DOES PRE-TESTING ENHANCE LEARNING WHEN IT IS DONE COVERTLY?

by

Ashley Berdelis

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

May 6, 2024

DOES PRE-TESTING ENHANCE LEARNING WHEN IT IS DONE COVERTLY?

Project Approved:

Supervising Professor: Uma Tauber, Ph.D.

Department of Psychology

Michelle Rivers, Ph.D.

Department of Psychology

Eric Cox, Ph.D.

Department of Political Science

ABSTRACT

Students wonder how to prepare for an upcoming exam, while instructors wonder how to best introduce materials to students. The present study combines the research areas of covert versus overt learning and pre-instruction testing: When taking a pretest, is it better to just think about the answer, or is it better to write it down? Undergraduate students at TCU were instructed to learn passages about the planet Saturn or Yellowstone National Park. For one of the passages, students simply read the passage (read-only condition). For the other passage, they were asked short-answer pretest questions about the passage prior to reading (pre-test condition). A random half of the participants completed their pretests overtly (typing their responses to the questions), whereas the other half of participants completed their pretests covertly (answering the questions in their mind). All participants completed a final multiple-choice test on the material they learned. Performance on this final test was higher for the pretest condition compared to the read-only condition, and this was true for both overt and covert pretesting. In strategy ratings made after the experiment, participants seemed to recognize that pretesting was more effective than learning because it helped them absorb relevant information while reading the passage. The results suggest that pretesting is an effective learning strategy, even when learners do not provide articulated responses.

Does Pretesting Enhance Learning When it is Done Covertly?

The question being investigated in this paper is: does pretesting (answering pre-instruction questions) enhance learning when it is done covertly? Practice testing tends to be more effective when students recall material verbally or write it down (overt learning) compared to when they only mentally recall the material (covert learning; Tauber et al., 2018). Other research evaluates the value of testing individuals before they learn subject matter. Studies in this field show that answering test questions before studying, even if people provide wrong answers, leads to beneficial learning outcomes compared to just studying material (Dunlosky et. al, 2013). I aimed to combine both research areas and investigate how covert and overt learning applies to pretesting to evaluate whether it is better to just think about the answer (covert response) or write it down (overt response) when completing pre-instruction questions.

Pre-instruction testing is when students are given test questions before learning that material (for a review, see Carpenter et al, 2023). For example, to introduce a new unit in class, a teacher may ask questions to students about the upcoming content. Or, students might complete a pre-test of their knowledge prior to learning about it in class to illustrate how much they can learn about the concepts. Pre-instruction testing is best described by the three-stage framework (Pan & Carpenter, 2023). The first stage is the actual pretesting, where questions about the material are presented. The second stage is the lesson or learning opportunity for the material, done through lecture, video, or text. The third and final stage is a posttest on the material in the previous two stages. Engaging in pre-instruction questions helps students learn that material better in the lesson relative to material that was not pretested. Answering pretest questions, even if students guess or answer the questions wrong, still leads to beneficial learning outcomes compared to just studying material.

Pan and Sana (2021) investigated if pretesting is competitive with posttesting. These researchers ran five experiments with variations on question type (short answer versus multiple choice), delays between the learning phases, and how or whether feedback was presented. Across all the experiments, participants read one or two text passages that were about 1,100 words long on the subject of Yellowstone National Park or the planet Saturn. Participants in Experiments 1-2 went through two blocks, and participants in Experiments 3-4 had some modifications of the general format. During the first block, the procedure followed the format of pretest, reading passage 1, five minute distractor test, and finally a criterial test on passage 1. The second block started with reading passage 2, then a posttest on the material, five minute distractor task, and finally a criterial test on passage 2. Experiment 5 had a pretest condition, a posttest condition, and a read-test-read condition. The results of Experiment 1 and 2 showed that pretesting has a significant effect on promoting greater learning than a posttest, whether the format was multiple choice or cued recall. Experiment 3, where feedback was given, did not show any effect. Experiment 4 (48-hour retention interval) showed an advantage of pretesting as opposed to posttesting, but only when it came to previously seen questions. Experiment 5 also showed an advantage of pretesting over posttesting. The results of this study show that there is a repeat advantage from pretesting on learning. The subject matter that showed a benefit in learning was also observed only for material covered on the pretest. Additionally, both pretesting and posttesting showed a learning benefit when compared to no testing at all.

To explain the theory behind the effects of pre-instruction testing, it is best to return to the three-stage framework (Pan & Carpenter, 2023). The first stage of the pre-instruction test uses psychological processes not triggered by non-testing methods. There are two explanations for this. First, students may be triggered into curiosity by the questions and material they do not

know the answer to. Second, the psychological processes may include creating a memory of the practice question and its contents, developing a mental framework for the information on the pretest for what is going to be learned, and coming up with possible answers to form specific memories or create retrieval routes for the answers. The second stage encompasses learning material to which the information is encoded into the long-term to a better extent than if no pre-instruction test was presented. Finally, the posttest shows improved performance when going through the previous two stages. This is due to better encoding in the second stage and/or assistance from specific memories or better retrieval routes formed during the first stage.

Most of the research in the pre-instruction test phase has been done overtly, but not much has been explored covertly. Overt learning is when students recall the material verbally or when they write it down, whereas covert learning is only mentally recalling the information silently. Investigating covert pre-instruction questions is important because we have already seen that pre-instruction testing can be effective, so it is important to investigate the reach and type of pretesting. Research on testing that occurs during learning reveals that practice testing tends to be more effective when students recall the material overtly, rather than covertly (Tauber et al., 2018). Gathering information about the effectiveness of covert and overt pre-instruction questions for enhancing student learning would be helpful for instructors to know if they were to employ the pretest method in their classes.

In this project, undergraduate students at Texas Christian University were instructed to learn passages about the planet Saturn or Yellowstone National Park. For one of the passages, students simply read the passage (read-only condition). For the other passage, they were asked short-answer pre-instruction test questions about the passage prior to reading (pre-test condition). Half of the participants completed their pre-instruction tests overtly (typing their responses to the

questions), whereas the other half of participants completed their pretests covertly (answering the questions in their mind). All participants completed a final multiple-choice test on the material they learned. The hypotheses account for the read-only control, overt pre-instruction questions, and covert pre-instruction questions. First, I predict that final test performance for the read-only control condition will not significantly differ between the covert pre-instruction and overt pre-instruction question groups because these conditions are identical. I also predict that overt pre-instruction questions will lead to significantly better performance on the final test for the same questions relative to new questions, which would replicate Experiment 1 of Pan and Sana (2021). I predict that overt pre-instruction questions will lead to significantly better performance on the final test for the same and for new questions relative to the read-only control condition. Pan and Sana (2021) did not report this comparison. Lastly, I hypothesize that covert pre-instruction questions may be less effective for learning relative to overt pre-instruction questions. Specifically, a smaller effect is expected for the benefit of pre-instruction questions on final test performance for the same questions relative to new questions and relative to control, and performance on the same questions may be lower relative to same question performance for the overt pre-instruction question group. Performance on new questions may not differ from the read-only control.

Method

Participants

Our target sample size was 176 participants (i.e., 88 participants per group). We planned to oversample slightly (no more than 10%) in anticipation of dropping some participants from analyses due to our predetermined exclusion criteria. Given that the outcomes from our research have applied implications, particularly in an educational context, we were only interested in

effects that are of at least moderate size. Thus, we powered for a medium effect. An a priori power analysis was conducted using G*Power 3.1.9.7 (Faul et al., 2007) to determine the necessary sample size to detect a medium effect (Cohen's $d = 0.5$) with 95% power and an alpha error probability of .05 for a one-tailed independent samples *t*-test. Exclusion criteria included: technology issues during the experiment (e.g., a glitch in the program, failure to save data, loss of internet connection); participants not finishing the experiment; significant distractions during the experiment (e.g., participant falling asleep, fire alarm).

Sixteen participants (8 from each group) were excluded from analyses due to an unfortunate coding error; these participants were pretested on the incorrect passage, making their data uninterpretable. An additional 14 participants were excluded from analyses (9 from the overt group, 5 from the covert group) due to a computer error and/or failure to finish the experiment.

My final sample included 165 students who were recruited through Texas Christian University's (TCU) Human Subject Pool. All participants were undergraduates at TCU and were within the age range of 16 years old through 39 years old (M age = 19.7, $SD = 2.06$). Students who were under 18 years of age obtained guardian permission to take part in this study. Most of the participants (72.7%) identified as a woman (26.1% as a man, and 1.2% as gender diverse or non-conforming). Most participants (63%) identified as White (6.1% as Asian, 7.3% as Black/African American, 13.9% as Hispanic or Latino, and 9.7% as a mixed-race ethnicity). All participants received partial course credit in Psychology courses. Data collection occurred in person in the Fall 2023 semester in the Tauber lab spaces on the TCU campus.

Design

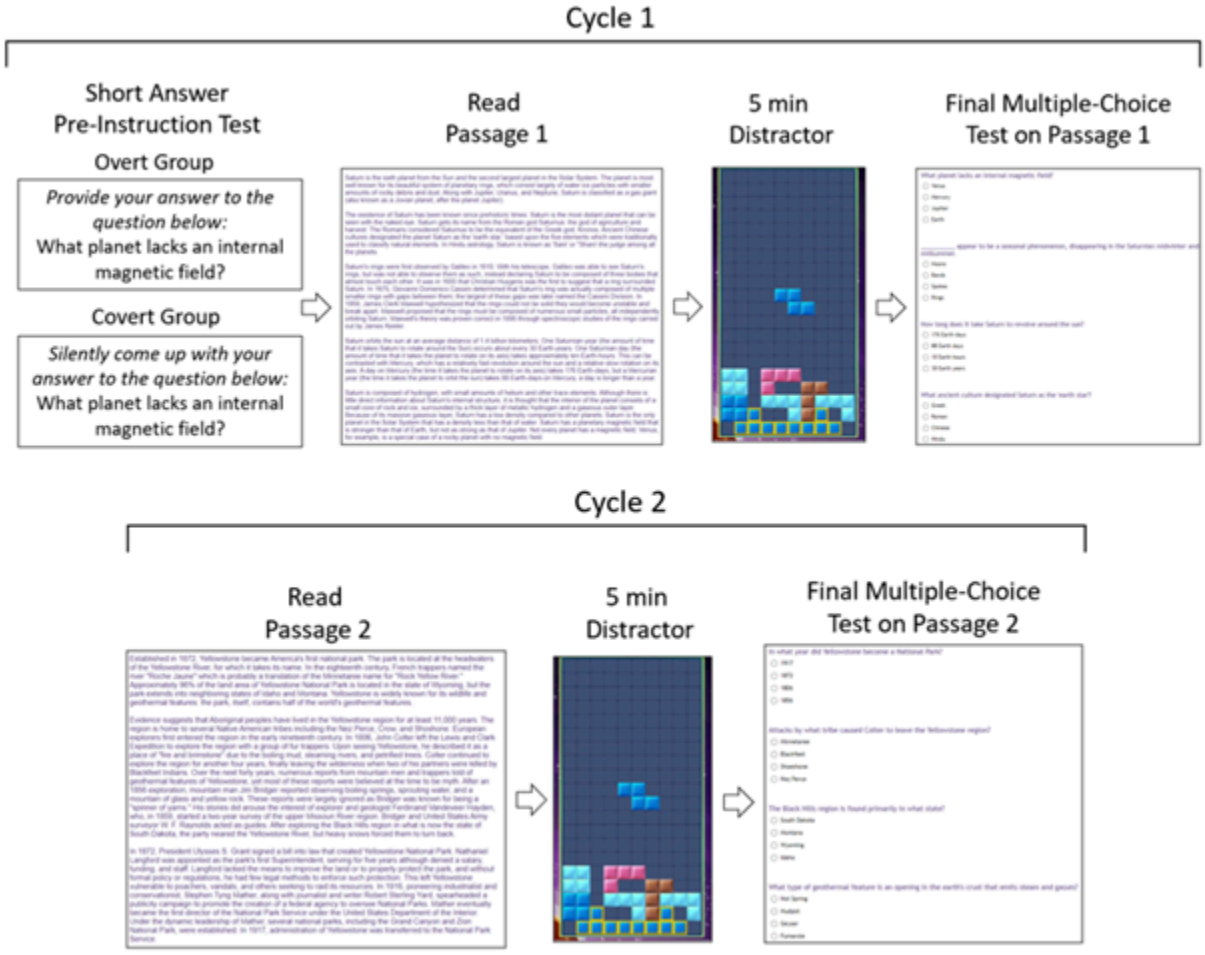
A 2 (pre-instruction modality: overt vs covert pre-instruction questions) x 3 (question type: same, new, read-only control) mixed-factor design was used, in which question type was manipulated within-participant and pre-instruction modality was manipulated between-participants. Thus, participants were randomly assigned to one of two groups: overt pre-instruction questions (n = 82) or covert pre-instruction questions (n = 83), and both groups completed a read-only control learning activity.

Materials

This study was conducted via Qualtrics programming software. Participants were given expository texts that were approximately 1,100 words long on Yellowstone National Park and the planet Saturn (taken from Pan & Sana, 2021). I also used pre-instruction test questions (ten per passage) and final test questions (twenty per passage) from the same source (Pan & Sana, 2021). See Figure 1 for examples of these materials and an overview of the study procedure.

Figure 1

Study Procedure



Note. All participants completed the two cycles. Passage order (Yellowstone or Saturn) and learning activity order (pretest or read-only control) were counterbalanced across participants. For simplicity, only one of two counterbalanced orders is shown.

Procedure

Participants completed the study individually on computers in the Tauber lab at TCU. They filled out the consent form and were then asked to rate their knowledge about Yellowstone

and Saturn. The order of presentation was counterbalanced across participants. They responded using a 5-point scale from “not knowledgeable at all” to “extremely knowledgeable.”

The experiment involved two cycles, illustrated in Figure 1 – and cycle order was counterbalanced across participants. During one cycle, participants completed a pre-instruction test phase. Specifically, participants read 10 short-answer pre-instruction test questions, and answered each overtly – by typing their responses – or covertly – by mentally thinking of their answers, depending on their group assignment.

Their instructions were to try their best, take a guess, and that there would be no penalty for incorrect responses. No feedback was provided for this activity, and participants were required to spend a minimum of 5 sec on each pre-instruction test question before they could move on. Next, participants entered the learning phase. During learning, participants were given a passage on Yellowstone or Saturn, which was counterbalanced across participants as well as cycles. In other words, participants read one passage per cycle so by completing both cycles, all participants read both passages, and passage order and associated learning task per passage depended upon group assignment. Participants were given unlimited time to read the passage, but they were required to read for a minimum of 3 min. Then, participants worked on a distractor task for five minutes, which was to play Tetris. Finally, participants entered the final test phase. This included 20 multiple-choice questions on the passage that participants read for the cycle. The test included questions that participants had previously answered during the pre-instruction test phase (i.e., same questions) as well as new questions that they had not previously encountered. This set of questions was counterbalanced across participants such that they were equally likely to appear as (a) both pre-instruction questions and final test questions, or (b) only final test questions. The question order was randomized anew per participant. The order of

response options was also randomized for each question. No feedback was provided on the final test.

During the other cycle, participants did not complete a pre-instruction test phase. Instead, they completed the read-only learning activity. Otherwise, this cycle had the same procedures, just with the other passage (Yellowstone or Saturn). Specifically, the read-only learning activity included a learning phase, distractor task, and final test phase.

The study ended with strategy beliefs and demographics. Participants answered questions about how effective they thought (1) answering the questions and then reading the passage, and (2) reading the passage without answering questions first on a 5-point scale from "not effective at all" to "extremely effective." Then, participants were asked why they made these ratings. Next, participants answered demographic questions (e.g., age, gender, race/ethnicity). Participants were told they could type "prefer not to answer" if they did want to answer these questions.

Results

The between-participant variable was whether participants received covert or overt pre-instruction questions. The within-participant variable was that all participants completed a read-only control learning activity and a pre-instruction question learning activity. My primary dependent measure of interest was performance on the final test that included the same and new questions relative to performance on the final test for the read-only control condition.

Mean Proportion Correct on Final Multiple-Choice Test

A 2 (pre-instruction modality group: overt vs covert pre-instruction questions) x 3 (question type: same, new, read-only control) mixed-effects ANOVA on the mean proportion correct on the final test revealed a large significant effect for question type (same questions: M = .54, SE = .02, new questions: M = .44, SE = .02, read-only control: M = .42, SE = .01), $F(2, 326)$

= 31.60, $p < .001$, $\eta^2_p = .16$. However, there was no significant main effect of the group, $F(1, 163) = 1.08$, $p = .30$, $\eta^2_p = .01$. Thus, covert and overt pre-instruction testing had approximately the same impact on final test performance. The interaction between question type and group was not significant, $F(2, 326) = 1.16$, $p = .31$, $\eta^2_p = .01$. Figure 2 represents these outcomes for both groups.

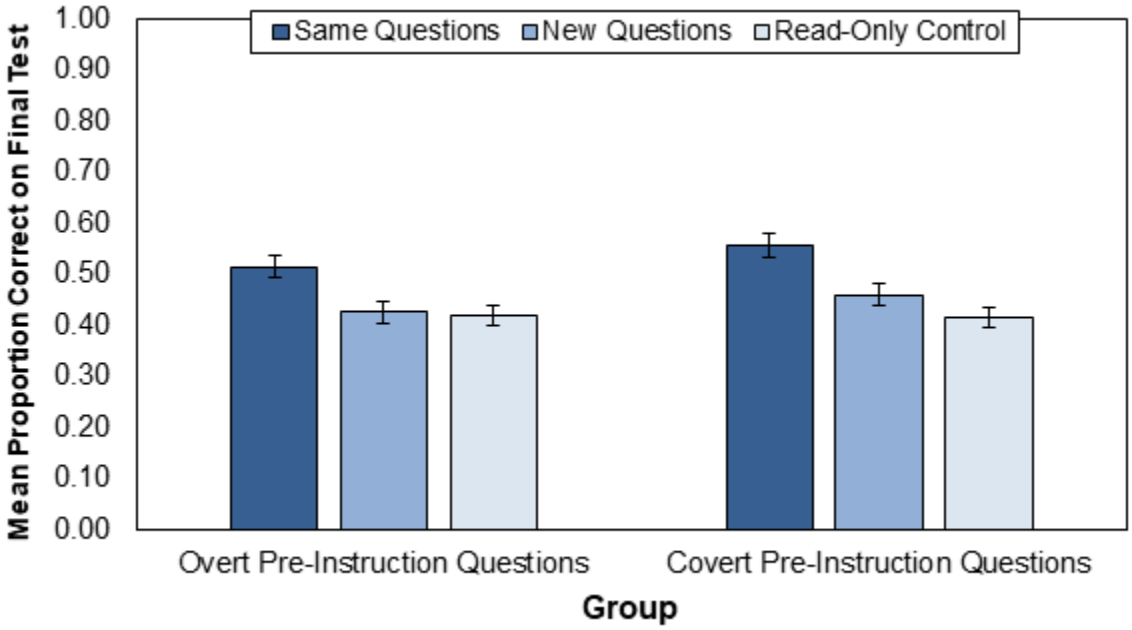
To better understand differences found for question type, we conducted dependent-sample *t*-tests for both the overt and covert groups. For the overt group, performance on the final test was higher for same questions compared to new questions, $t(81) = 3.88$, $p < .001$, $d = 0.43$. Performance was also higher for the same questions compared to the read-only control, $t(81) = 3.95$, $p < .001$, $d = 0.44$. However, performance did not differ for new questions and the read-only control, $t(81) = 0.29$, $p = .39$.

For the covert group, performance on the final test was higher for same questions compared to new questions, $t(82) = 4.56$, $p < .001$, $d = 0.50$. Performance was also higher for the same questions compared to the read-only control, $t(82) = 7.00$, $p < .001$, $d = 0.77$. Finally, performance was higher for new questions and the read-only control, $t(82) = 2.30$, $p = .02$, $d = 0.25$.

In sum, overt and covert pre-instruction questions are an effective way to learn, and both are more beneficial than just reading.

Figure 2

Performance on Same Questions, New Questions, and Read-Only Control Questions for the Overt and Covert Groups



Note. Error bars represent the standard error of each mean.

Strategy Effectiveness Ratings

We ran a mixed-factor ANOVA to look at the strategy effectiveness ratings. The within subject factor is the strategy type (reading or pretesting). The between subject factor was the group (covert or overt pretest group). The dependent measure was average ratings on a 5-point scale from "not effective at all" to "extremely effective."

For the within subject factor of strategy type, the measure revealed a large significant effect (pre-instruction testing: $M = 3.29$, $SE = .09$, reading: $M = 2.20$, $SE = .08$), $F(1, 163) =$

68.31, $p < .001$, $\eta^2_p = .30$. Essentially, pre-instruction testing was rated as "moderately effective," whereas reading was rated as "slightly effective." There was no significant difference in the ratings of strategy effectiveness between the two groups (overt: $M = 2.76$, $SE = .08$, covert: $M = 2.73$, $SE = .08$), $F(1, 163) = 0.09$, $p = .76$, $\eta^2_p = .001$. There also was no significant interaction between strategy and group, $F(1, 163) = 0.26$, $p = .61$, $\eta^2_p = .002$.

Overall, 69% of participants rated pretesting as more effective than reading, 16% of participants rated reading as more effective than pretesting, and 15% of participants showed no difference in their ratings.

Prior Knowledge Performance

Participants rated their knowledge on Yellowstone National Park and the planet Saturn on a 1-5 scale ranging from "not knowledgeable at all" to "extremely knowledgeable." We looked at the knowledge with a *t*-test with the variables being Yellowstone and Saturn. Self-rated prior knowledge was low for both Yellowstone ($M = 1.79$, $SE = .06$) and Saturn ($M = 1.58$, $SE = .06$). The ratings of prior knowledge were not significantly different between the overt and covert groups for Yellowstone, $t(163) = 0.12$, $p = .90$, or Saturn, $t(163) = 0.39$, $p = .70$. These outcomes indicate that my sample of participants did not have much existing knowledge about Yellowstone or Saturn before taking part in the experiment and most important, self-rated knowledge did not differ between the groups.

Pre-Instruction Test Performance

We can only measure the outcome of the pre-instruction test performance for the overt group because the covert group recalled answers in their head. Pre-instruction test performance

was very low for the overt group (M = .04, SE = .01). This provides more evidence that participants had little knowledge about the target concepts (Yellowstone and Saturn) prior to reading about them in the experiment.

Time Spent Reading and Completing Pre-Instruction Questions

The time spent during pre-instruction testing was evaluated by using an independent samples t-test, with the test variable being the pretest time (seconds spent per question) and the grouping variable being the group. The overt group took significantly more time (M = 12.78, SE = 0.57) to answer pre-instruction questions than did the covert group (M = 8.21, SE = 0.24), $t(163) = 7.36$, $p < .001$, $d = 1.15$. This suggests that covert pre-instruction testing may be more efficient relative to overt pre-instruction testing.

The amount of time spent reading the passage was analyzed in two ways. First, we used an independent samples t-test with time spent reading after pretesting versus without pretesting, and there was no significant difference between the overt versus covert groups for both pretesting (overt group M = 250.38, SE = 8.92, covert group M = 258.81, SE = 9.27): $t(163) = 0.66$, $p = .51$, and reading only (overt group M = 252.48, SE = 8.74, covert group M = 255.26, SE = 9.00): $t(163) = 0.22$, $p = .83$. We also ran a paired samples t-test because there were no group differences, and there was not a significant difference when measuring the time spent overall in seconds on the passages after pretesting (M = 254.62, SE = 6.42) versus reading only (M = 253.87, SE = 6.26): $t(164) = 0.14$, $p = .89$.

When comparing the amount of time spent reading the Yellowstone passage versus the Saturn passage, we ran a paired samples t-test, and there was not a significant difference between the time spent with each passage (Yellowstone passage M = 254.89, SE = 6.36, Saturn passage M

= 253.60, *SE* = 6.32): $t(164) = 0.25$, $p = .81$. These outcomes suggest that differences in test performance between the conditions are unlikely to be attributable to study time of the passage.

Discussion

When it comes to the overt and covert effects from this study, I found that there is a significant benefit to pretesting over the read only control. Questions that were seen on the pre-instruction test had a mean proportion correct that was significantly better than both the new questions and the read only control. As was shown in Experiment 1 of Pan and Sana (2021), I hypothesized that overt pre-instruction questions would lead to significantly better performance on the final test for the same questions relative to new questions. This prediction was correct, as we did see significantly better performance on the final test for the same questions relative to new questions for the overt group. However, I also saw this effect for the covert group, which was not what I hypothesized. I predicted that covert pre-instruction questions would be less effective for learning relative to overt pre-instruction questions. In this study, there was not a significant difference between the overt and covert pre-instruction groups. Additionally, the covert group took significantly less time in the pre-instruction phase than the overt group. Due to the fact that the results on the final test were about the same, we may be able to conclude that covert pre-instruction testing is a more efficient way to run a pre-instruction test.

These findings have implications for how students learn from their instructors. Overt pre-instruction could be presented in a variety of ways. First, instructors could give students a written practice test before teaching the material to them. Second, instructors could present one or multiple questions about the material and have students answer them verbally either to the whole class or in discussion groups. Covert pre-instruction does not require any writing or talking on the students' part. So, instructors could give students a list of questions about the

material for students to mentally think about. Instructors could also verbally ask students questions about the material and give students a few moments to quietly think about the answers themselves. One important takeaway that instructors need to keep in mind when designing their pre-instruction test/method is to ensure their questions encompass big ideas and concepts they want their students to really understand and remember. This is because of the finding that material presented on the pre-instruction test had better outcomes than new material.

Future research should make minor manipulations to the design of this study. First, I would want to see what the results would be if short answer questions (instead of multiple-choice questions) were used for the final test and see if the same effects are found. I am curious to see if students would be able to recall the answers from the passage as well if there were no clues of the multiple-choice options. I would hypothesize that we would still see a benefit from the "same" questions as seen on both the pre-instruction test and final test. This would replicate both the findings of this study and Experiments 1 and 2 from Pan and Sana (2021).

One other manipulation I would make in future research on this topic is a greater time length between learning the material (reading the passage) and the final test. In this study, participants had a five-minute distractor of Tetris, but it would be interesting to compare overt versus covert pre-instruction testing with more time in between reading the passage and taking the final test. In Experiment 4, Pan and Sana (2021) ran a 48-hour delay to the final test. This 48-hour retention interval showed an advantage of pretesting as opposed to posttesting, but only when it came to previously seen questions. I hypothesize that the findings would be similar to the study. Eventually after running a 48-hour experiment, it would be fruitful to run an experiment where participants are tested one week after being pre-tested and reading the passage. This

would further show the benefit in the classroom because typical class units last longer than an hour or a couple days.

Additionally, about 70% of participants stated that they thought pre-instruction testing was a more effective learning strategy relative to just reading the passage. This is an exciting finding because it not only aligns with the results the participants are showing, but also that they understand they are gaining a benefit from pre-instruction testing. The implication of this specific finding is that students feel a benefit from pre-instruction testing, which is another reason why instructors should begin to implement this concept. Pre-instruction testing gives students another memory and cue for the information, and it shows them what information is important to look for. When asked to explain why they thought one strategy was more effective than another, many participants answered that pre-instruction testing was better because they knew what to focus on, and it guided their learning in the passage. Instructors should look at doing this with important information and key takeaways from lessons.

References

Carpenter, S. K., King-Shepard, Q., & Nokes-Malach, T. (2023). The prequestion effect: Why it is useful to ask students questions before they learn. In C. E. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In their own words: What scholars and teachers want you to know about why and how to apply the science of learning in your academic setting* (pp. 74–82). Society for the Teaching of Psychology.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest, 14(1)*, 4-58.

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and pretesting effects: a review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review,* 35(4), 97.

Pan, S. C., & Sana, F. (2021). Pretesting versus posttesting: Comparing the pedagogical benefits of errorful generation and retrieval practice. *Journal of Experimental Psychology: Applied,* 27(2), 237–257.

Tauber, S. K., Witherby, A. E., Dunlosky, J., Rawson, K. A., Putnam, A. L., & Roediger III, H. L. (2018). Does covert retrieval benefit learning of key-term definitions?. *Journal of Applied Research in Memory and Cognition,* 7(1), 106-115.