A Comparison of Study Aids on Exam Performance

Michele C. Baranczyk
Department of Psychology
Kutztown University
USA

Matthew J. Gray
Department of Psychology
Saint Joseph's University
USA

Abstract

Data were collected from 453 students enrolled in General Psychology courses over two semesters to evaluate the effectiveness of textbook pedagogical aids. Students who volunteered for a study skills intervention were randomly assigned to either a study skills intervention containing a packet of paper-and-pencil worksheets and online quizzes; or a computer program condition requiring meta-cognition. Performance data via class exams were recorded for both experimental groups. Performance data from students who did not participate in any intervention served as a control group. Longitudinal results indicate that the Packet condition showed an increase in exam scores after the intervention, while the computer Metacognitive Program and control group showed a decrease in exam scores after the intervention. In addition, the Metacognitive Program condition resulted in higher self-reports of program utility compared to the paper-and-pencil condition. Implications of this research suggest that relatively simple instructor intervention may be effective in improving exam grades.

Keywords: online quizzes, teaching of psychology, student learning

1. A Comparison of Study Aids on Exam Performance

How can instructors of introductory courses best help their students learn new material? Many textbooks now include tools such as online quizzes, flashcards, and other study aids. These tools may offer an avenue for instructors to provide additional aids to students without incurring an increase in workload, such as additional grading. However, the use of these ancillary aids often comes with an increased price in textbook cost. Are these study aids valuable to students, and worth the increased cost? The current study utilizes experimental methodology in a classroom setting to examine the effectiveness of various study aids.

1.1 Textbook Cost and Use

The price of textbooks has become a cause of concern for students and instructors alike. The material historically provided in textbooks has changed, now including more ancillary materials such as websites, online homework and interactive demonstrations (Carbaugh & Ghosh, 2005). These extra materials likely drive up the cost of textbook production as well as consumer price. The question of whether these tools help increase student learning is a valuable research question. Both instructors and students desire ancillary materials, such as website activities and test banks, citing that textbooks need to increase the amount of materials provided (Carbaugh & Ghosh 2005; Landrum & Hormel, 2002).

However, previous research suggests that availability of textbook ancillary packages and computerized test banks was not related to student learning outcomes (Landrum & Hormel, 2002). In other words, while these extra materials are desired by students, they may not serve to increase student learning.

Previous research on how students utilize textbooks has been examined by student surveys reporting how and why they used textbooks in studying (Gurung, 2003; Gurung, 2004). Research on this topic has not found a direct link between textbook characteristics such as pedagogical aids and student performance (Landrum, Gurung & Spann, 2012).

Yet, textbook publishers claim that their products, such as the McGraw Hill Connect product, increase outcomes such as grades (McGraw Hill Connect Effectiveness study, 2013). This specific program is designed as an interactive, meta-cognitive textbook supplement that uses students' confidence as well as correct answer choices in quizzing. However, the McGraw Hill paper is a compilation of case studies, and use of the tool varies by instructor. For some courses, students' use was optional, while for other courses, it appears that the online program was assigned to all students, resulting in potential confounds. One case study (p. 6) indicated that students who used the program scored one letter grade higher than those who did not. However, it appears students were able to self-select into using the tool. Therefore, it is possible that more motivated students chose to use the tool, therefore student motivation may account for difference in student performance. Secondly, it appears as though the options were to use the program, or not, rather than using the program versus an alternative study tool. Would such a program still have benefit if it were compared to an alternative?

1.2 Research in Study Techniques and Aids

Testing is an important aspect to the process of learning in addition to assessment of learning (Roediger & Karpicke, 2006). The testing effect, a phenomenon that occurs when a test is given and subsequently produces better effects on retention, may be useful to bolster student performance. Students may benefit from quizzes preceding an exam because this requires students to recall prior information, which has been established to be more effective than simply reading over the material multiple times (Karpicke, 2012). This body of research indicates that a study intervention of any kind using quizzing strategy should be beneficial to students' performance on exams, though continued study on which specific methods of retrieval are most effective is still needed (Karpicke, 2012). The purpose of the current study is to compare the impact of two types of quizzing materials on students' in-class exam scores.

Indeed, studies have found quizzing to be an effective tool. According to Daniel and Broida (2004), online quizzes were just as effective, but not more effective, as in-class quizzes. This suggests that web-based aids are useful, but do not exceed benefits of in-class aids. Grabe and Sigler (2002) allowed students to voluntarily use online study tools and measured subsequent performance on course exams, finding that study tools increased performance on exams. However, potential confounds exist in each of these studies. The first study (Daniel & Broida, 2004) was a quasi-experimental design in which condition was dependent on course section; it is possible that students in one section would have outperformed students in the second section regardless of type of quiz (in-class or online). In the second study (Grabe and Sigler, 2002), a potential confound is student self-selection; more motivated students may have self-selected into the research. As such, the increased exam performance may have been due to the study tools, student motivation, or a combination of both.

However, not all research has found online quizzes to be effective in improving students' grades. Brothen and Wambach (2001) found that using online quizzes alone is an ineffective study technique. In fact, the authors suggest that computerized quizzes may be distracting, because students merely look up quiz answers, and take quizzes numerous times to get a full score; this tactic diverts students from mastering the information. Additionally, some suggest that the over use of ancillary aids may compromise student's long-term success (Daniel & Poole, 2009). For example, continued use of ancillary aids does not teach students to learn how to develop skills on their own without these programs.

One potential reason for the mixed findings regarding the use of quizzing is that the type of quiz may matter. Gurung, Weidert and Jeske (2010) determined that there are four major categories of study behaviors, one of which are metacognitive abilities (see Gurung et al. (2010) for details of the other categories of application behaviors, elaboration, and resource use behaviors). Metacognitive abilities assess individual self-knowledge about certain material (Gettinger & Seibert, 2002). Metacognition includes students planning a method of learning information, reflecting on the information, and accurately determining whether this information has been learned successfully (Zhao, Wardeska, McGuire, & Cook, 2014).

Metacognitive abilities are essential to studying effectively; the use of metacognitive strategies is related to greater learning results and can be a predictor of academic performance. For example, students who were taught to use metacognition effectively performed better on course exams compared to previous exams (Zhao, Wardeska, McGuire, & Cook, 2014). Metacognition used in conjunction with other methods, such as connecting new information with prior information, also provides beneficial results. Thus, quizzes utilizing meta-cognition may explain why some but not all students show an increase performance.

A second reason for the difference in quiz effectiveness findings may be lab versus field studies. Daniel and Poole (2009) highlight the importance of recognizing differences in lab versus field studies with regard to memory. Some of the previous findings, such as the testing effect, were laboratory-based findings (Roediger & Karpicke, 2006). Ultimately, results found in a lab may not be generalizable to a classroom setting due to differences in motivation or other factors that differ between a lab study and a classroom. Thus, testing lab findings in an applied setting is an important extension.

The current manuscript provides a new approach to studying metacognitive pedagogical aids versus more traditional quiz and practice methods. While previous studies allowed self-selection (Grabe& Sigler, 2002), or assigned particular sections of courses to a certain condition (Daniel & Broida 2004), the current study examines student performance on exams while being assigned one of two programs, and also includes a control group of students not using any extra study aids. Previous research has not addressed the direct comparison of two competing study aids on student performance outcomes using experimental methodology, as suggested in previous research (Weiten, Daniel, Erick, & Sewell, 1999). In addition, this research occurs as a field, rather than lab, study.

Given the previous research on student learning through testing, retrieval, and use of textbook and study aides discussed previously, the current study addresses a comparison of two study methods. Both utilize quizzes and practice questions, thereby including the testing effect. The difference lies in whether the practice is part of an online program that is metacognitive in nature, or if the study program is comprised of more traditional quizzes and homework assignments. Based on previous research findings of the testing effect and retrieval (Daniel & Broida, 2004; Grabe & Siegler, 2002; Karpicke, 2012), it is hypothesized that:

Students self-selecting into the study skills program will have higher exam scores than those who do not participate in the study skills program.

However, it is unknown whether or not a metacognitive program will be more or less effective than a traditional intervention. Therefore, we propose a research question:

Will students assigned to a traditional online quiz condition perform differently than students assigned to a meta-cognitive quiz condition?

2. Methods

2.1 Participants

Data were collected at a state university with an overall enrollment of approximately 10,000 students, located in the Mid-Atlantic United States. Participants in this study were enrolled in the first author's General Psychology class in either the spring or fall of 2012. The majority of students were freshman (68%) followed by sophomores (21%), juniors (8%), and seniors (3%). All students were invited to participate in a study on study strategies. A total of 68 students participated in the study, with 32 randomly assigned to the Packet condition and 36 randomly assigned to the Metacognitive Program condition. Participants who did not complete the materials were excluded from the analyses, resulting in 28 students in the Packet condition and 25 students in the Metacognitive Program condition. Performance data were collected from students in the two experimental groups as well as the remaining students enrolled in class who did not participate in any specific study condition (n=374), providing a control group comparison.

2.2Procedure

In the spring and fall semesters of 2012, the first author had two sections of General Psychology each semester. Institutional Review Board (IRB) approval was obtained prior to the study, and included recording data from all students enrolled in the course.

Throughout the course, performance was assessed through five multiple-choice examinations as a large portion of the course grade. All students enrolled in the course also were required to complete a weekly online quiz through the course management system as part of their course grade. After Exam 3, an intervention was offered. Interested students were offered extra credit (Spring 2012) or research credit as part of their course grade (Fall 2012) for participation. Therefore, Exam 3 measured pre-intervention performance and Exam 4 measured post-intervention performance.

Interested students were told to read, print, and sign informed consent forms found on the course website if they wanted to participate. By requiring students to print and return the informed consent forms (as opposed to an instructor handing out material in class), coercion and/or the potential for students feeling forced to participate was limited.

Interested students met with a research assistant who randomly assigned the student to one of two experimental conditions, thereby controlling for factors such as student motivation. Additionally, use of the research assistant ensured that the instructor was blind to participants and assigned conditions until after the course ended and data analyses began.

Students in both experimental conditions completed activities as instructed by the research assistant, but students in the control group did not perform additional activities beyond class assignments. Then, performance data in the form of exam grades were collected from all students, whether or not they participated in the research study. This resulted in 2 experimental groups and 1 control group: 1) participants randomly assigned to the Packet condition, 2) participants randomly assigned to the Metacognitive Program condition, and 3) students who did not participate in the research study.

2.3 Measures

2.3.1 Independent Variable

The independent variable of study strategy had two conditions: Packet vs. Metacognitive Program. In the Packet condition (n=28), students received a packet of worksheets related to the course content (personality and psychological disorders). Students were instructed to complete the paper and pencil worksheets on their own time and return them to the research assistant prior to Exam 4. The worksheet material was not graded and students did not receive any feedback about their performance. In addition, students in this condition were given five on-line quizzes. The quizzes were multiple-choice, contained between five and fifteen questions, and were completed online through the course management system. Students did receive feedback through percentage correct on the quizzes; the average score across all quizzes was 69.5%.

In the Metacognitive Program condition (n=25), students were given access to an online program that utilized metacognitive components. They were instructed to complete quizzes online through the assigned Metacognitive Program. Five of the quizzes were a traditional quiz in which students selected an answer; these quizzes contained 5-20 questions per quiz. The average score on the quizzes in the Metacognitive Program across all quizzes was 69.0%.The Metacognitive Program also included two metacognitive online quizzes as part of the manipulation. The number of quiz questions varied according to individual performance. If students did not perform well in a particular area (for example, defense mechanisms) additional quiz questions would be given to them until performance improved. The quizzes contained questions that were multiple choice or fill-in-the-blank. Before submitting each question, the student had to respond regarding his or her confidence in choosing the correct answer. For each question, students marked their confidence as follows: "I know it;" "Think so" "Unsure" and "No idea." Asking the student to assess their own knowledge and confidence is a metacognitive strategy, as it requires them to consider their knowledge level. By combining the student's confidence level along with whether the student marked the correct answer, four categories are possible (the average percentage across students is noted in the parentheses). These categories include "Correct and Aware" (31%); "Correct and Unaware" (19%); "Incorrect and Aware" (32%); and "Incorrect and Unaware" (18%). Activities in both the Metacognitive Program and the Packet conditions were designed to take approximately 1-2 hours to complete.

A third group, a control, was comprised of students who did not participate in this research, but were a part of the course (n=374). As a part of the course, students were required to complete nine scheduled online quizzes through the course management system and take five scheduled exams. The performance of these students was tracked in addition to the students in the experimental groups.

2.3.2 Exam Performance

Exam performance was assessed through multiple-choice exams containing 50 questions per exam. Exam 3 was used as the pre-intervention measure, and assessed learning in the topics of development, intelligence, emotion and motivation. Exam 4 was used as the post-intervention measure as the first opportunity after the intervention to assess change. Exam 4 covered the topics of personality theories and psychological disorders and treatment.

2.3.3 Utility Reaction

Utility reaction, or how useful the participants found the study materials, was also assessed for the two experimental groups. In line with Holton, Bates, and Ruona (2000), three items were created for inclusion in a utility reaction scale. The items included the following statements:

1. The individual worksheets and activities were effective in improving my understanding of the material
2. I expect this study program to increase my next exam score compared to my previous exams
3. After using the resource(s) I had a good understanding of the material.

Students rated items on a 1 (*Strongly Agree*) to 5 (*Strongly Disagree*) scale. However, upon creating a scale for the items, internal consistency could not be established. Instead, all three of the items were used individually as dependent variables. While single-item dependent variables are often not desirable, it does provide data regarding the usefulness of the tools according to the student user.

3. Results

3.1 Exam Performance

3.1.1. Hypothesis Testing

To test the hypothesis that students self-selecting into the study skills had higher exam scores than those who did not participate in the study, *t*-tests were conducted to compare post-intervention exam scores of those who participated in either the Packet or the Metacognitive Program (these groups were combined into one group) to students who did not participate in any intervention. Participants with missing exam scores were deleted list wise from all analyses. Results from a one-tail independent measures *t*-test indicated a significant difference between groups for Exam 4 ($t(434) = 1.91$, $p = .028$), the first exam after the intervention,. On average, students who participated in either experimental group (Packet or Metacognitive Program) scored a higher percentage ($M = 68.4$, $SD = 14.5$) than the control group of students who did not participate ($M = 64.4$, $SD = 14.2$). The results indicate that the hypothesis was supported; students participating in one of the experimental groups scored higher than non-participants on Exam 4.

3.1.2 Research Question Analyses

Three analyses are used to analyze the research question asking if students assigned to a traditional online quiz condition performed differently than students assigned to a meta-cognitive quiz condition. First, *t*-tests were used to compare participant scores in the Metacognitive Program and the Packet conditions. Secondly, a chi-square analysis was conducted to examine the frequency of students in each group who increased or decreased in performance measured by Exam 4. Finally, data were analyzed longitudinally with repeated-measures Analysis of Variance (ANOVA) to examine relationships among the conditions.

The results of *t*-tests indicated that there were no significant differences between the Packet condition ($M = 69.5$, $SD = 13.2$) and the Metacognitive Program condition ($M = 67.3$, $SD = 15.8$) on Exam 4 scores ($t(48) = 0.57$, $p = .57$). Thus, participants in both conditions performed comparably to each other on Exam 4 (the first exam after the intervention). This difference was not present for the pre-intervention measure ($t(53) = .40$, $p = .69$), indicating that the groups scored comparably prior to intervention.

Next, a chi-square analysis was utilized to compare the frequency of students in each experimental group increasing performance on Exam 4. The results of the chi-square analysis were not significant ($\chi^2(1) = 3.13$, $p = .077$). In the Metacognitive Program condition, 72% of students received the same or a lower score on Exam 4 compared to Exam 3, leaving 28% of students in the Metacognitive Program scoring higher on Exam 4. In the Packet condition, 48% of students did not improve their Exam 4 score, while 52% of students in the Packet condition showed improvement in score between Exams 3 and 4.

Finally, the research question was tested using repeated measures analyses. Exam3 was used as the pre-intervention measure, and Exam 4 was used as the post-intervention measure.

The results of a 2 (Condition: Packet and Metacognitive Program) x 2(Time: Pre- and Post-Intervention) repeated-Measures Analysis of Variance (ANOVA) indicate no main effect of time ($F(1, 50) = 1.021$, $p = .317$) or condition ($F(1, 50) = .013$, $p = .91$). However, a significant interaction of time and condition was present ($F(1,50) = 7.79$, $p = .007$), such that participants in the Packet condition increased exam score from pre-intervention ($M = 70.3$) to post intervention ($M = 71.9$). Participants in the Metacognitive Program condition decreased exam score from pre-intervention ($M = 72.4$) to post-intervention ($M = 69.0$).

To further address Research Question 1, data were analyzed including the control group of participants. Exam 3 was used as the pre-intervention measure, and Exam 4 was used as the post-intervention measure. Results of a 3 (Condition: Control, Packet, and Metacognitive Program) x 2 (Time: Pre- and Post-intervention) repeated-measures ANOVA indicate no main effect of Condition ($F(2, 424)=.54$, $p=.585$). No main effect of Time was detected ($F(1, 424)=3.55$, $p=.06$). A significant Time x Condition interaction ($F(2, 415)=4.35$, $p=.013$) was detected. Participants in both the control and Metacognitive Program conditions showed a decrease in score from pre- to post-intervention. Students in the Packet condition, however, showed an increase between pre- and post-intervention scores. Figure 1 shows this interaction.

3.2 Utility Reaction

Utility reaction could not be combined into a reliable scale. Instead, individual items were analyzed. There was one significant difference ($t(46)=2.33$, $p=.024$) noted in utility reaction between the Packet and the Metacognitive Program group for the first item: "Activities were effective in increasing my understanding." Students in the Metacognitive Program group reported higher agreement ($M=1.52$) than the packet group ($M=2.12$) on the item. No differences were found in the other two utility analysis items.

4. Discussion

The purpose of the current study was to compare the effect of two types of quiz study strategies on classroom exam performance. In addition, a control group of students was obtained for performance comparison. It was hypothesized that students who self-selected into the study would perform better on the subsequent exam; this hypothesis was supported. Students in either experimental condition scored higher on Exam 4 than the control group after the intervention was introduced. This finding may be due to differences in student motivation. Students choosing to participate in additional class-related activities are likely to have higher motivation than students who choose not to participate in additional activities.

In order to eliminate student motivation as the sole reason for differences in post-intervention exam score, the two experimental groups (Packet and Metacognitive Program) were analyzed as individual conditions in further analyses. Main effects on exam score between the three groups did not exist; there were no statistically significant differences between the three conditions (control, Packet, or Metacognitive Program) on any individual exam. Thus, students in the Packet and the Metacognitive Program condition scored similarly on exams. Perhaps more surprisingly, the students in the control condition also scored similarly to students included in the manipulation. Students who self-selected into the study may be assumed to have higher course motivation, but this is not supported by grade differences before or after the intervention. Rather, exams scores for participants are the same as exam scores for non-participants. As such, neither the Packet nor the Metacognitive Program condition showed a main effect on exam scores when compared to each other, or even when compared to a control group.

However, when analyzed longitudinally, an interaction was present. This interaction was such that students in the Packet condition increased exam performance after the intervention, while both the Metacognitive Program and control group students decreased exam performance after the intervention. This decrease in exam score had been noted in previous semesters, which is why the intervention was used at this time. (The reason for the decrease in scores is unknown.) The Packet condition consisting of worksheets and non-adaptive quizzes appears to have buffered the decrease in exam score. The Metacognitive Program condition did not have this buffering effect on exam scores.

The question of why students in the Metacognitive Program did not perform as well as those in the Packet condition is intriguing. It is possible that this highlights some methodological variables in the literature. As stated previously, applied studies on metacognition use surveys (Gurung, Weidert and Jeske, 2010), or have utilized a metacognition versus no manipulation design (Zhao, Wardeski, McGuire, & Cook, 2014).

By comparing metacognitive studying to another type of studying, the effects do not appear as strong as a metacognitive versus control experimental design.

Despite the exam performance increase in the Packet condition, it was the students in the Metacognitive Program intervention who reported that higher effectiveness of the study aid. In other words, students in the Metacognitive Program condition reported that it was very effective, significantly more so that the Packet condition. However, participants in the Metacognitive Program condition in fact performed worse on the following exam.

This finding is similar to other studies in which student reports of study aid usefulness were not related to performance measures (Balch, 2002; Gurung, 2003; Gurung and Martin, 2011), and suggests that students are not accurate in determining which study strategies are most effective for learning material.

Though textbook publishers present data regarding the effectiveness of interactive tools, the current results suggest that student learning and performance can be increased with a simpler solution. An intervention of paper and pencil worksheets and non-adaptive quizzes was effective in improving exam performance compared to both a Metacognitive Program and a control group of no intervention in large general psychology classrooms. The findings of this research can be applied to general psychology instructors seeking to facilitate their students' learning. By giving out additional practice problems and collecting these problems without providing feedback, exam scores improved. Students should not be discouraged from using pedagogical aids including metacognitive programs if it is included with their textbook, but it is not a necessary component of increasing exam grades. This may be useful information for instructors when choosing required or optional course materials.

This study also adds to the body of literature of the testing effect (Karpicke, 2012) in an applied classroom setting. In these settings, benefits of the testing effect using classroom quizzes have had mixed findings (i.e., Daniel & Broida, 2004; Grabe & Sigler, 2002; Grimstad & Grabe, 2004; Landrum, Gurung, & Spann, 2012). This study suggests that after controlling for student motivation, the metacognitive program was not as effective as a more traditional quiz and homework approach in improving exam grades.

A major strength of the study is the experimental design applied. Students self-selected into the study, but were then randomly assigned to one of two study conditions. Much of the previous research relies on survey (Landrum, Gurung, & Spann, 2012) or allows students to self-select into the study (Daniel & Broida, 2004; Grabe & Sigler, 2002; Grimstad & Grabe, 2004). Random assignment in this study allowed the factor of student motivation to be controlled by the interested students being assigned to one of two conditions. This is an important confound to be controlled in classroom settings.

As a second strength, the data were collected over two different semesters (four total courses) and a large total sample size was obtained, ensuring that the findings are not limited to a single class' experience. Finally, the course instructor was blind to the study during its duration, as research assistants assigned conditions and collected data. These three factors lead to a strong experimental design, supporting the reason for the difference being the intervention rather than other extraneous variables.

In addition to the methodological strength, attention was also given to ecological validity and generalizability. The activities chosen for this study were activities that strongly related to the content area of the course. Additional multiple-choice questions similar to exam questions, and examples related to class content were given. The performance outcome measure was exam and quiz data. Using these outcome measures ensured that students were motivated to perform well, as the scores on these factored into their overall course grade.

A final strength of the study is that a practical concern for instructors is addressed; namely, simple ways to improve student performance in large introductory courses. This intervention has the advantage of being relatively simple to introduce; providing practice worksheets and quizzes that the instructor does not need to grade. This intervention is also low cost to students.

4.1 Limitations

Despite being offered extra credit or research credit as part of a course requirement over the course of two semesters, only 53 students participated in the study to its completion. This is a small number compared to the control group. Secondly, some of the questions and activities in the Packet condition were taken from an ancillary that was related to the required textbook. This approach may have led to more consistency with terms from the course, resulting in more repetition for students in this condition. The content in the Metacognitive Program condition was the same, but phrasing may have varied more than the Packet condition.

However, average quiz scores were remarkably similar between conditions (69.0% in the Metacognitive Program vs. 69.5% in the Packet condition), suggesting similar performance between conditions. As such, both conditions should have helped students retrieve and review concepts contained in the exam.

4.2 Contributions and Conclusion

The current study combined experimental rigor and practical concern. While many previous studies on this topic have been done using surveys and quasi-experimental designs, the advantage of this study is the use of an experimental design in a classroom setting and collecting classroom performance data. The data were collected in four different sections of General Psychology over two semesters, providing evidence that the effect is consistent.

The results of the study suggest a simple way for instructors to help students perform better on exams in large, introductory courses. The transition from high school to college courses can be difficult for some students, and helping students learn study skills early in their college careers may be one way to foster student success.

References

Carbaugh, R., & Ghosh, K. (2005). Are college textbooks priced fairly? *Challenge, 48*(5), 95-112.

Daniel, D. B., & Broida, S. (2004). Using web based quizzing to improve exam performance: Lessons learned. *Teaching of Psychology 31*(3), 207-208. doi:10.1207/s15328023top3103_6

Daniel, D. B., & Poole, D. A. (2009). Learning for life: An ecological approach to pedagogical research. *Perspectives on Psychological Science. 4*(1). 91-96. doi:10.1111/j.1745-6924.2009.01095.x

Gettinger, M., & Seibert, J., K. (2002). Contributions of study skills to academic competence. *School Psychology Review, 31*(3), 350-365.

Grabe, M., & Sigler, E. (2002). Studying online: Evaluation of an online study environment. *Computers & Education, 38*, 375-383. doi:10.1016/s0360-1315(02)00020-9

Grimstad, K., & Grabe, M. (2004). Are online study questions beneficial. *Teaching of Psychology, 31*(2), 2004.doi:10.1207/s15328023top3102_8

Gurung, R. A. R., & Martin, R. C. (2011). Predicting textbook reading: The textbook assessment and usage scale. *Teaching of Psychology, 38*(1), 22-28.doi:10.1177/0098628310390913

Gurung, R. A. R. (2004). Pedagogical aids: Learning enhancers or dangerous detours? *Teaching of Psychology, 31*, 164-166. doi:10.1207/s15328023top3103_1

Gurung, R. A. R. (2003). Pedagogical aids and student performance. *Teaching of Psychology, 30*(2), 92-95.doi:10.1207/s15328023top3002_01

Gurung, R. A. R., Weidert, J. & Jeske, A. (2010). Focusing on how students study. *Journal of the Scholarship of Teaching and Learning, 10*(1), 28-35.

Gurung, R.A.R., Daniel, D. B. & Landrum, R. E. (2012). A multisite study of learning in introductory psychology courses. *Teaching of Psychology, 39*(3), 70-75. doi:10.1177/0098628312450428

Holton, E.F., Bates, R.A., & Ruona, W.E.A. (2000). Development of a generalized learning transfer system inventory. *Human Resource Development Quarterly, 11(*4), 333-360. doi:10.1002/1532-1096(200024)11:4<333::AID-HRDQ2>3.0.CO;2-P

Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*(3), 157-163. doi:10.1177/0963721412443552

Landrum, E. R., Gurung, R. A. R., & Spann, N. (2012). Assessments of textbook usage and the relationship to student course performance. *College Teaching. 60*, 17-24. doi:10.1080/87567555.2011.609573

Landrum, E. R., Hormel, L. (2002). Textbook selection: Balance between the pedagogy, the publisher, and the student. *Teaching of Psychology, 29*(3), 245-248.

McGraw Hill Education Connect Effectiveness study (2013). Evaluating the impact on student engagement, test scores, course grades, attendance and retention rates at 34 U.S. higher education institutions. http://connect.customer.mcgraw-hill.com/wp content/uploads/2013/08/MHE_ConnectEffectivenessStudy_2013.pdf

Roediger, H. L. III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long term retention. *Current Directions in Psychological Science, 17*(3), 249-255. doi:10.1111/j.1467-9280.2006.01693.x

Wambach, C.,& Brothen, T. (2001). The relationship of conscientiousness to metacognitive study strategy use by developmental students. *Research and Teaching in Developmental Education*, 18(1), 25-31.

Weiten, W., Daniel, D., Erika, R., & Sewell L. (1999).University, community college, and high school students' evaluations of textbook pedagogical aids. *Teaching of Psychology 26*(1), 19-21. doi:10.1207/s15328023top2601_3

Zhao, N., Wardeska, J.G., McGuire, S.Y., and Cook, E. (2014). Metacognition: An effective tool to promote success in college science learning. *Journal of College Science Teaching, 43* (4), 48-53.

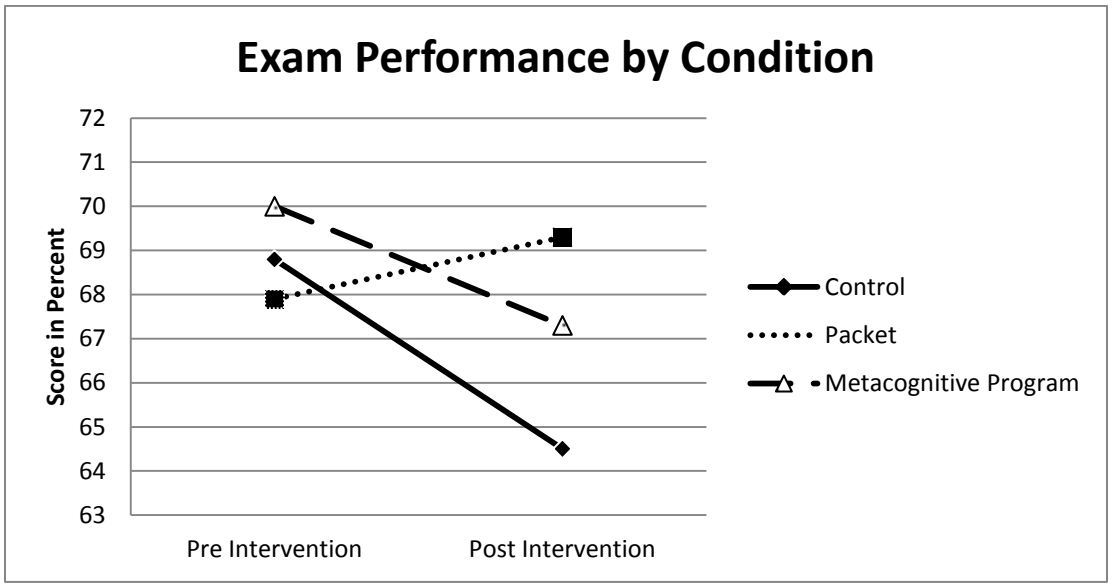

Figure 1: Study Condition by Exam Interaction. Mean exam score in Percent by Condition and Time.