Running head without "Running head:"

pg. #

Methods

Centered, bolded

Participants

Left-justified, bolded

Indent

Participants were 145 undergraduate students from Colorado State University who participated in exchange for course credit. Twenty-three participants were removed from analyses due to technical difficulties or not completing the experiment ($n = 4$), already seeing the passages ($n = 7$), or not following the experimental instructions ($n = 12$). Therefore, data from 122 participants (65 in the self-test condition, 57 in the simultaneous condition) were used in analyses. A sensitivity analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) indicated this sample size was sufficient to detect an effect size of $d = .51$, assuming an alpha of .05, power of .80, and a two-tailed test. Participants (40 men, 81 women, 1 non-binary) were between 16 and 27 ($M = 18.85$, $SD = 1.33$) years old. Two participants did not provide their age.

Design

This experiment used a 2 (strategy: simultaneous, self-test) x 2 (final test question type: factual, conceptual) x 2 (passage: ice age, monetary policy) design. Strategy was manipulated between-subjects, and question type and passage were manipulated within-subjects. Participants' final test performance was used as a measure of overall learning of the passage materials.

Materials

Two short passages on the topics of monetary policy (549 words) and ice ages (1052 words) were used for study materials (Thiede, Wiley, & Griffin, 2011). Ten multiple-choice questions for each passage developed by Thiede et al. (2011) were used for the final test. Because some previous self-testing studies have shown differences between conceptual and factual questions (e.g., Bugg & McDaniel, 2012; Davey & McBride, 1986; Denner & Rickards, 1987), it was important to consider these two types of questions separately in the current study.

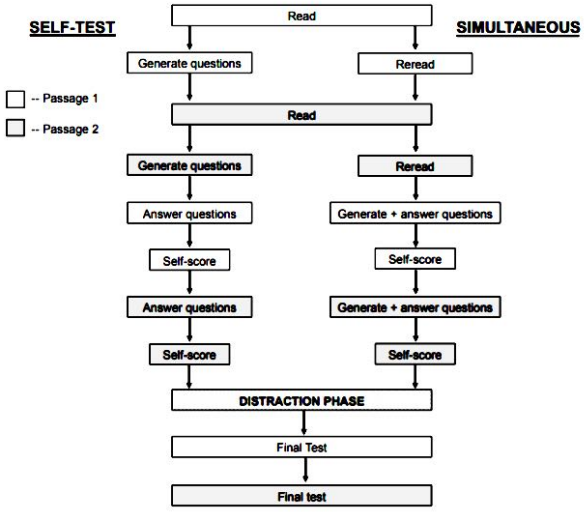
Therefore, five of the final test questions had answers that were explicitly stated within one sentence of the text (i.e., factual questions). The other five questions required participants to integrate information across at least two sentences (i.e., conceptual questions). Final test questions were originally classified as factual and conceptual by Thiede et al. (2011). However, based on the definitions used in the current study[1], one factual question on the final test was re-scored as conceptual for each passage, and one conceptual question for each passage was changed to factual. All activities of the experiment took place in Qualtrics. Both experiments were approved by the Colorado State institutional review board before data collection began.

Procedure

After providing consent, participants were randomly assigned to either generate questions and write the answers immediately while having access to the passage (simultaneous condition) or to generate questions and answer them after a delay without the passage (self-test condition). Figure 1 depicts the procedure used in Experiment 1. First, participants were given 5 minutes to read one of the two passages. Then, participants in the self-test condition were told they would create their own questions based on what they thought would be helpful for them to practice key information from the passage. After reading these instructions, participants in the self-test condition had 7 minutes to type in their questions and could scroll to see the passage underneath the response box. To equate exposure, participants in the simultaneous condition reread the passage during this time. All participants then had 5 minutes to read the second passage and 7 minutes to generate their own questions (self-test condition) or reread the passage (simultaneous condition). Order of passages was counterbalanced across participants so that half the

[1]Factual questions were defined as questions that could be answered using one sentence of the passage, whereas conceptual questions were defined as questions that required participants to integrate information across two or more sentences.

Figure 1. Diagram of Experiment 1 procedure for the self-test (answer generated questions after a delay) and the simultaneous (generate and answer questions simultaneosuly) conditions. White boxes represent tasks for Passage 1. Gray boxes represent tasks for Passage 2.



Results ← Centered, bolded

Global JOLs ← Left-justified, bolded

Indented → A 2 (strategy: self-test, simultaneous) x 2 (passage: ice age, monetary policy) mixed-factor ANOVA was conducted to analyze participants' JOL ratings, with passage manipulated within-participants and strategy manipulated between-participants. Overall, participants predicted they would perform better on the ice age final test (M = 63.59, SE = 1.60) than the monetary policy final test (M = 55.80, SE = 1.71), $F(1, 120) = 33.21$, $p < .001$. On average, participants' predictions did not differ between the self-test (M = 58.08, SE = 2.07) and simultaneous (M = 61.32, SE = 2.21) conditions, $F(1, 120) = 1.14$, $p = .29$. The passage x strategy interaction was significant, $F(1, 120) = 7.71$, $p = .01$. Follow-up t-tests indicated that, for the ice age passage, participants' predictions in the self-test (M = 63.85, SE = 2.30) and simultaneous (M = 63.33, SE = 2.20) conditions did not differ, $t(120) = 0.16$, $p = .87$, $d = 0.03$. However, for the monetary policy passage, participants in the simultaneous condition (M = 59.30, SE = 2.44) predicted they would perform significantly better than those in the self-test condition (M = 52.31, SE = 2.39), $t(120) = 2.04$, $p = .04$, $d = 0.37$.

Final Test Performance

A 2 (strategy: self-test, simultaneous) x 2 (type of question: conceptual, factual) mixed-factor ANOVA was conducted on participants' final test performance, with strategy manipulated between-participants and type of question manipulated within-participants (see Figure 1). Overall, participants answered more factual questions correctly (M = 79.28, SE = 1.48) than conceptual questions (M = 58.92, SE = 1.41), $F(1, 120) = 149.10$, $p < .001$. On average, participants in the self-test (M = 67.77, SE = 1.61) and simultaneous (M = 70.44, SE = 1.72) conditions did not differ, $F(1, 120) = 1.28$, $p = .26$. The type of question x strategy interaction

participants in each condition read the ice age passage first and half read the monetary policy passage first.

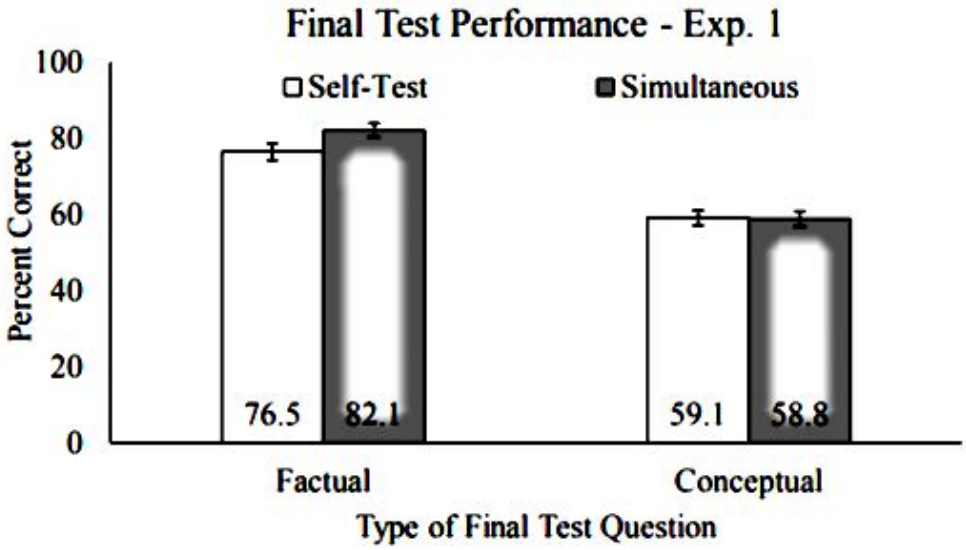
Next, participants in the self-test condition were shown their generated questions over the first passage on the screen (although they did not have access to the passage). They were instructed to answer their questions from memory and were encouraged to guess if they could not remember an answer. Participants in the simultaneous condition saw the first passage again and were instructed to generate questions and answers over the first passage. Instructions were the same as the self-test condition, except that participants in the simultaneous condition were told to type their questions with the answers. All participants had 7 minutes for these activities. After this, all participants saw their questions, answers, and the first passage for 5 minutes. During this time, participants were instructed to self-score their questions using the passage as a means of feedback (Agarwal et al., 2008). They were asked to type "C" for each question they believe they answered correctly and "I" for each question they believe they answered incorrectly. After scoring their questions for the first passage, participants in the self-test condition answered the questions they generated over the second passage while participants in the simultaneous condition generated questions and answers for the second passage for 7 minutes. After that, both groups had 5 minutes to score their answers.

Participants were then told they would take a ten-question multiple-choice test over each passage about 5 minutes later. They were asked to predict how many of the ten final test questions they would answer correctly for each passage (providing a global JOL)[2]. After then completing a 3-minute distraction phase of solving math problems, participants received a final

[2]Participants provided JOLs after completing the study activities for both passages. Thus, more time had elapsed between their study activities and JOL for the first passage they read than the second. This was corrected in Experiment 2.

test over the first passage they read. Questions were presented in a unique random order for each participant, and the final test was self-paced. All participants then completed a self-paced final test over the second passage. Since the ice age and monetary policy passages have been used in other Colorado State experiments, one final question asked participants if they had seen or read these passages before. Participants were then debriefed and released.

also did not reach conventional significance, $F(1, 120) = 3.18$, $p = .07$. However, planned comparisons were conducted to compare performance between the two study strategies for factual and conceptual questions separately. A follow-up t-test showed that, for factual questions, participants in the simultaneous condition numerically outperformed those in the self-test condition, although this difference was not significant, $t(120) = 1.90$, $p = .06$, $d = 0.35$. For conceptual questions, those in the simultaneous and self-test conditions did not differ, $t(120) = 0.11$, $p = .91$, $d = 0.02$.



Italicized graph name

Figure 1. Percent of factual and conceptual questions answered correctly on the final test for those in the self-test and simultaneous conditions. Error bars represent one standard error of the mean.

Figure caption