of the 24 sets of stimuli was pseudo-randomized, constrained by stem fact groups, through counterbalancing across versions.

Learning Phase I

During Learning Phase I, stem facts were shown. 4 of the 8 presentation versions included only “A” facts and 4 of the 8 presentation versions included only “B” facts (see Fig. 1). Either 22 or 23 facts were presented during Learning Phase I. For the “A” presentation versions, stem facts were counterbalanced between A1 and A2, meaning 2 of 4 versions presented A1 facts during Learning Phase I, and 2 of 4 versions presented A2 facts during Learning Phase I. For the “B” presentation versions, stem facts were counterbalanced between A/B1 and B2, meaning 2 of 4 versions presented A/B1 facts during Learning Phase I, and 2 of 4 versions presented B2 facts during Learning Phase I. The facts not shown from this set (A1 or A2, or A/B1 or B2) during Learning Phase I were shown during Learning Phase II. Each stem fact sentence appeared on the screen for 400 ms/word. At the end of Learning Phase I, a test cross was displayed on the screen. An unrelated buffer activity was then presented for approximately 10 minutes.

Learning Phase II

Learning Phase II proceeded as Learning Phase I but with different stimuli. 15 of the stem facts shown during Learning Phase II were the complementary stem-fact pair for the 15 stem facts in the 2-stem condition shown during Learning Phase I. The rest of the stem facts presented during this phase were the facts in the 1-stem condition that had not yet been seen. An unrelated buffer activity was then presented for approximately 6 minutes.

Test Phase

The test phase consisted of all of the integration questions for the “A” fact set or the “B” fact set being presented on a Powerpoint (corresponding to the fact sets presented during Learning Phase I and Learning Phase II). The order of the integration questions was randomized for each version. A sheet of paper was handed out to all participants with 30 blank spaces to write down their answers in an open-ended format. Each question was displayed on the screen for 15 seconds. At the end of the Powerpoint, a fixation cross was displayed on the screen and all sheets of paper were collected from participants. The same questions were then handed out to participants on a sheet in a forced-choice (multiple-choice) format. There were 3 possible answer choices for each question that were all encountered at some point during the protocol. The next part of the test phase consisted of all of the stem questions shown during Learning Phase I and Learning Phase II presented on a Powerpoint. The order of the stem questions was randomized for each version. A sheet was handed out to all participants with 45 blank spaces to write down their answers in an open-ended format. Each question was displayed on the screen for 15 seconds. Sheets were collected from all participants at the conclusion of the Powerpoint. Forced-choice questions were not asked for the stem questions in order to keep the study session to a reasonable length and to avoid participant fatigue.

Data Scoring and Analysis

The test phase and both buffer tasks were scored binarily. Participants received 1 point for a correct response and 0 points for an incorrect response. The maximum score that participants could receive for open-ended questions during the test phase was 75 because the first part of the test phase consisted of 30 open-ended integration questions and the final part of the test phase consisted of 45 open-ended stem questions. The maximum score that participants could receive for forced-choice questions during the test phase was 30 because the 30 integration questions were asked in a forced-choice format.

Data were analyzed using paired *t*-tests to compare 1-stem and 2-stem condition performance on the integration questions. This allowed for individual participant scores to be compared across the 1-stem and 2-stem conditions. Two separate analyses were run for open-ended answers and for forced-choice answers.

Results and Discussion

All analyses were conducted using Excel and the SPSS Statistics package (Version 26). As stated above, stimuli were included in Experiment 1B if participants performed at 15% or lower in the 1-stem condition and at least 40% correct in the 2-stem condition, and/or if participants performed twice as well in the 2-stem condition than in the 1-stem condition. Of the total stimuli that were tested, 18 stimulus sets met at least one of these criteria. These 18 stimuli were further analyzed to determine difference between 1-stem and 2-stem testing.

Because of the exclusion of fact sets that did not meet the criteria stated above, participants performed better on open-ended questions in the 2-stem data condition ($M = 0.34$, $SD = 0.47$) than the 1-stem data condition ($M = 0.17$, $SD = 0.38$). To determine whether this difference in performance between the 1-stem and 2-stem conditions in the open-ended questions was significant, a paired 2-sample t-test was performed. Performance in the 1-stem and 2-stem conditions were paired by participant. Participants performed significantly better in the 2-stem condition ($t(29) = 3.91$, $p = 0.00003 < 0.05$, $d = 0.40$). Similarly, participants performed better on forced-choice questions in the 2-stem data condition ($M = 0.75$, $SD = 0.43$) than the 1-stem data condition ($M = 0.62$, $SD = 0.49$). When paired by participant and compared using a 2-sample *t*-test, this difference was significant ($t(29) = 3.07$, $p = 0.005 > 0.05$, $d = 0.28$). These results show that the participants need to use memory integration in order to successfully answer the integration questions for the fact sets analyzed. As such, these 18 stimulus sets were validated for use in Experiment 1B. 6 stimulus sets that were tested in a similar method 1-stem and 2-stem in prior experiments were included in Experiment 1B as well.

Experiment 1B

Method

Participants

Participants in Experiment 1B were from the same source and thus represent the same populations as Experiment 1A. However, none of the participants participated in Experiment 1A. The final sample of participants for Experiment 1B included 66 (63.49% female) undergraduate students enrolled in introductory psychology classes at a mid-sized private university. Participants ranged in age from 18 to 23 years ($M_{age} = 19.19$ years, $SD = 0.96$). As for race, 28.57% identified as Asian, 11.90% identified as Black or African American, 56.35% identified as White or Caucasian, and 3.17% did not report their race. 7.94% were Hispanic or Latinx, 90.48% were not Hispanic or Latinx, and 1.59% did not report their ethnicity. Written informed consent was obtained for all participants. Protocols and procedures were approved by the university Institutional Review Board. All participants were compensated for their participation by receiving course credit for their introductory psychology class as in Experiment 1A.

Materials and Measures

Stimuli

In this study, we used 24 stimulus sets comprised of 4 novel, related facts ("stem facts") as in Experiment 1A. 18 of the stimuli used were tested 1-stem in this study (see Experiment 1A) and the other 6 of the stimuli used had been previously 1-stem tested in prior studies, making for 24 total stimulus sets. In this experiment, A/B1 was never explicitly presented to participants; rather, participants were asked an integration question in the "Self-Derivation" condition in order to self-derive the A/B1 fact.

All stimuli used exhibited lower performance in 1-stem than in 2-stem, suggesting that integration is necessary for high performance. The stimulus sets used in this study spanned the topics of science, geography, art, and history.

Buffers

Three buffer activities were used to separate different phases of the experiment. Two of the tasks were used in Experiment 1A. The paper folding task, which was adapted from Ekstrom et al. (1976), was the third buffer task.

Procedure

Experiment 1B consisted of an in-lab portion and an online portion. The second phase was conducted online for the ease of the participants. The order of each in-lab session consisted of: Learning Phase I, Buffer I, Learning Phase II, Buffer II, Test Phase I, and Learning Phase III. There were 6 versions of the test. The order of presentation of the 24 sets of stimuli was pseudo-randomized through counterbalancing across versions. Each buffer task took about 7-10 minutes to complete. The online portion of the study consisted of Test Phase II in the form of an online survey. Overall, the in-lab portion of the task took from 45 minutes to 60 minutes to complete and the online portion of the task took from 45 minutes to 60 minutes to complete.

Learning Phase I

During Learning Phase I, the first stem fact in the "A" stem-fact pair was shown on a Powerpoint presentation. Participants were instructed to pay close attention as the sentences were shown. Versions were counterbalanced such that each fact in a pair (A1 or A2) appeared equally often in Learning Phase I and Learning Phase II. Each stem fact sentence appeared on the screen for 400 ms/word. At the end of Learning Phase I, a test cross was displayed on the screen. See Experiment 1A for details on the first buffer task.

Learning Phase II

Learning Phase II proceeded as Learning Phase I. The second buffer between Learning Phase II and Test Phase I took a little over 6 minutes to complete. Sheets of paper were collected at the end of the task.

Test Phase I

Test Phase I was administered after Buffer II. To counterbalance the design, the stimuli sets were randomly divided into 3 equal groups of 8 stimulus sets – Group 1, Group 2, and Group 3. Groups were split equally between different conditions. Groups were labelled as "Self-Derivation," "Retrieval Practice," or "Control" for each of the 6 versions of the test. For the stimuli groups labeled "Self-Derivation," the "A" integration questions were presented. Integration questions were used in this condition in order to promote the cognitive process of self-derivation through integration. For the stimuli groups labeled "Retrieval Practice," 1 of the 2 stem questions for that stimulus set were presented. Stem questions were used in this condition in order to promote the cognitive process of recall. Only 1 question was presented in order to make sure only 1 of the stem facts was being recalled, and in order to limit the amount of self-derivation through integration being performed. For the stimuli groups labeled "Control," no questions were presented. Participants each received a sheet of paper with 16 numbered blank spaces. The 16 blank spaces corresponded to the 8 stem questions (retrieval practice condition) and the 8 “A” integration questions (self-derivation condition) that were presented during Test Phase I. Each question was displayed on the screen for 15 seconds, during which participants wrote down their answers to the open-ended question. See Experiment 1A for details on the third buffer task.

Learning Phase III

During Learning Phase III, all “B2” facts were shown on a Powerpoint presentation as were “A” facts in Learning Phase I and Learning Phase II.

Test Phase II

Test Phase II was counterbalanced to be given either 2 days or 7 days later between participants. Test Phase II delay was a between-subjects manipulation. 34 participants were assigned to the 2-day delay condition and 32 participants were assigned to the 7-day delay condition. This phase was administered through an online Qualtrics survey that was emailed out to all participants. Participants were instructed to complete the questionnaire within 12 hours of the email being sent out, and were incentivized to complete the survey within this timeframe with an additional course credit. Questions on the Qualtrics survey first included all “B” integration questions during the first block (see Fig. 2) in order to measure knowledge extension as well as all “A” and “B” stem questions during the second block (see Fig. 2) in order to measure long-term retention. The questions were randomized within each block through the Qualtrics system. Participants were required to answer all of the questions in order to complete the survey, and all questions were presented in an open-ended format.

Data Scoring and Analysis

Test Phase I was scored binarily as in Experiment 1A. Participants received 1 point for a correct response and 0 points for an incorrect response. The maximum score that participants could receive during test phase I was 16 points because there were 24 stimulus sets presented and there were 8 questions in the control condition, for which no questions were presented (24 – 8 = 16). 8 of the 16 points were from stem questions for the "A" fact pair (retrieval practice condition), and the other 8 of the 16 points were from integration questions for the "A" fact pair (self-derivation condition). Similarly, participants received 1 point for each correct response and 0 points for each incorrect response on Test Phase II (the Qualtrics survey). Questionable answers were presented to multiple researchers in order to determine their accuracy. The maximum score that participants could receive for test phase II was 96 points because there were 3 stem questions presented for each of the 24 stimulus sets (3*24 = 72) and 1 "B" integration question presented for each of the 24 stimulus sets, making for a total of 96 (72+24 = 96) points.

The main results of the study were analyzed using a 2 x 3 MANOVA, with delay (2 days vs. 7 days) and test condition (Self-Derivation vs. Retrieval Practice vs. Control) as the independent variables, and accuracy of response for "B" delayed integration questions and accuracy of response for "A" and "B" delayed stem questions as the dependent variables. Differences in performance between the retrieval practice and self-derivation conditions during Test Phase I were analyzed, as well as bivariate correlations between performance in Test Phase I and Test Phase II.

Data Reduction

Before conducting data analysis, a number of exclusion factors had to be accounted for. Participants who failed to answer at least 1 question correctly in the retrieval practice condition during Test Phase I were excluded, since they have shown no evidence of encoding any information during the session (*n* = 4). Participants who had taken part in previous related experiments were also excluded (*n* = 7). Some participants were excluded from analysis due to experimental error (*n* = 5). Participants who did not answer the delayed integration and stem questions were excluded from delayed question analysis, but were still included for analysis involving only Test Phase I (*n* = 11). The final sample size consisted of *n* = 66 participants for Test Phase I analysis, and *n* = 55 participants for all other analyses.

One stimulus set was excluded from analysis because it was found to not require memory integration. There also were some questions that were answered correctly but did not use retrieval of information for stem questions or self-derivation through integration for integration questions to answer. Answers fell into this category when they were true, but not derived from the memory integration task. Prior knowledge was the most likely source of these responses. Any answers that were judged to fall into this category by at least 2 researchers were excluded from analysis. A total of 81 number of responses across Test Phase I (25) and Test Phase II (56) were excluded from analyses.

Results

Initial Test Performance: Test Phase I

All analyses were conducted using Excel and the SPSS Statistics package (Version 26). Descriptive statistics for performance on the retrieval practice and self-derivation conditions during Test Phase I are shown in Figure 4. As predicted, initial performance was greater in the retrieval practice condition ($M = 0.57$, $SD = 0.50$) than in the self-derivation condition ($M = 0.49$, $SD = 0.50$). Although general performance hovered around 50% for both conditions, performance was highly variable across conditions as shown by the standard deviation bars in Figure 4. To determine whether the difference in performance between conditions was significant, an independent samples *t*-test was conducted. Participants performed significantly better on retrieval practice questions ($M = 0.57$, $SD = 0.496$) than integration questions ($M = 0.49$, $SD = 0.500$) during Test Phase I ($t(822) = 2.271$, $p = 0.023 < 0.05$, $d = 0.161$). This was a small effect since $d < 0.2$, but indicates that on average, students performed better on surface processing questions than deep processing questions with a meaningful difference of 12%. Performance during Test Phase I did not differ between participants in the 2-day delay period ($M = 0.52$, $SD = 0.50$) and participants in the 7-day delay period ($M = 0.52$, $SD = 0.50$) ($t(66) = 0.11$, $p > 0.05$, $d = 0.01$). This means that participants placed in the 2-day delay period performed comparably to participants placed in the 7-day delay period, and thus can be treated as equivalent populations for analysis during the next test phase.

Delayed Test Performance: Test Phase II

Performance on integration questions during Test Phase II (by condition and delay) is presented in Figure 5. As hypothesized, when collapsed across delay period, performance on integration questions during Test Phase II was highest in the self-derivation condition ($M = 0.23$, $SD = 0.43$), followed by the retrieval practice condition ($M = 0.19$, $SD = 0.39$) and the control condition ($M = 0.17$, $SD = 0.38$). Within each condition, performance in the 2-day delay period was higher than performance in the 7-day delay period.

Performance on stem-fact questions during Test Phase II (by condition and delay) is presented in Figure 6. When collapsed across delay period, performance on stem-fact questions during Test Phase II was highest in the self-derivation condition ($M = 0.46$, $SD = 0.50$), followed by the retrieval practice condition ($M = 0.42$, $SD = 0.49$) and the control condition ($M = 0.36$, $SD = 0.48$). Within each condition, performance in the 2-day delay period was higher than performance in the 7-day delay period.

A 2 (delay: 2 or 7 days) x 3 (condition: control, retrieval practice, self-derivation) MANOVA was first used to analyze the data with the two dependent factors of delayed integration question performance and delayed stem-fact question performance. An initial MANOVA was used instead of separate ANOVAs for stem-fact questions and integration questions because the interaction between delayed stem-fact question performance and delayed integration question performance was a factor of interest. There was a main effect of condition, $F(4, 316) = 2.603$, $p = 0.036 < 0.05$; Wilk's $\Lambda = 0.937$, partial $\eta^2 = 0.032$. This means that across delayed integration questions and delayed stem-fact questions, there was a significant difference in performance as a function of condition (control, retrieval practice, or self-derivation). Similarly, there was a main effect of delay that approached significance, $F(2, 158) = 2.821$, $p = 0.063 > 0.05$; Wilk's $\Lambda = 0.966$, partial $\eta^2 = 0.034$. As expected, participant performance was nominally higher in the 2-day delay than in the 7-day delay. However, there was clearly no interaction between condition and delay ($F(4, 316) = 0.298$, p = 0.880; Wilk's $\Lambda = 0.993$, partial $\eta^2 = 0.004$). This result means that the effect of condition did not depend on the delay period and the effect of delay period did not depend on the condition.

A follow-up two-way 2 (delay: 2 or 7 days) x 3 (condition: control, retrieval practice, self-derivation) ANOVA was conducted to analyze the data with the dependent variable of delayed integration question performance. When examining delayed integration question performance as a function of condition, univariate analysis for condition did not yield statistically significant findings ($F(2, 159) = 1.482$, $p = 0.230$). Thus, there was no statistically significant effect of condition for dependent variable of delayed integration test performance. Likewise, when examining delayed integration question performance as a function of delay, univariate analysis did not yield statistically significant findings ($F(1, 159) = 3.001$, $p = 0.085$).

Thus, there was no statistically significant effect of delay for dependent variable of delayed integration test performance. Additionally, there was no statistically significant interaction between condition and delay on delayed integration question performance ($F(2, 318) = 0.125$, $p = 0.882$).

A follow-up two-way 2 (delay: 2 or 7 days) x 3 (condition: control, retrieval practice, self-derivation) ANOVA was conducted to analyze the data with the dependent variable of stem-fact question performance. Univariate analysis for condition yielded a statistically significant effect of condition on delayed stem-fact memory ($F(2, 159) = 5.292$; $p = 0.006 < 0.05$) such that the self-derivation condition had the highest performance on delayed stem-fact questions ($M = 0.46$, $SD = 0.48$), the retrieval practice condition had the next highest performance on delayed stem-fact questions ($M = 0.42$, $SD = 0.49$), and the control condition had the lowest performance on delayed stem-fact questions ($M = 0.36$, $SD = 0.50$). A post hoc Tukey test showed that the control condition and self-derivation conditions differed significantly on the dependent measure of delayed stem-fact performance ($p = 0.005 < 0.05$). There were no other significant differences between means. This result was unexpected because it was inconsistent with a prior study that demonstrated that higher-order processing questions and a restudy condition yielded equally low performance on delayed fact questions (Agarwal, 2019). Thus, this result means that self-derivation is better than a no-test condition at promoting the recall of information. Univariate analysis for delay yielded a statistically significant effect of delay on delayed stem-fact memory ($F(1, 159) = 5.296$; $p = 0.023 < 0.05$) such that participants in the 2-day delay had the higher performance on delayed stem-fact questions ($M = 0.45$, $SD = 0.50$) and participants in the 7-day delay had the lower performance on delayed stem-fact questions ($M = 0.38$, $SD = 0.49$).

However, there was no statistically significant effect of interaction between condition and delay on delayed stem-fact memory ($F(2, 318) = 0.213$, $p = 0.808$).

Relations between Initial and Delayed Test Performance

To test the correlation in individual performance in both test phases, bivariate correlations between Test Phase I performance and delayed stem-fact question performance separated by condition are represented using a scatter plot in Figure 7, and bivariate correlations between Test Phase I performance and delayed integration question performance separated by condition are represented using a scatter plot in Figure 8. Overall, these bivariate correlations provide evidence that those who performed well during Test Phase I also performed well during Test Phase II. Specifically, there was a significant positive correlation between performance on Test Phase I and performance on delayed stem-fact questions ($r = 0.636$, $p = 0.000 < 0.05$). This means that those who processed information in both shallow or deep conditions were more likely to retain this information, whereas those who failed to process information were more likely to forget this information. When separated by condition (retrieval practice vs. self-derivation) during Test Phase I, there was a significant positive correlation between retrieval practice performance on Test Phase I and delayed stem-fact performance ($r = 0.597$, $p < 0.01$) and a significant correlation between self-derivation performance on Test Phase I and delayed stem-fact performance ($r = 0.736$, $p < 0.01$). These results mean that participants who performed well during Test Phase I (on both retrieval practice and self-derivation conditions) also tended to perform well during delayed stem-fact questions across both 2-day and 7-day delay periods. These significant positive correlations imply that in order to perform well on later tests of recall and integration, it is important to remember and perform well on questions asked during the first test phase.

There was also a significant positive correlation between performance on Test Phase I and performance on delayed integration questions ($r = 0.239$, $p = 0.012 < 0.005$). When separated by condition during Test Phase I, there was a significant correlation between self-derivation performance on Test Phase I and delayed integration performance ($r = 0.340$, $p < 0.05$) but no significant correlation between retrieval practice performance on Test Phase I and delayed integration performance ($r = 0.176$, $p = 0.197 > 0.05$). These results mean that those who performed well during Test Phase I on the self-derivation condition also tended to perform well during delayed integration questions across both 2-day and 7-day delay periods. However, those who performed well during Test Phase I on the retrieval practice condition did not also tend to perform well during delayed integration questions across both delayed periods. Thus, these data provide evidence that retrieval practice does not improve individual performance on integration questions.

Discussion

The results from Experiment 1B show that participants perform significantly better on retrieval practice questions than self-derivation questions during Test Phase I. This data provides evidence that the process of self-derivation through integration is more difficult than the process of retrieval practice. Another finding from Experiment 1B was that there were significant correlations between performance on Test Phase I and performance on Test Phase II integration questions, as well as performance on Test Phase I and performance on Test Phase II stem-fact questions. This finding supports the claim that individual performance stayed consistent over either a 2-day or 7-day delay period. Those who remembered information during the learning phase had a higher likelihood of retaining information and self-deriving new information in the future. Additionally, the data show that placement of stimuli in the self-retrieval condition compared to placement of stimuli in a no question (control) condition significantly enhances future performance on stem-fact questions. This finding provides initial evidence that self-derivation through integration has the potential to enhance memory.

General Discussion

The researchers of the current study aimed to investigate the effects that different levels of processing –retrieval practice and self-derivation through memory integration – have on future test performance on both stem-fact questions and integration questions. This question was important to investigate because although retrieval practice has been shown to enhance memory (Agarwal et al., 2017; Agarwal, 2019; Tran, Rohrer, Pashler, 2015; Whiffen & Karpicke, 2017), the role of different levels of processing in enhancing future retention and learning opportunities has yet to be determined.

Experiment 1A identified stimulus sets that required the use of memory integration for use in Experiment 1B, and Experiment 1B examined the effects of different levels of processing on future test performance. There were four major findings from the study. First, we found that college students answered more retrieval practice questions correctly than self-derivation questions during Test Phase I. Second, we found that those who performed well during Test Phase I also tended to perform well on both delayed integration questions and delayed stem-fact questions. Third, we found that in the control condition, performance on stem-fact questions declined over the longer delay period, but performance on integration questions remained stable over a longer delay period. Fourth, we found that self-derivation through integration was the most effective at promoting stem-fact recall compared to retrieval practice and control conditions. In other words, these findings support our initial hypothesis that self-derivation through integration, a deeper level of processing, promotes the retention of information more than retrieval practice.

The first major finding from the study was that college students answered significantly more retrieval practice questions correctly than self-derivation questions during Test Phase I. This finding has broad applications for the fields of psychology and education. Since fewer self-derivation questions were answered correctly than retrieval practice questions, this implies that self-derivation questions are harder to answer and to answer them correctly, participants may need a deeper understanding of information. Thus, the finding is consistent with the conjecture that self-derivation requires a deeper level of processing than retrieval practice (Bauer & Varga, 2017; Varga, Esposito, & Bauer, 2019). For educators, this means that using self-derivation questions on examinations may be more effective at assessing students for deeper understanding of course material.

Another major finding from the study was that those who performed well during Test Phase I also tended to perform well on delayed integration questions and delayed stem-fact questions. This means that individual performance tended to stay consistent throughout the study period. Participants who showed evidence that they encoded information during the learning phases as measured by performance in the retrieval practice condition during Test Phase I also tended to have a high test performance on Test Phase II, and participants who failed to encode as much information during learning phases tended to have a low test performance on Test Phase II. Although this positive correlation was significant between the self-derivation condition and delayed integration questions, this positive correlation was not significant between the retrieval practice condition and delayed integration questions. The lack of significance in this instance may be due to measurement invariance. The measures for successful self-derivation through integration in Test Phase I were similar to the measures for successful self-derivation through integration for Test Phase II, whereas the measures for successful retrieval in Test Phase I were more dissimilar to the measures for successful self-derivation through integration for Test Phase II. In addition, the quantity of integration questions asked during Test Phase II was only a third of the quantity of stem-fact questions asked during Test Phase II. Thus, less data was available to analyze for integration questions during Test Phase II, representing more variance and a larger standard error.

Although there were no significant differences between performance in the 2-day and 7-day delay periods by condition, there was a trend for performance on both integrations questions and stem-fact questions to decline over the longer delay period. However, this trend was not evident for performance on the integration questions in the control condition. Even though performance declined between the 2-day and 7-day delays for stem-fact questions in the control condition, performance on integration questions for the control condition in both the 2-day and 7-day delay periods was relatively similar. This finding indicates that performance on integration questions tends to be more consistent and persist over longer periods of time than performance on stem-fact questions. Performance on integration questions may be a better indicator of long-term memory than performance on stem-fact questions.

The fourth major finding was that self-derivation through integration was the most effective at promoting stem-fact recall compared to the control condition, followed closely by the retrieval practice condition and differing greatly from the control condition. In other words, self-derivation is significantly better than a control in improving memory retention. This result is inconsistent with findings from Agarwal (2019) that retrieval practice with higher order questions – which is comparable to self-derivation through integration – did not improve performance on delayed fact tests.

One explanation for the difference in these results is that this study examined the process of self-derivation through integration, whereas Agarwal (2019) studied a number of higher order processes. These higher order processes included analysis, synthesis, and evaluation, as per Bloom et al. (1956). It may be the case that certain higher order processes, like self-derivation through integration, have a higher propensity to promote the retention of facts whereas other higher order processes, like analysis and evaluation, do not support this retention of facts. All higher order processes may not be equally effective at promoting the retention of facts. Since Agarwal (2019) used multiple forms of higher order processes but did not include self-derivation through integration, they may have not found a significant association between higher order processes and performance on fact questions.

Additionally, Agarwal (2019) used a multiple-choice format for all questioning, whereas in the current study we used an open-ended format followed by a forced-choice format in Experiment 1A. The researchers of the current study used an open-ended format since it is important to test and understand performance on open-ended tasks as much of knowledge extension that is done outside the laboratory are open-ended, and there are no options to choose from. Therefore, this study may be more consistent with conditions outside of the laboratory (e.g., in educational settings). However, the current study was consistent with Agarwal's (2019) findings in providing evidence against Bloom's Taxonomy.

Bloom's Taxonomy is one of the first and prevailing theories about the process of learning and it consists of six hierarchical categories: knowledge, comprehension, application, analysis, synthesis, and evaluation. On the lower end, knowledge entails remembering or memorizing information without necessarily understanding its meaning, and on the higher end, evaluation involves being able to predict outcomes based on information. Bloom claimed that lower-order processing (like knowledge and comprehension) must occur before being able to reach a higher-order processing (like analysis, synthesis, and evaluation). This theory contributes to one of the prevailing attitudes in the educational field that strong foundational knowledge is necessary for higher-order learning. Because Agarwal (2019) found that fact questions involving lower-order learning fared worse than higher-order questions in promoting higher-order processes, she found evidence that undermines Bloom et al. (1956). Similarly, the current study also found that retrieval practice fares worse than integration questions in promoting self-derivation through integration (although this difference was not statistically significant).

The findings from Agarwal (2019) were explained using Barnett and Ceci's (2002) taxonomy that distinguishes between "near" and "far" transfer of knowledge and supports a transfer-appropriate processing framework, but the data from the current study support a desirable difficulties framework originally proposed by Bjork (1994). Bjork claims that more difficult and complex mental processing of information tends to increase both the future retention and application of knowledge (Bjork, 1994). This means that the more that a task related to the retrieval of information contains "desirable difficulty," the more the task promotes the remembering of this information. Bjork identified factors such as varying the conditions of practice, providing contextual interference, giving distributed practice sessions, reducing feedback to the learner, and using tests as learning events as benefitting the learner in terms of the long-term retention of information (Bjork, 1994). Some more concrete examples of these "desirable difficulties" include randomizing the time between trials, having more spaced trials, and employing retrieval practice. As self-derivation through memory integration represents a deeper level of processing than retrieval practice, it requires the use of more complex mental processing and thus may represent a "desirable difficulty" that increases knowledge recall and application skills.

The data from the current study adds to the retrieval practice literature by displaying the utility of higher order processes (such as self-derivation through integration) in not only future performance on inference questions as demonstrated by Tran et al. (2015), but also on future fact recall performance. These results imply that self-derivation through integration can promote performance on a variety of information testing methods in addition to further integration of information. The researchers of the current study also propose an idea that performance by individuals on integration questions may be a better measure of long-term retention than performance on fact questions. This idea is proposed because performance on fact questions was not a significant predictor of long-term retention, whereas performance on integration questions was a significant predictor of long term retention. However, this conclusion has yet to be supported by more statistically significant data.

Limitations

Although the current study contributes important finding to the literature on both retrieval practice and self-derivation, it is also important to note some limitations of this study. First, the second test phase was sent out as an online survey, not administered to participants in a laboratory setting. Although this format was chosen for the ease of participants, it may have introduced a number of confounding variables that would not have been present in a more controlled setting. For example, participants had access to the internet and so they could have looked up answers. However, the participants were not being graded or given any advantage for accuracy, so there was little motivation for them to use outside resources. Nonetheless, for future research on this topic, it may provide more compelling evidence to support the hypothesis that self-derivation through integration promotes learning more than retrieval practice if participants were tested in a controlled environment for both test phases.

Sample bias is another limitation of this study. The participants involved in this study were all undergraduate college students enrolled in an introductory psychology course. Thus, the results of this study are not generalizable to other student bodies, those of the same age who do not attend college, etc. In order to generalize the study to different populations, the results of this study must be tested for replication using a number of populations. Replicating the findings from this study in other populations would provide more compelling evidence for underlying cognitive processes implicated within the study.

Future Directions

In addition to reducing some of the limitations discussed in the previous section, future research can focus on examining the effects of retrieval practice and self-derivation through integration on future test performance in individuals from a larger age range. The current study only involved college students as participants, and it would be useful to see how the process of self-derivation through integration develops throughout childhood and whether it is maintained for older adults. This information could shed light on the typical ages for developing cognitive processes that allow for self-derivation through integration. For example, if the link between self-derivation through integration and long-term memory is not found in younger children and older adults, then the findings would provide evidence that the process of self-derivation through integration varies across the lifespan.

A question that is raised by the results of the current study is how self-derivation through integration facilitates the retention of information. In order to study the mechanisms through which self-derivation acts on memory, it may be valuable to conduct fMRI or PET scans to see if specific brain areas are related to the process of self-derivation through integration. Investigating whether there are any neurobiological patterns during self-derivation through integration may reveal more about the brain structures involved with long-term learning. By identifying these structures, researchers may have the potential to target these regions for drug development and therapies aimed at enhancing memory and learning.

Conclusion

This study provides evidence that self-derivation through integration is a useful tool for enhancing memory retention as well as possibly enhancing further integration of learned information. Additionally, it may be a good predictor for long-term retention. However, these conclusions must be interpreted carefully as this study was conducted using only college students at a mid-sized private university. Additionally, there were other limitations of this study (see limitations section). Despite these limitations, self-derivation through integration shows promise in enhancing learning. This process may even help students perform better on exams and apply classroom knowledge in the real world. As such, it would be useful for educators to employ more novel teaching methods that utilize self-derivation through integration in order to enhance learning in the classroom. It would also benefit students to practice for final exams using self-derivation through integration rather than retrieval practice such as repeated recall tests. Although there are limitations to this study and further methods of testing (other than retrieval practice and self-derivation through integration) have yet to be explored, evidence suggests self-derivation through integration is an efficient method of learning and retaining new information.

References

Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology, 111*(2), 189-209. DOI: 10.1037/edu0000282

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger III, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory, 25*(6), 764-771. DOI: 10.1080/09658211.2016.1220579

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., . . . Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (abridged ed.). New York, NY: Addison Wesley Longman.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128,* 612– 637. DOI: 10.1037/0033-2909.128.4.612

Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of Experimental Psychology, 107,* 452–465. DOI: 10.1016/j.jecp.2010.05.012

Bauer, P. J., & Varga, N. L. (2017). Similarity and deviation in event segmentation and memory integration: Commentary on Richmond, Gold, & Zacks. *Journal of Applied Research in Memory and Cognition, 6,* 124-128. DOI: 10.1016/j.jarmac.2017.01.006

Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *The taxonomy of educational objectives: The classification of educational goals* (Handbook 1: Cognitive domain). New York, NY: David McKay Company.

Brunyé, T. T., Smith, A. M., Hendel, D., Gardony, A. L., Martis, S. B., & Taylor, H. A. (2019). Retrieval Practice Enhances Near but Not Far Transfer of Spatial Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. DOI: 10.1037/xlm0000710

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*. *23*(6), 760–771. DOI: 10.1002/acp.1507

Coane, J. H., (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition*. *2*(2), 95–100. DOI: 10.1016/j.jarmac.2013.04.001

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

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal behaviour, 11*, 671-684.

Hanawalt, N.G. (1937). Memory trace for figures in recall and recognition. *Arch. Psychol*. 31 (216), 5–89.

Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology,* 7(350). DOI: 10.3389/fpsyg.2016.00350

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. DOI: 10.1111/j.1467-9280.2006.01693.x

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

Varga, N. L., Bauer, P. J. (2017). Young adults self-derive and retain new factual knowledge through memory integration. *Memory & Cognition, 45*, 1014–1027. DOI: 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. DOI: 10.1037/xge0000581

Varga, N. L., Stewart, R. A., & Bauer, P. J. (2016). Effects of delay and reminders on the accessibility of self-generated knowledge in 4-year-old children. *Journal of Experimental Child Psychology, 145,* 48–63. DOI: 10.1016/j.jecp.2015.11.015

Whiffen, J., & Karpicke, J. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1036–1046. DOI: 10.1037/xlm0000379

A1: A disease spread by fleas is typhus.
A2: Typhus is caused by the Rickettsia bacteria.
A/B1: Fleas transmit the Rickettsia bacteria.
B2: The organisms that transmit the Rickettsia bacteria cannot hear.
B: Fleas cannot hear.

Figure 1. Example of Stimulus Set used in Experiment 1A. A1 and A2 are “A” stem facts, A/B1 and B2 are “B” stem facts. Fact “A” can be self-derived through the integration of A1 and A2 stem facts, and fact “B” can be self-derived through the integration of A/B1 and B2 stem facts.

A1: Pumice is a type of volcanic rock.
A1 SFQ: What type of rock is pumice?
Answer: Volcanic rock
A2: Volcanic rock floats in water.
A2 SFQ: What kind of rock floats in water?
Answer: Volcanic rock
A IQ/B1 SFQ: What does pumice do when you put it in water?
A/B1: Pumice floats in water.
B2: Rocks that float in water have air bubbles inside them.
B2 SFQ: What do rocks that float in water have inside them?
Answer: Air bubbles
B IQ: What is inside pumice?
B: Air bubbles are inside pumice.

Figure 2. Example of Stimulus Set used in Experiment 1B. A1 and A2 are “A” stem facts, A/B1 and B2 are “B” stem facts. Fact “A” can be self-derived through the integration of A1 and A2 stem facts, and fact “B” can be self-derived through the integration of A/B1 and B2 stem facts. (IQ = Integration Question, SFQ = Stem-Fact Question)

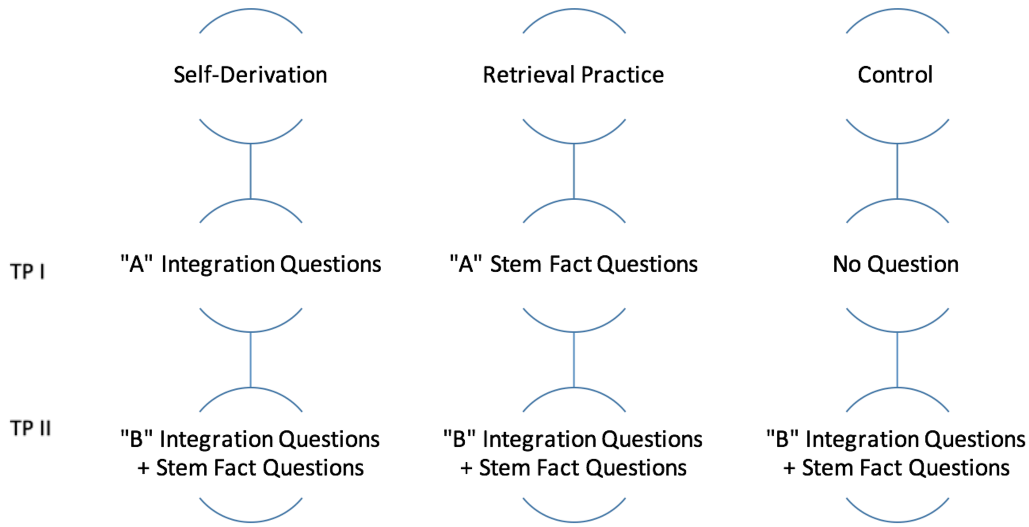

Figure 3. Illustration of Questions Presented during Test Phases for Experiment 1B. 8 stimulus sets in the "Self-Derivation" condition were presented with the "A" integration question during test phase I, 8 stimulus sets in the "Retrieval Practice" condition were presented with a stem question during test phase I, and 8 stimulus sets in the "Control" condition were presented with no question during test phase I. A total of 16 questions were presented during test phase I. (TP = Test Phase)

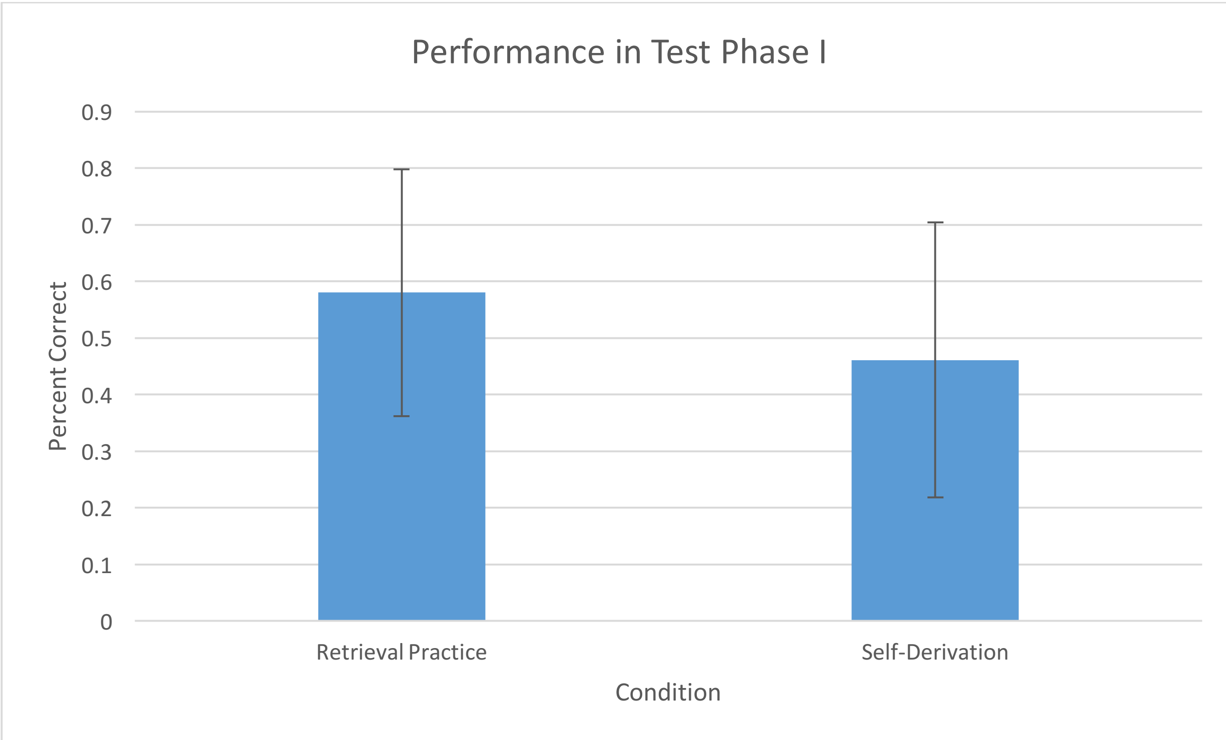

Figure 4. Performance in Test Phase I. Error bars represent standard deviation.

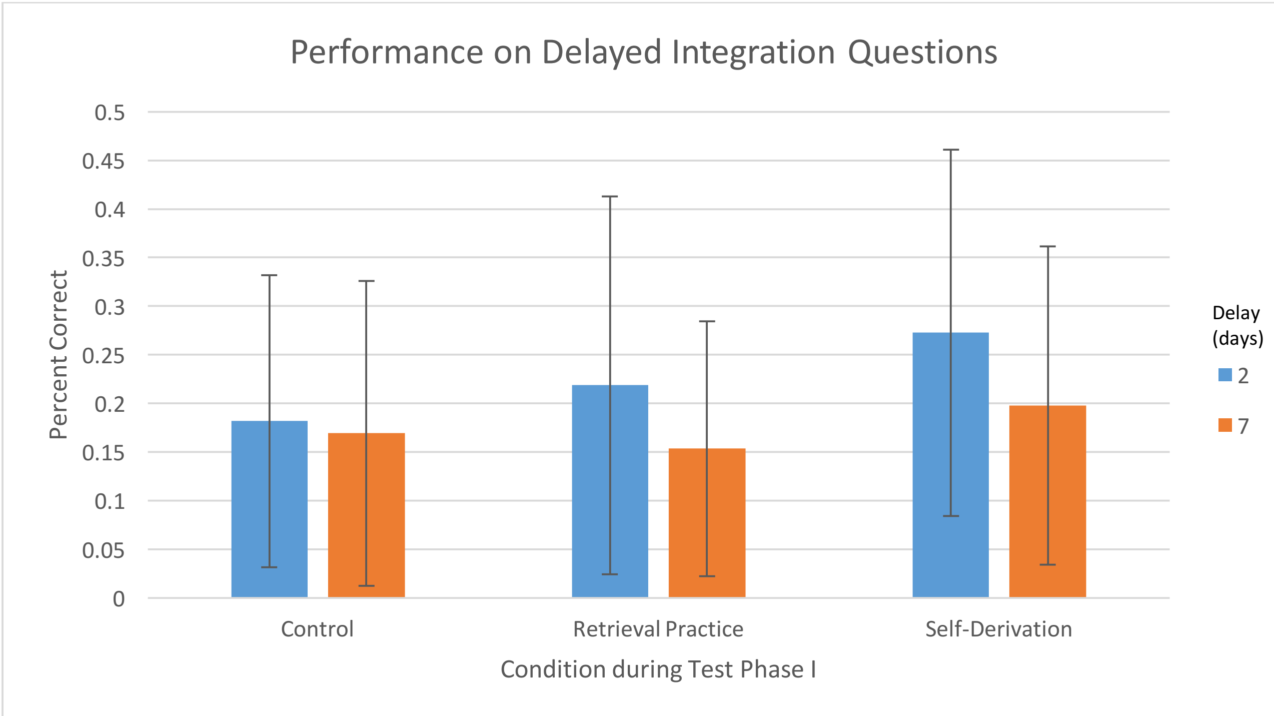

Figure 5. Performance on delayed integration questions separated by condition during Test Phase I and delay period.

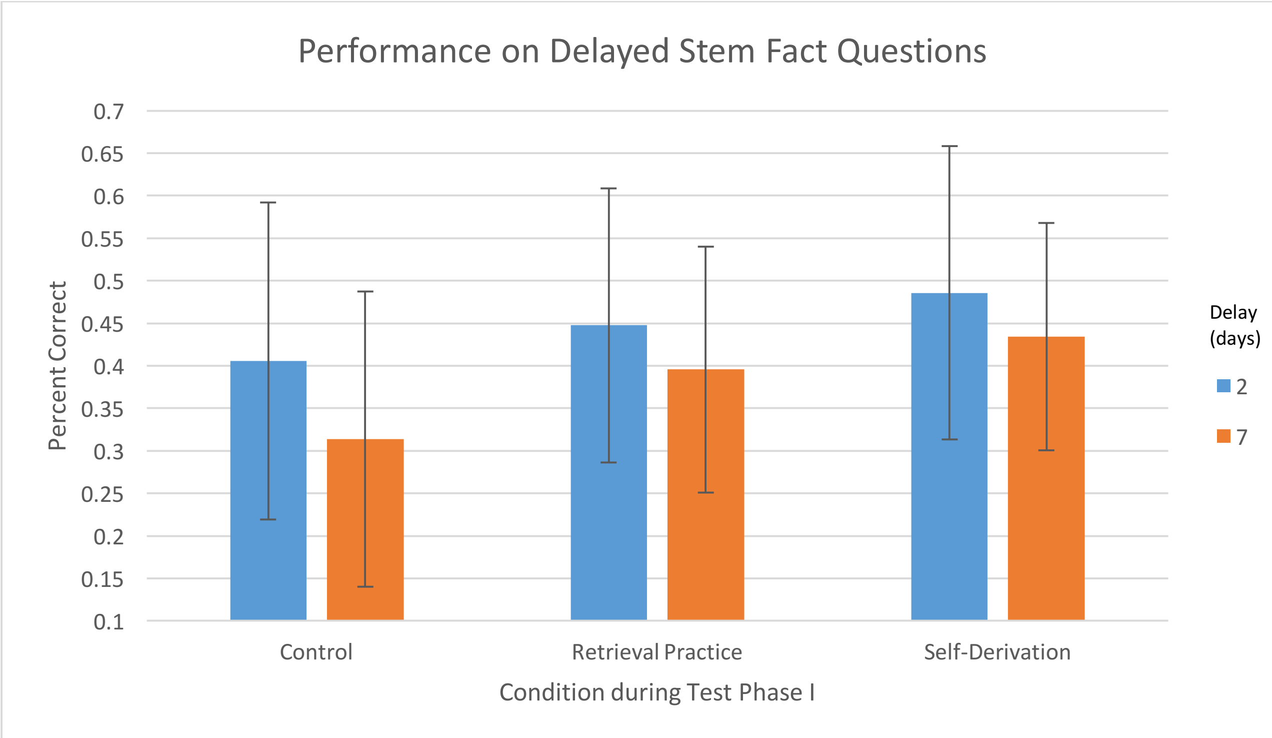

Figure 6. Performance on delayed stem questions separated by condition during Test Phase I and delay period.

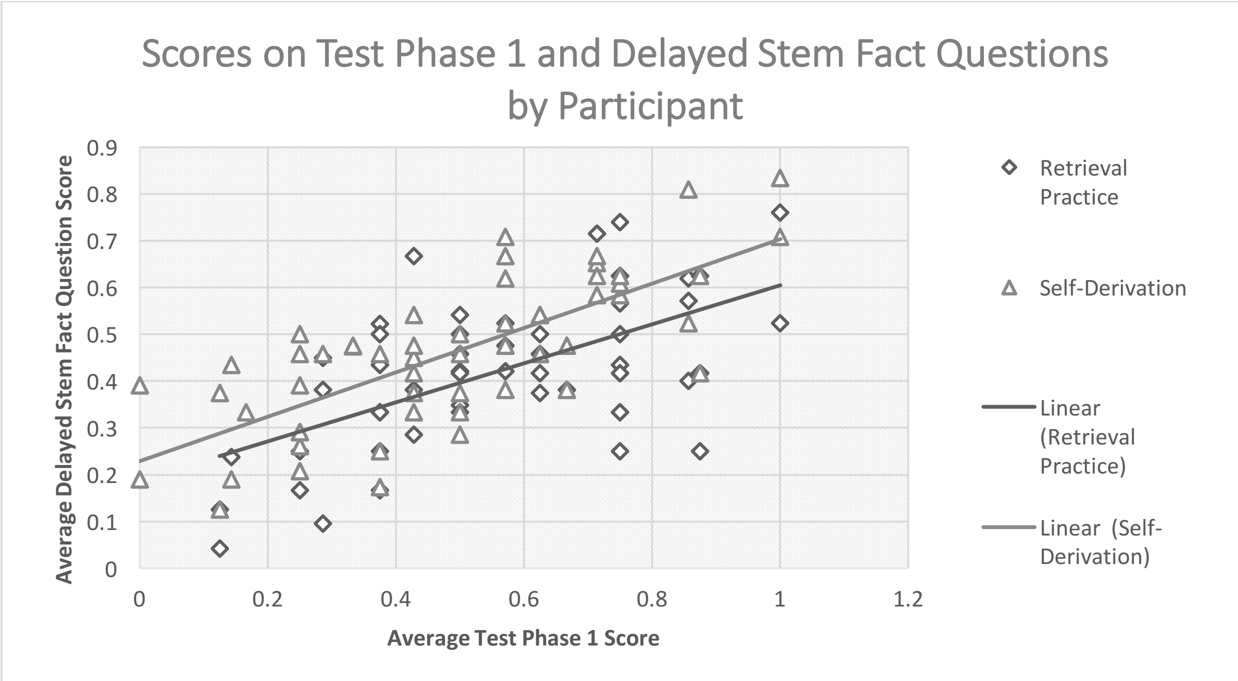

Figure 7. Bivariate correlations between Test Phase I scores and Delayed Stem Fact Scores.

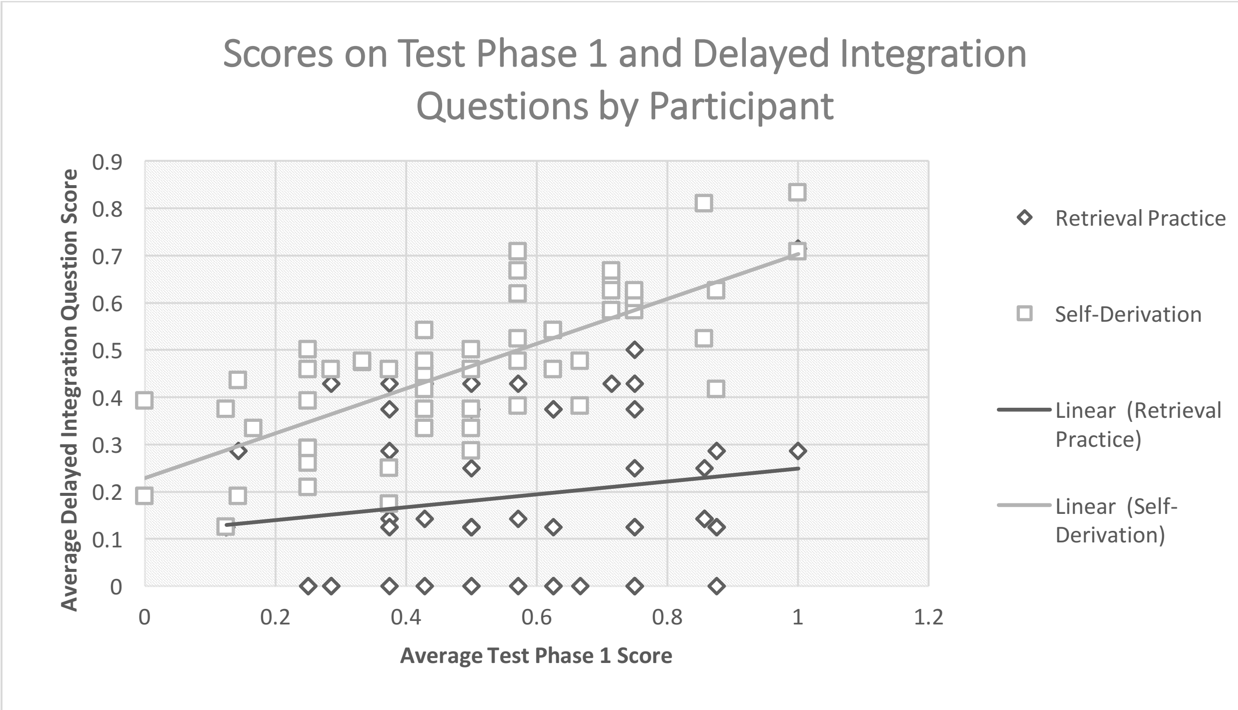

Figure 8. Bivariate correlations between Test Phase I scores and Delayed Integration Scores.