THE IMPACT OF SELF-REGULAION AND RETRIEVAL PRACTICE ON CATEGORY LEARNING

by

Kennadi Cook

Submitted in partial fulfillment of the
requirements for Departmental Honors in
the Department of Neuroscience
Texas Christian University
Fort Worth, Texas

May 8, 2023

THE IMPACT OF SELF-REGULAION AND RETRIEVAL PRACTICE ON CATEGORY LEARNING

Project Approved:

Supervising Professor: Uma Tauber, Ph. D.

Department of Psychology

Tracy Centanni, Ph. D.

Department of Psychology

Mikaela Stewart, Ph. D.

Department of Biology

ABSTRACT

Retrieval practice typically improves learning and memory performance of basic information (Rowland, 2014). Much less research has evaluated the degree to which retrieval practice results in better test performance of more complex information such as category learning. In one case, retrieval practice led to superior classification performance relative to restudy conditions (Jacoby et al., 2013); however, in another, it did not (Babineau et al., 2022). One important component that may contribute to retrieval practice effects in category learning is whether the learning process is self-regulated. We systematically explored this issue with the goal to establish when retrieval practice benefits learning of complex categorical information. In this experiment, students learned organic chemistry compounds by completing study or test trials. During learning, study order was controlled by the researcher (experimenter-controlled) or participant (self-regulated). As predicted, I found that classification performance was best when students completed test trials. Further, this testing benefit was unaffected by study order (researcher-controlled vs. self-regulated), and most students reported the belief that testing would be beneficial for their learning. The data from this experiment help establish effective ways for students to learn difficult material.

TABLE OF CONTENTS

INTRODUCTION

Retrieval practice is considered one of the most effective ways to improve performance on memory tasks. Specifically, retrieval practice often results in better test performance than does restudying the same material (Carpenter et al., 2022; Roediger & Karpicke, 2006). Retrieval practice is a strategy in which a learner tests themselves over the material rather than restudying it. When taking an organic chemistry class, students may complete practice tests or use flash cards to practice retrieving information from memory and reinforce their knowledge of the material, or they may review their notes or reread the book to restudy the material. The benefit of retrieval practice over restudying has also been found for natural category learning in one study (Jacoby et al., 2013); however, not in another study (Babineau et al., 2022).

Jacoby and colleagues evaluated the retrieval practice effect on natural categories (bird species) over three experiments. Participants learned to classify eight categories of birds in an initial training phase. Participants were shown an image of the bird with either the correct name or a test prompt with several options of bird names presented underneath. The training phase consisted of 45 exemplars shown four times each for a total of 160 trials. Next, participants entered the study phase. Students in the repeated study groups were shown an image of the bird along with the correct name underneath which was read aloud. Students in the testing groups were shown an image of the bird along with question marks below the image. Participants were prompted to select the name they thought matched the bird displayed. After their selection was made, feedback was given on if their response was correct or incorrect. Students then entered the final testing phase where they were shown 10 exemplars (5 novel, 5 studied) from each of the eight bird families they had learned to classify. As each exemplar was shown, participants were prompted to select the name to which it belonged. The researchers found that retrieval practice resulted in better classification performance than restudying on an immediate test and on a delayed test.

Babineau and colleagues (Babineau et al., 2022) evaluated the retrieval practice effect on natural categories (rock categories) as well. In this experiment, participants learned to classify three rock categories, each consisting of four subcategories. After a prior knowledge test was completed, students entered a self-regulated learning phase. For students in the study group, the exemplar was shown along with the correct classification name underneath. Students in the testing group were shown the exemplar along with the three category names. Participants were prompted to select the category they thought the exemplar belonged to. After their selection was made, feedback was given on if their response was correct or incorrect. After each trial was completed, students in both groups were given the option of what type of rock category they wanted to study next, either a rock from the same category they just studied or a rock from a different category. The study phase continued until participants had classified at least 40 exemplars. They then entered a final testing phase. Students were given both a novel and studied exemplar tests consisting of 36 trials each. As each exemplar was shown, participants were prompted to select the name to which it belonged. Performance on the classification tests indicated that the study-only group and the retrieval practice group did not significantly differ. Thus, the results were inconsistent with the outcomes from Jacoby et al. (2013).

Why do the outcomes from prior work differ with respect to the impact of retrieval practice on category learning? One important difference to address between these studies is the introduction of self-regulation during learning. We believed that self-regulating one’s learning verses having learning controlled by an experimenter might be important for the retrieval practice effect. When students are able to self-regulate their learning by choosing the order of exemplars during study, some of their choices may reduce the need to retrieve information from memory. For example, if a student completes a practice test trial for a specific category and they answer correctly, they receive feedback telling them that their selection was correct. If they decide to study another exemplar from the same category, they do not need to engage as fully in the retrieval process. They already know what the correct answer will be because they got that information during the last trial and decided to stay within the same category. Alternatively, if a student selects to go to a different category, they do not automatically know what the correct answer will be. However, they are able to eliminate the category they just reviewed as a possible answer choice. Thus, the student can reject one answer option without needing to retrieve information about the new exemplar or category membership for it. By contrast, when learning is controlled by the experimenter, students never know which category they will see on an upcoming trial. This forces them to engage in the retrieval process more often relative to students who self-regulate during learning.

The goal of this experiment was to investigate how retrieval practice and the learning context (self-regulated or experimenter-controlled) impacts category learning. To meet this goal, we asked students to learn how to categorize organic chemistry compounds (materials provided by Eglington and Kang, 2017). All participants studied the same exemplars but were randomly assigned to either study or take practice tests. Critically, practice tests encourage retrieval practice, whereas study does not. Additionally, participants were randomly assigned to either a self-regulated or experimenter-controlled group. Participants in both groups were allowed to study the exemplars for as long as needed, and they were able to move on to the classification tests after studying 72 exemplars. However, the self-regulated group made decisions about the order of exemplars during study whereas the experimenter-controlled group did not. All participants took a studied and novel exemplar test followed by a short post-test questionnaire. I hypothesized that when category learning is controlled by researchers, classification performance will benefit from retrieval practice, which would replicate Jacoby et al. (2013). By contrast, when category learning is controlled by participants (i.e., is self-regulated), classification performance will not benefit from retrieval practice. This outcome would be consistent with Babineau et al. (2022). As such, the primary measure of interest was test performance between the self-regulated groups, as this was one of the main discrepancies between the Jacoby et al. and Babineau et al. studies. Additionally, I expected that the test groups would perform better on the studied and novel exemplar tests relative to the study groups, which is a prediction consistent with the larger literature on test-enhanced memory (Carpenter et al., 2022; Roediger & Karpicke, 2006).

METHOD

Participants and Design

Two hundred and thirty-four undergraduate students from Texas Christian University were recruited to participate, and they received extra credit in their psychology courses. Prior to data collection we established a stopping rule of 50 participants randomly assigned to each group for a total of 200 participants. We intentionally overcollected data to account for possible data collection issues (e.g., experimenter error, complications with technology, unexpected on-campus incidences) to ensure that there would be sufficient data to analyze. There were no technical difficulties or errors with the first 200 participants' data; therefore, none of the additional data was included in the study or analyses. Critically, after the 200 participant criterion was met, the additional data were not viewed and did not influence any of our decisions about analyses or the obtained outcomes. Further, this experiment and sample size was preregistered with the Open Science Framework (OSF) prior to data collection (https://osf.io/f6etk/), and all data and materials are freely available.

This study utilized a 2 (learning context: self-regulated, experimenter-controlled) x 2 (strategy: study, test) between-participant design. Participants were randomly assigned to one of four groups: SRL Study group (n = 50), the SRL Test group (n = 50), the Experimenter-controlled Study group (n = 50), or the Experimenter-controlled Test group (n = 50). Four participants self-identified as a chemistry or biochemistry major. One participant self-identified as a chemistry or biochemistry minor. Students' ages did not differ between groups (SRL Study, M = 19.18, SE = .17; SRL Test, M = 19.63, SE = .24; Experimenter Controlled Study, M = 19.56, SE = .26; Experimenter Controlled Test, M = 19.26, SE = .19), $F(3, 195) = 1.06$, $p = .369$. Most students identified as a woman (n = 154, 77% of sample), 18.5% of students identified as a man (n = 37), and 4.5% of students identified as a man and a woman, a woman and gender diverse, or gender diverse (n = 9). The gender distribution did not differ between groups, χ^2 (12) = 8.51, p = .744. Participants were able to report their race/ethnicity freely. Most students identified as Caucasian (n = 132, 66% of the sample), 10.5% of students identified as two or more ethnicities (n = 21), 8% of students identified as Hispanic and/or Latino (n = 16), 5.5% of students identified as African American and/or Black (n = 11), 4% of students identified as Asian and/or Asian American (n = 8), 3.5% of students identified as from a specific region or country (n = 7), and 2.5% of students preferred not to report their race/ethnicity (n = 5). The race distribution did not differ between groups, χ^2 (21) = 20.96, p = .461. Further, few participants reported experience as majors or minors related to chemistry (see Table 1). There were no significant differences in majors or minors related to chemistry between groups, χ^2s $\leq$ 3.02, ps $\geq$ .389, and the groups did not significantly differ in participants' self-rated knowledge of chemistry, *F*(3, 196) = .08, *p* = .973.

Table 1 *Participants' responses to each prior knowledge question*

Question	No SRL Study Group (*n* = 50)	No SRL Test Group (*n* = 50)	SRL Study Group (*n* = 50)	SRL Test Group (*n* = 50)
1. Are you currently or have you ever been a Chemistry/Biochemistry Major?	2	0	1	1 (1 no response)
2. Are you currently or have you ever been a Chemistry/Biochemistry Minor?	0	0	1	0
3. What is your own rated level of expertise at identifying chemical compounds? (1-7)	2.08 (.20)	2.06 (.18)	2.18 (.19)	2.10 (.20)

Note. The number of participants to said "yes" to questions 1 and 2 are provided. For question 3, the mean rating is provided with the standard error in parentheses.

Materials

Participants learned to classify organic chemistry stimuli that were obtained from Eglington and Kang (2017). Stimuli were images of exemplars from six chemical categories: epoxide, nitrile, organochloride, organophosphate, pyrethroid, and sulfone. Each exemplar was a two-dimensional chemical compound in black and white, and each category had 20 exemplars (120 total). For each category, 12 of the exemplars were randomly assigned to the study phase and studied classification test (72 total exemplars), and the remaining 8 exemplars from each category were assigned to the novel classification test (48 total exemplars).

Procedure

Participants first completed a consent form agreeing to participate in the study. They then began the experiment with prior experience questions. The prior knowledge questions asked participants to report if they were or have ever been a chemistry or biochemistry major or minor. Participants were also asked to rate their current level of chemistry expertise on a scale of 1 (no knowledge) to 7 (expert) (see Table 1). Participants then completed a chemistry prior knowledge test that consisted of three free-response questions. The three questions aimed to determine participant's general knowledge about organic chemistry principles. For each question participants typed their answer or entered "I don't know" and each question was self-paced and presented one-at-a-time.

After completing the prior knowledge questions, participants completed a chemistry tutorial (Babineau & Tauber, 2022). The goal of the tutorial was to familiarize participants with 2-dimensional representations of organic chemistry compounds, and it presented information on how molecules and bonds are represented in chemical structures. Participants were given the option to return to the beginning of the tutorial at any time if they wanted to review the material again. At the end of the tutorial, participants were given six exemplars from chemical categories that were not included in the study, practice classification, or test phases of the experiment and asked to identify key elements of them (e.g., *click on the nitrogen molecule*, *click on the three-member ring*). Participants were required to correctly answer each question before proceeding to the main task. By correctly answering the tutorial questions, participants demonstrated basic understanding of 2-dimensional depictions of organic chemistry compounds, which is necessary for the primary learning task.

Following the tutorial, participants began the study phase. Participants in all groups were instructed to study the exemplars in preparation for a novel classification test. Participants assigned to the Experimenter-controlled Study group were shown the 72 exemplars assigned to the study phase. The exemplars were displayed one-at-a-time and were shown in the center of the screen with their correct compound name displayed beneath. Participants viewed the exemplar for 3 seconds before they could proceed to the next exemplar by selecting the “next” button that appeared at the bottom of the screen. This was implemented to ensure that participants spent an adequate amount of time viewing each exemplar and allowing them to study the exemplar for longer than 3 seconds if they desired to do so. There was a brief 0.5 second inter-stimulus-interval before the next exemplar was presented. The exemplars were displayed in a new interleaved order for each participant. For the interleaved order, the exemplars are arranged so that participants never saw two exemplars from the same category back-to-back.

Participants assigned to the Experimenter-controlled Test group were tested on the 72 exemplars assigned to the study phase. The exemplars were displayed one-at-a-time and were shown in the center of the screen with 6 potential category names displayed on buttons beneath in alphabetical order. Participants answered the question, “*What type of compound is this?*” by selecting the button corresponding to their answer. After answering the question, participants received corrective feedback above the exemplar, (e.g., *Correct this is a Nitrile*; or *Incorrect, this is a Nitrile.*). Participants viewed the feedback for 3 seconds before they could proceed to the next exemplar by selecting the “next” button that appeared at the bottom of the screen. There was a brief 0.5 second inter-stimulus-interval before the next exemplar was presented. The exemplars were displayed in a new interleaved order for each participant.

For participants in the SRL Study group, the study phase was nearly identical to the Experimenter-controlled Study group. For both groups, the first exemplar presented was randomly selected by the program for each participant; however, after that, participants in the SRL Study group selected the study order. Thus, the key difference between the SRL Study group and the Experimenter-controlled Study group was that for the SRL study group, study order was determined by each participant, whereas for the Experimenter-controlled Study group, exemplars were presented in an interleaved order. Participants in the SRL Study group made a study order decision following each exemplar. The study decision screen asked participants, "*What type of compound do you want to study next?*" with buttons below labeled "*Same type of compound*", "*Different type of compound*" or "*I am ready for the test*". Participants who selected to study the same type of compound (i.e., to stay within the category) were shown a different exemplar from the same compound category, and participants who selected to study a different type of compound (i.e., to switch to a different category) were shown an exemplar from a different, randomly selected compound category. Participants proceeded through the study phase until they had completed at least 72 trials, in any order. Participants could continue studying after the 72 trials if they desired to do so. However, once the 72 trials were completed, the option to proceed to the test appeared. This was done to ensure that participants in the Experimenter-controlled groups and participants in the SRL groups had the same number of study opportunities.

For participants in the SRL Test group, the study phase was nearly identical to the Experimenter-controlled Test group. For both groups, the first exemplar presented was randomly selected by the program for each participant; however, after that, the participants in the SRL Test group selected the study order. Participants in the SRL test group made a study order decision following each exemplar. The study order decision procedure was identical to that used in the SRL Study group such that participants selected to study an exemplar of a different category (i.e., switch) or the same category (i.e., stay) on the next trial.

Once participants completed the study phase, they were directed to the test phase. Participants completed a novel classification test and a studied classification test, and the order of the tests was counterbalanced between participants. During the novel classification test, a never-before-seen exemplar belonging to one of the studied compound categories was shown on the screen alongside the possible compound names. Participants classified each novel exemplar one-at-a-time and in a random order. Feedback was not provided during the novel classification test, and participants had no time limit to respond. During the studied classification test, the exemplars that were learned during the study phase were shown on the screen alongside the possible compound names. Participants classified each exemplar one-at-a-time and in a random order per participant. Feedback was not provided during the studied classification test, and participants had no time limit to respond. After completing the test phase, participants completed a brief follow-up question about their study strategy beliefs. Specifically, participants were provided with a description of the study group and the test group and asked to indicate which study strategy they thought would be best for their leaning. Participants selected from two options (study strategy or test strategy) that were shown in a random order and participants had no time limit to respond. Finally, participants were debriefed, thanked for their time, and granted credit for their participation.

RESULTS

Participants' Prior Knowledge

Prior to any experimental manipulation, participants completed a test evaluating their prior knowledge in organic chemistry. Responses to the three questions were scored for accuracy and combined to create a composite prior knowledge score (0%-100% correct) for each participant. The composite prior knowledges scores were low (Experimenter-controlled Study, *M* = 2.67, *SD* = 11.35; Experimenter-controlled Test, *M* = 4.00, *SD* = 12.85; Self-Regulated Study,

$M = 6.00$, $SD = 17.42$; Self-Regulated Test, $M = 4.00$, $SD = 17.35$), and they did not significantly differ between the groups, , $F(3, 196) = .42$, $p = .738$. Based on these prior knowledge data, we concluded that all groups had little prior knowledge of organic chemistry and their prior knowledge organic chemistry did not statistically differ.

Participants’ Test Performance

As evident from Figure 1, participants’ performance on the studied exemplar test significantly differed between the study ($M = 0.71$, $SE = 0.02$) and test ($M = 0.82$, $SE = 0.01$) groups, $F(1, 196) = 18.87$, $p < .001$, $\eta^2 = .088$. However, performance did not significantly differ based on whether students self-regulated their learning ($M = 0.76$, $SE = 0.02$) or if it was controlled by the experimenter ($M = 0.77$, $SE = 0.02$), $F(1, 196) = .08$, $p = .773$. Further, there was not a significant interaction between learning context and strategy, $F(1, 196) = 1.37$, $p = .243$, (SRL Study, $M = 0.73$, $SE = 0.03$; SRL Test, $M = 0.81$, $SE = 0.02$; Experimenter Controlled Study, $M = 0.69$, $SE = 0.03$; Experimenter Controlled Test, $M = 0.83$, $SE = 0.02$).

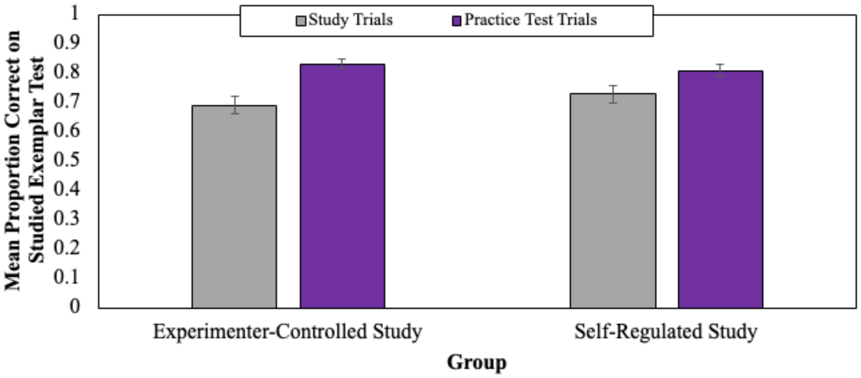

Figure 1 Performance on studied exemplar test. Error bars represent standard error.

Similarly, participants’ performance on the novel exemplar test significantly differed between the study ($M = 0.69$, $SE = 0.02$) and test ($M = 0.80$, $SE = 0.01$) groups $F(1, 196) = 18.39$, $p < .001$, $\eta^2 = .086$. However, performance was not influenced by whether study was self-regulated (*M* = 0.75, *SE* = 0.02) or experimenter-controlled (*M* = 0.75, *SE* = 0.02), *F*(1, 196) < .001, *p* = .993 (see Figure 2). Further, there was not a significant interaction between learning context and strategy, *F*(1, 196) = 1.22, *p* = .272 (SRL Study, *M* = 0.71, *SE* = 0.03; SRL Test, *M* = 0.79, *SE* = 0.02; Experimenter Controlled Study, *M* = 0.68, *SE* = 0.03; Experimenter Controlled Test, *M* = 0.82, *SE* = 0.02).

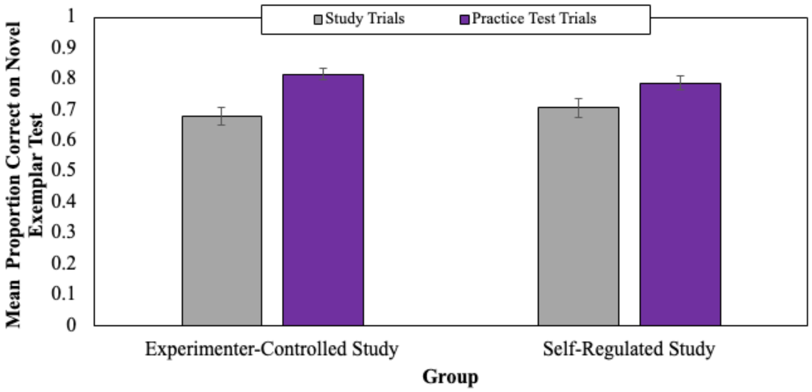

Figure 2 Performance on novel exemplar test. Error bars represent standard error.

Time Spent Studying and Test Performance

There was a significant difference in time spent studying each exemplar between the study (*M* = 5.739, *SE* = 0.232) and test (*M* = 4.429, *SE* = 0.222) groups *F*(1, 196) = 15.95, $p < .001$, $\eta^2 = .075$. Specifically, participants who were in the study groups spent longer studying each exemplar than did participants who were in the practice test groups. The amount of time studying was not influenced by whether the learning was self-regulated (*M* = 5.31, *SE* = .232) or experimenter-controlled (*M* = 4.86, *SE* = .232), *F*(1, 196) = 1.83, *p* = .178. There was also a significant difference in time spent exposed to the correct exemplar and category name pairing between the study (M = 5.739, SE = 0.239) and practice test (M = 8.56, SE = 0.239) groups, *F*(1, 196) = 69.59, $p < .001$, $\eta^2 = .26$.

Participants in the practice test condition received corrective feedback after answering each question. The time spent reviewing this feedback did not significantly differ based on whether students self-regulated their learning (M = 4.20, SE = .08) or if it was controlled by the experimenter (M = 4.05, SE = .06), *t*(98) = 1.51, *p* = .13. Participants in the self-regulated groups were given a choice about which category they wanted to study after each trial was completed. The time spent making this decision did not significantly differ between the study (M = 1.44, SE = .12) and practice test (M = 1.34, SE = .05) groups, *t*(98) = .72, *p* = .47.

Due to the study time difference between the study and test groups, I conducted two multiple linear regressions to explore the degree to which the study time difference impacted the practice test effect on participants' test performance. One multiple linear regression model was conducted on participants' novel test performance and the other was conducted on participants' studied exemplar test performance. Participants' self-paced study time during learning was uncentered. Each model included two dummy coded predictors and the interaction between them. Learning strategy was dummy coded (study = 0, practice test = 1) as was learning context (experimenter-controlled = 0, self-regulated = 1). Thus, each model included study time (uncentered) to control for differences in time between the groups, strategy (dummy coded), learning context (dummy coded), and the interaction between learning strategy and context as predictors.

For performance on the studied exemplar test, the study groups saw exemplars and the correct category name for longer than did the practice test groups, and this difference in study times significantly impacted their novel test performance, *b* = -.011 (*SE* = .005), *t* = 2.14, *p* = .033, R^2 = .005. Thus, relative to the study groups, the testing groups did significantly better on the novel test due to additional time spent with the exemplars during learning. However, after controlling for study time, the practice test groups outperformed the study groups (see Figure 1). Specifically, relative to the study groups, the test groups performed significantly better on the studied exemplar test while controlling for self-paced study times during learning, $b = .174$ ($SE = .04$), $t = 4.46$, $p < .001$, $R^2 = .14$. Studied exemplar test performance was not predicted by learning context (experimenter-controlled versus self-regulated), $b = .044$ ($SE = .04$), $t = 1.23$, $p = .222$, $R^2 = .002$, and learning context did not interact with study strategy, $b = -.061$ ($SE = .05$), $t = 1.21$, $p = .226$, $R^2 = .02$. In sum, students in the practice test groups spent more time with exemplars during learning, which benefitted their learning of those materials relative to the study groups. Even so, after taking into account study time, practice testing continued to benefit students' test performance with studied exemplars.

For novel test performance, even though participants in the study groups were exposed to the exemplar and the correct category name for longer than participants in the practice test groups, study times did not significantly impact their novel test performance, $b = -.009$ ($SE = .005$), $t = 1.64$, $p = .102$, $R^2 = .007$. After controlling for study time, the impact of completing practice tests on learning maintained (see Figure 2). Specifically, relative to the study groups, the test groups performed significantly better on the novel exemplar test while controlling for self-paced study times during learning, $b = .163$ ($SE = .04$), $t = 4.17$, $p < .001$, $R^2 = .13$. Novel test performance was not predicted by learning context (experimenter-controlled versus self-regulated), $b = .033$ ($SE = .04$), $t = .92$, $p = .360$, $R^2 < .001$., and learning context did not interact with study strategy, $b = -.057$ ($SE = .05$), $t = 1.13$, $p = .259$, $R^2 = .016$. Thus, the practice test effect on novel test performance is attributable to completing practice tests during learning and not to exposure time with exemplars during learning.

Participants' Beliefs about Testing

After both the novel and studied tests were completed, participants were asked about their beliefs regarding studying and self-testing (See Table 2). Participants answered a qualitative question indicating what they thought would be best for their learning: studying, testing, or no difference between the two. Most participants (*n* = 157, 78.5% of sample) reported they believed practice tests would be best for their learning. The distribution of participants' responses about whether practice tests or study was better for their learning did not significantly differ between groups, χ^2 (3, $N = 200$) = 2.46, $p = .483$.

Table 2 *Participants' Beliefs about Testing and Restudying*

Question	**No SRL Study Group (*n* = 50)**	**No SRL Test Group (*n* = 50)**	**SRL Study Group (*n* = 50)**	**SRL Test Group (*n* = 50)**
Study	7 (14%)	11 (22%)	13 (26%)	12 (24%)
Test	43 (86%)	39 (78%)	37 (74%)	38 (76%)

Note. The number of participants who indicated each response is provided separately by group. Percentages reflect the percent of participants within group.

DISCUSSION

In the current study, I sought to investigate how retrieval practice and the learning context (self-regulated or experimenter-controlled) impacts category learning and overall performance on a classification test. Participants who utilized practice tests during learning performed better on both novel and studied classification tests compared to participants who utilized study trials (see Figures 1 and 2). My hypothesis that retrieval practice would result in better test performance was confirmed. However, my hypothesis that participants in the experimenter-controlled conditions would perform better than participants in the self-regulated conditions was not confirmed.

An important variable that was further examined was the amount of time spent learning each exemplar. Participants in the study groups spent approximately 5.7 seconds looking at each compound before moving on to the next trial. Participants in the practice test groups spent approximately 8.6 seconds looking at each exemplar before moving on to the next trial. This significant increase in time was due to the testing group receiving corrective feedback after making their selection. This time difference warranted further analyses to investigate if more time spent during learning impacts test performance. This was completed by performing two linear regression analyses. For the novel classification test, time did not significantly predict performance. This result was expected due to the nature of the novel test. This test contained images that participants had never seen before. Thus, it would not have mattered how long they spent during learning, as those were not the images they would later be tested over. By contrast, for the studied classification test, more time spent during learning was associated with better test performance. However, after statistically controlling for timing, participants in the practice test groups still outperformed the study groups. This result confirms that testing was the strongest predictor of classification performance, not timing. These analyses indicate that the amount of time spent looking at the material is not indicative of overall performance on a classification test. The act of taking a practice test and forcing active recall reinforces the information in a much more effective way, which yields better performance. Our results align with current literature regarding the benefits of retrieval practice (Jacoby et al., 2013). Utilizing practice testing during learning results in better performance on a classification test. This can aid both students and educators to implement more beneficial study techniques. Students can begin taking practice tests to better learn material. Educators can create practice tests or active recall scenarios to help students reinforce their knowledge.

Surprisingly, there was no significant difference in overall test performance between participants who regulated their own learning, and participants whose learning was regulated by the experimenters. Our hypothesis that participants in the experimenter-controlled conditions would perform better than participants in the self-regulated conditions was not confirmed. These results offer an interesting new point regarding retrieval practice. One of the major differences between the Jacoby et al. (2013) study and the Babineau et al. (2022) study was the introduction of self-regulation. When students self-regulate their learning, their choices may limit the need to actively recall information from memory. We believed that self-regulation was an important component of the testing benefit, leading us to further manipulate this variable in the current study. Our findings indicate that learning context, self-regulated or experimenter-controlled, do not significantly impact test performance. This result can transfer well into a real-world application. Regardless of if students design their own practice tests or if they utilize tests made by a teacher or a third party, they will still receive the testing benefit.

Because our learning context hypothesis was not supported, more work is needed to explore why the prior research show different practice test effects. Another discrepancy between the Jacoby et al. (2013) study and the Babineau et al. (2022) study was the number of categories that participants learned. The Jacoby et al. (2013) study has participants learn 8 categories, while the Babineau et al. (2022) study only had participants learn 3 categories. It is possible that participants in the Babineau et al. (2022) study did not learn enough categories to fully get the testing benefit. A future study could utilize a similar design to the current study, but have participants learn only 3 categories. This decrease from 6 categories to 3 categories may yield different results between participant groups.

Another measure of interest in this study was participants' responses to the preferences and beliefs question posed at the end of the study. A total of 78.5% of the sample reported that they believed testing would be more beneficial for their learning relative to restudying. This was a unique finding because students believe testing will be beneficial for them, and testing is beneficial for them as evidenced by Figures 1 and 2. Thus, students' thoughts and opinions about practice testing are consistent with scientific evidence about the effectiveness of using this learning strategy. As mentioned above, students can continue to utilize practice tests and educators can design these tests along with other active recall scenarios to help students better learn material.

Collectively, the results of this study reveal that when learning complex natural categories, utilizing practice tests rather than study trials will yield better results on a classification tests. This includes classification tests of the exact examples used during learning and classification tests that require transferring knowledge to never-before-seen examples. We also found that practice testing was a significant predictor of overall test performance after controlling for the amount of exposure time to the examples during learning. Future research should aim to understand the relationship between self-regulation, retrieval practice, number of categories learned, and final test performance.

REFERENCES

Babineau, A. L., Witherby, A. E., Ariel, R., Pelch, M. A., & Tauber, S. K. (2022). Do Domain Knowledge and Retrieval Practice Predict Students' Study Order Decisions?. *Journal of Intelligence*, *10*(4), 122. https://doi.org/10.3390/jintelligence10040122

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1441–1451. https://doi.org/10.1037/a0020636

Eglington, L. G., & Kang, S. H. K. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition, 6*(4), 475–485. https://doi.org/10.1016/j.jarmac.2017.07.005