Journal of Effective Teaching in Higher Education

Comparison of Two Online Courseware Instructional Methods Using Propensity Score Matching

Renée L. Babcock, Central Michigan University, babco1rl@cmich.edu
Jean Georgiou, New Jersey City University

Abstract. This study used propensity score matching to compare two online courseware systems intended for use as supplements to either face-to-face or online courses. One courseware system focused on adaptive features available to students while completing assignments; whereas the other system was an interactive courseware that provided embedded media supplements and assessments. Participants were students enrolled in a psychology course taught by an instructor who used the adaptive courseware in Fall 2013, switching to the interactive courseware in Fall 2015 and 2016. Propensity score matching was used to match students in 2015 with students in 2013. A separate matching process was done for the 2016 versus 2013 students. The use of propensity score matching was successful in allowing comparison across groups. However, the only consistent finding across the two matched samples was that students appeared to spend less time in the interactive courseware than the adaptive courseware. The most likely reason for the difference is a function of the requirements of the two courseware systems.

Keywords: Propensity Score Matching, digital learning systems, adaptive courseware, interactive courseware

A common assumption about millennials is that they thrive in a digital world that allows them to instantly interact with endless amounts of material at their fingertips. This assumption is not unfounded: Pew Research Center found that 97% of millennials (ages 22-37) use the internet regularly (Jiang, 2018). Capitalizing on that information, the use of online learning materials to improve student learning has been increasingly adopted by college instructors, as publishers compete to offer digital supplements with their textbooks. Pearson Publishers, McGraw-Hill Education, Wiley, Cengage, and others offer access to digital products intended to draw students more deeply into the material and offer various strategies to do so. Such web-based systems can enhance instructor-student interactions by providing additional resources for the student to explore in-depth those topics that can typically be explored only at a broader level in classroom settings. Understanding whether these courseware systems are effective in learning is important for instructors in decisions on the amount and type of material to assign. That is, knowing whether the digital learning systems are efficacious in knowledge acquisition is important for instructors as they choose whether or not to adopt them for use in the classroom. One method of determining efficacy is to use propensity score matching when comparing systems. As noted, there are a range of types of programs available, with two common strategies being adaptive learning and interactive programs that provide multiple types of material all within one source. The current study used propensity score matching to compare two such programs offered by Pearson Publishers.

Online Adaptive Courseware

The core learning design element of this courseware is its adaptive function that includes learning activities and features designed to measure learners' strengths and weaknesses on topics and guide them to study the topics that require practice. That is, this type of courseware provides additional material and quizzing on topics that the student has not yet mastered and limits the presentation of material that the student has already mastered. The goal of such systems is to optimize a student's study time and effort. The online adaptive courseware was designed based on research in learning principles. The key adaptive features include scaffolding, feedback, and a reduction in cognitive load.

One of the components of an online adaptive courseware is an instructional approach known as scaffolding. Often useful in assisting in the construct of appropriate levels of cognitive load, scaffolding offers assistance or support to the learner as needed (Bunch, 2009; Sharma & Hannafin; 2007; Van Merrienboer, Kirschner, & Kester; 2003). However, once the goal is achieved, or the learner no longer needs the assistance, the support disappears (Bunch, 2009). While utilizing different scaffolding approaches, several researchers noted positive findings based on their achievement of complex cognitive skills across a variety of domains reducing the overall cognitive load (Bunch, 2009; Gerjets, Scheiter, & Catrambone; 2004; Ong & Tasir, 2015; Xie et al., 2018; Yaman, Nerdel, & Bayrhuber; 2008). The adaptive courseware here provides students with scaffolding - or technology-based support - on specific learning tasks, such as concept mastery through practice quizzing. The adaptive features embedded in the courseware were based on recommendations from research on scaffolding in educational technology (Bunch, 2009; Sharma & Hannafin, 2007).

In this online adaptive courseware, students complete a pretest at the beginning of each unit to determine what they have already mastered and what they have not yet mastered. Students are then provided with feedback that suggests where they most need to focus their attention when reviewing the course material. Suggestions might include reviewing sections of the text, viewing a video, completing a simulation, or testing on concepts. After students have completed the suggestions, they complete a post-test to determine if they have mastered the material. After the post-test, students again receive feedback on what they may want to continue studying. Research suggests that feedback from the pre- and post-tests may boost students' long-term retention and confidence (Hattie, 2009, 2012). One of the assessments that Reddy, Labutov, and Joachims (2016) utilized in their study provided learners with an opportunity to improve their skills over time by requiring a prerequisite. When a learner successfully completed a module and was assessed with a pass, he/she was then able to move forward, hence receiving positive feedback. In addition to using pre and post-tests, Ong and Tasir (2015) provided those in their study with a cognitive load measurement for self-reporting. This self-reporting scale measured the users invested mental effort. The scale ranged from "very, very low mental effort" to "very, very high mental effort" (p. 508). While this instrument may have been chosen for its reliability, simplicity, and practicality, it helped the learner as well as the researchers receive valuable feedback regarding the learner's overall invested mental effort while their cognitive load was being measured.

In cognitive psychology, cognitive load refers to the total amount of mental effort being used in working memory. This includes extraneous cognitive load (i.e., mental effort spent on distracting elements that are not relevant to the learning). Research shows that if one can reduce extraneous cognitive load for students when they are reading or studying, one can improve the effectiveness of the students' ability to process the important information and move it from working memory to long term memory (Pociask, DiZazzo-Miller, & Samuel, 2013; Miller, 1956; Sweller, 1988). Ong and Tasir (2015) discuss the importance of incorporating certain aspects of cognitive psychology and Cognitive Load Theory to reduce the overall cognitive load and the beneficial effects it has on "training time, improved performance on tasks using learned knowledge/skills, and improved performance on other tasks where the learned information can be applied" (p. 503). When properly accomplished the learner's cognitive resources are released perhaps freeing them for another learning experience.

Online Interactive Courseware

The online interactive courseware was created to provide a more engaging experience for the students. As students read through the text, they encounter several opportunities to deepen their learning by watching videos, manipulating points on graphs, and taking brief quizzes. A benefit of this digital product was that everything was contained within a single source: the e-text, interactive charts/graphs, videos, and assessments. Unlike the adaptive learning system described above, students do not have to leave the e-text to study as everything is contained with the e-text. Additionally, this courseware does not change or adapt based on a student's performance. The interactive courseware was designed utilizing research-based design principles intended to help students learn. The key learning design principles included reduced cognitive load, multi-media presentation of content, and embedded assessments.

Similar to the adaptive courseware, the interactive courseware attempts to lessen extraneous cognitive load by clearly organizing materials into topics and subtopics. Each chapter is divided into modules that cover a small amount of material. At the close of each subtopic, students complete a brief (3-5 question) quiz, thus minimizing the amount of material that needs to be mastered at any given point. Overall, this interactive courseware provides students with a centralized learning environment with easy access to assignment instructions and other resources needed to complete any assignment.

Mayer and Moreno (2003) define meaningful learning as "deep understanding of the material, which includes attending to important aspects of the presented material, mentally organizing it into a coherent cognitive structure, and integrating it with relevant existing knowledge" (p. 43). When material is delivered through multimedia technology the benefits may be as limitless as the technology itself.

This may be due in part to the similarities of multimedia and a person's innate way of learning or their cognitive processing as it is often referred to (Jereb & Smitek, 2006; Mayer, 2002; Ong & Tasir, 2015; Xie et al., 2018). Xie et al. (2018) were able to show positive results using the coordination of visual and auditory cueing in multimedia learning. Their results indicated that by incorporating dual modality cues (visual and auditory) to the online lessons, the learners were able to achieve an increase in overall retention span with online learning in addition to a more time efficient learning process (Xie et al., 2018). The online interactive courseware utilized in this study attempts to follow research-based multimedia learning principles so that students can apply new learning and assessment strategies in ways that would not be possible with a printed textbook (Mayer & Moreno, 2003). Specifically, the interactive courseware contains narrative text that is combined with interactive elements. Interactive media provides many more options for how to present information and ideas than static text, which in turn leads to opportunities to more clearly present those ideas and information. Often when working with interactive media, students can choose to pause what they are reading so that they have the opportunity to stop and process information, which is intended to help them connect the media to the text and build a richer knowledge base (Craik & Tulving, 1975; Mayer, 2002; Virk, Clark, & Sengupta, 2015).

The interactive courseware contains embedded assessments, the purpose of which is to allow students to check their understanding and receive immediate feedback. Jereb and Smitek (2006) stress the importance of using self-assessment instruments and recommend creating three assessments: One to be used in the beginning, one for the middle, and one for the end of the learning material so that the learner can gain the most from their multimedia learning experience (Jereb & Smitek). Reddy et al. (2016) utilized several embedding techniques in their study, one of which is pass/fail assessments that allow the learner to improve by working on modules over a period. Additional research suggests that such embedded assessments and the feedback they provide can help build long-term retention and increase students' confidence and motivation (Hattie, 2009, 2012). The spacing of assessment opportunities is aligned with research findings on how to optimize memory (e.g., Clark & Bjork, 2014).

In general, both online courseware systems are similar in their learning design principles. The main difference between them is that one utilizes an adaptive learning approach whereas the other utilizes an interactive learning approach. The objective of this study is to use propensity score matching to examine whether there are differences in students' learning outcomes and behaviors in an introductory psychology course when utilizing the two types of online courseware. Thus, this study investigates whether students do better if they prepare using interactive tools or if they have adaptive features to help them with their assignments.

Methods

Participants

This study took place at a public university in the Midwest. The instructor in the study taught an introductory psychology course during Fall semesters. In Fall 2013, the instructor used the online adaptive courseware with 200 students. Then in Fall 2015, the instructor switched to the interactive courseware with 174 students and continued to use the interactive courseware in Fall 2016, with 191 students.

Measures

The two online supplemental courseware systems described above were utilized in the study. For each, there are data that come from the online courseware system as well as data from performance in the course.

System log data for the interactive courseware and the adaptive courseware were extracted. These data came from students' use of the courseware and included variables on the average percent correct that the students achieved in the assignments within the courseware, the total time spent in the courseware, and the total items attempted in the courseware. These system log data variables were the outcomes that were examined after matching. The system log data were merged with the exam grades the students obtained in the course. These data were then merged with data on students' baseline characteristics that are described below and that were used in the matching. In addition to the system log data variables that were examined as outcomes, the average grade across all exams was also examined as an outcome.

To account for differences between students who were enrolled in the different semesters, this study used propensity score matching. Propensity score methodology is a recommended method when the goal is to estimate causal effects using observational (nonrandomized) data because it mimics some of the particular characteristics of a randomized study (Rosenbaum & Rubin, 1984), which is aligned to the research design of this study.

The matching algorithm used was one-to-one matching with replacement, that is, each student in the interactive courseware class was matched to one adaptive courseware student, where the adaptive courseware students could be matched more than once, if they were the closest match to the interactive courseware student. Unlike normal matching, in which participants are typically matched on a single characteristic, propensity score matching utilizes a predicted probability of group membership based on observed predictors. Thus, propensity scores are used to reduce selection bias by equating groups when randomization isn't possible.

The baseline characteristics that were being matched on include ACT composite score, gender, minority, freshman standing, and cumulative GPA at the beginning of the semester. With ACT composite and cumulative GPA, differences in prior achievement (an important covariate to consider) were addressed in the study.

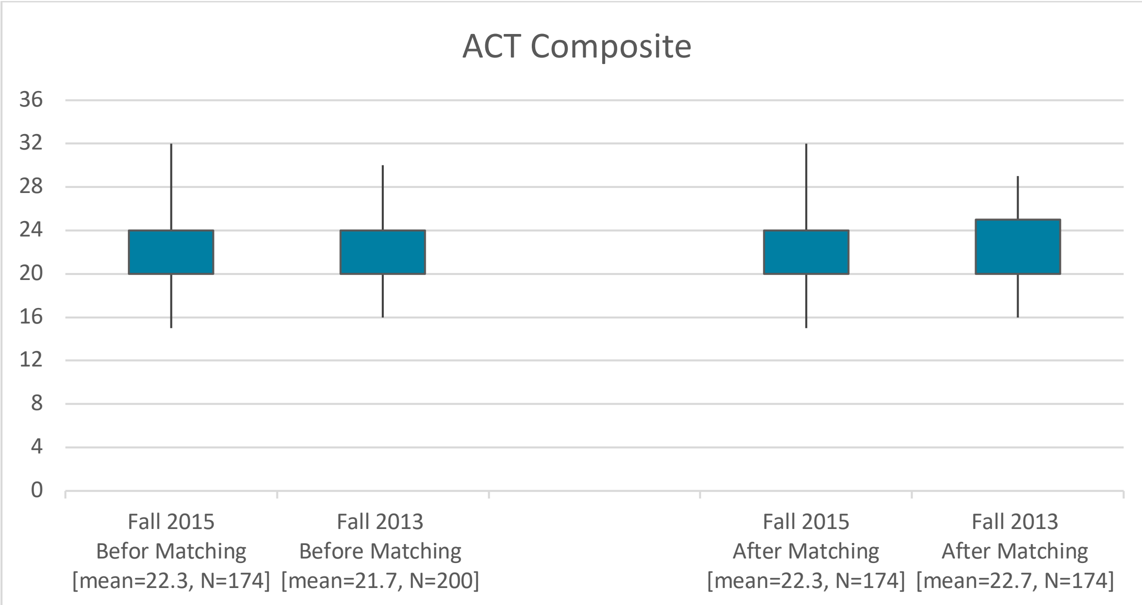
Procedure

Students in the semester utilizing the adaptive courseware were required to complete pre-tests for each chapter of the book. The pre-tests were used to provide an individualized study plan for each student. Students were encouraged, but not required to complete all of the assignments suggested within their unique plan. At the end of each chapter, students were then required to complete a post-test for the chapter. The post-tests were presented with a mastery approach such that the highest score achieved on the post-test was the score recorded for their grade. Students in the semesters utilizing the interactive courseware were required to complete brief quizzes at the end of each module of a chapter and at the end of each chapter. Scoring of the quizzes encouraged students to interact with the material prior to responding to each question. That is, if the student chose the correct response to the question on the first attempt, three points were earned. If the student needed two attempts, then two points were earned; and, if the student needed three attempts, only one point was earned. Thus, students were encouraged to look back through the material prior to responding to each quiz question. In all three semesters, students were required to complete either five or six high-stakes exams in the classroom.

Results

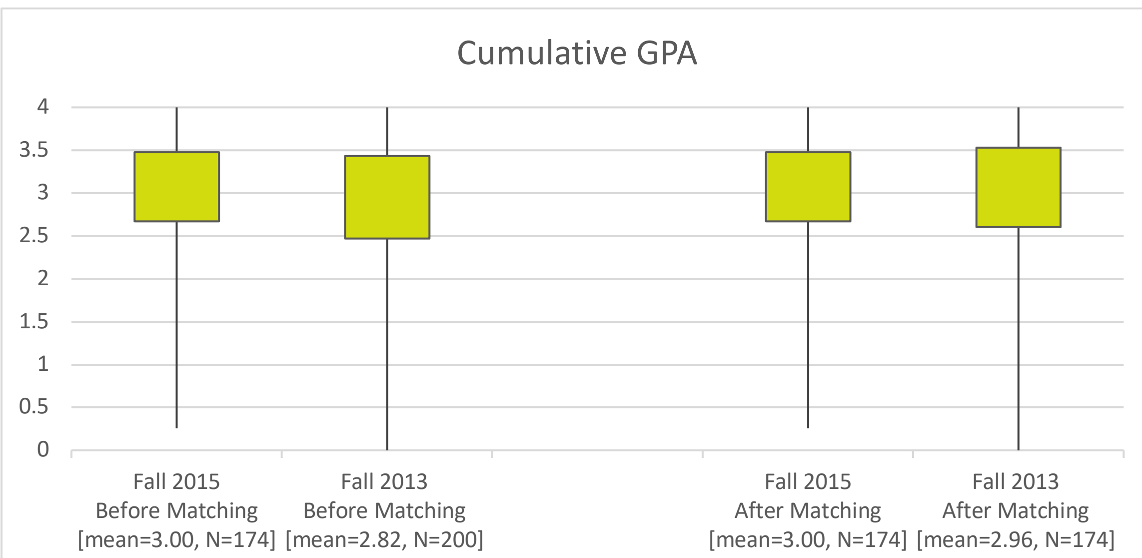
This study compared students using the interactive courseware to the students using the adaptive courseware using propensity score matching. Two comparisons were made. First the Fall 2015 interactive courseware class was compared to the Fall 2013 adaptive courseware class. The second comparison was between the Fall 2016 interactive class and the Fall 2013 adaptive class.

Figures 1a to 2c show the descriptive statistics of the samples before and after matching. Overall, for the Fall 2013 (adaptive courseware) and Fall 2015 (interactive courseware) samples, the descriptive statistics for ACT Composite, cumulative GPA, female, non-white, and freshman showed that the one-to-one matching did a good job and the differences between the two groups were smaller after matching, except for only one demographic characteristic—non-white.



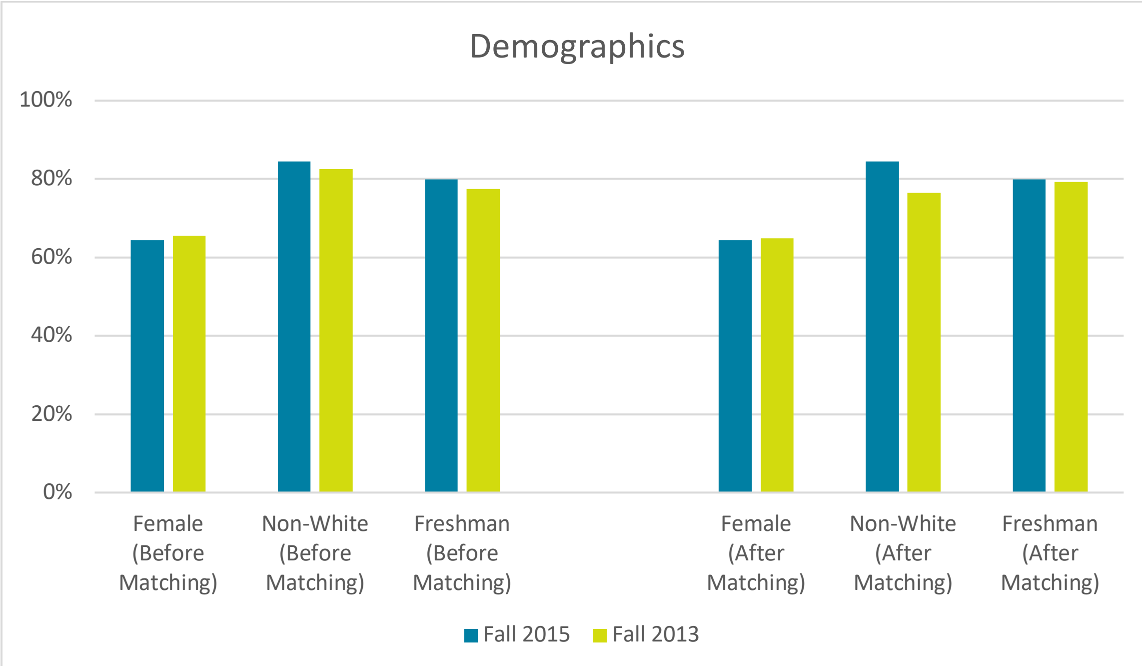
Note: After matching, the outliers are reduced in the adaptive courseware group (Fall 2013) and the mean is closer to the interactive courseware (Fall 2015) mean.

Figure 1a: ACT Composite Descriptive Statistics for the Adaptive Courseware Group (Fall 2013) and the Interactive Courseware Group (Fall 2015)



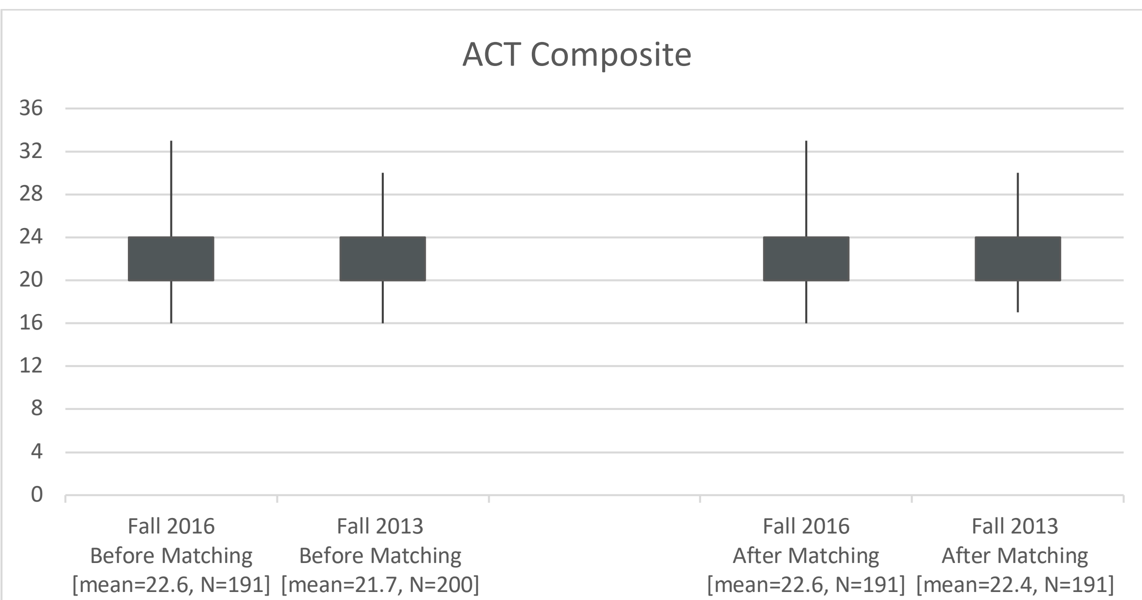
Note: After matching, the interquartile range and mean in the adaptive courseware group (Fall 2013) are closer to the interactive courseware group (Fall 2015).

Figure 1b. Cumulative GPA Descriptive Statistics for the Adaptive Courseware Group (Fall 2013) and the Interactive Courseware Group (Fall 2015)



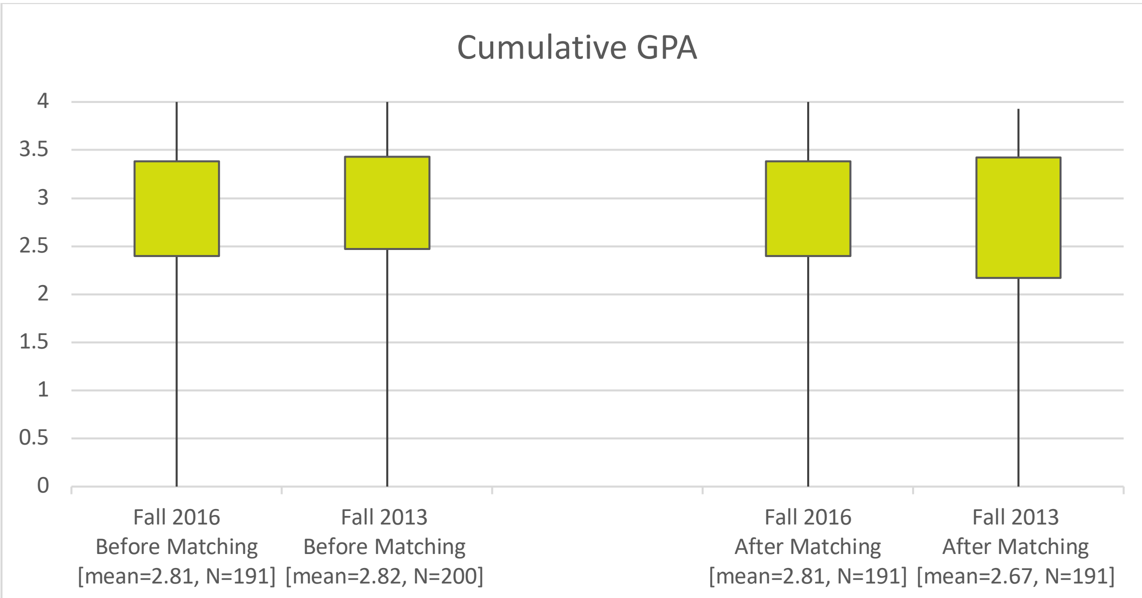
Note: The percentages between the two groups are closer after matching except for non-white.

Figure 1c: Demographic Descriptive Statistics for the Adaptive Courseware Group (Fall 2013) and the Interactive Courseware Group (Fall 2015)



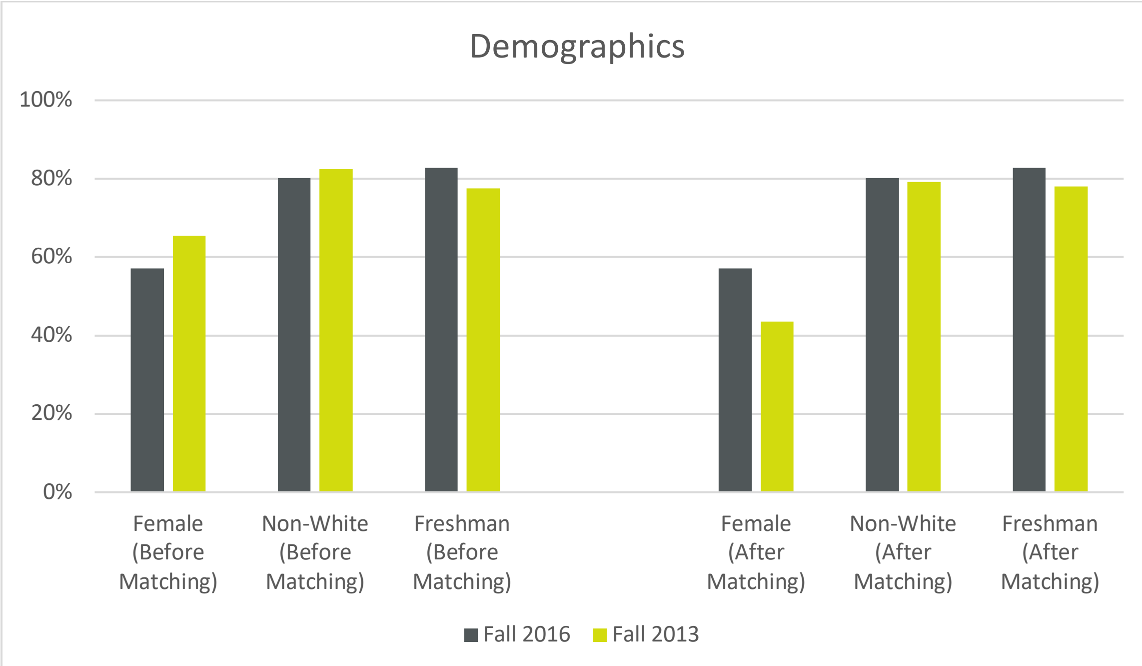
Note: After matching, the adaptive courseware group (Fall 2013) mean is closer to the interactive courseware group (Fall 2016) mean.

Figure 2a: ACT Composite Descriptive Statistics for the Adaptive Courseware Group (Fall 2013) and the Interactive Courseware Group (Fall 2016)



Note: The distributions of the cumulative GPA between the two groups are almost identical before matching. Hence, it is not surprising that, after matching and removing some of the unmatched students in the adaptive courseware group (Fall 2013), the distribution is a little more disperse, though there is still a high degree of overlap after matching. Thus, balance is not an issue here.

Figure 2b: Cumulative GPA Descriptive Statistics for the Adaptive Courseware Group (Fall 2013) and the Interactive Courseware Group (Fall 2016)



Note: The percentages between the two groups are closer after matching except for female.

Figure 2c: Demographic Descriptive Statistics for the Adaptive Courseware Group (Fall 2013) and the Interactive Courseware Group (Fall 2016)

Similarly, for the Fall 2013 (adaptive courseware) and Fall 2016 (interactive courseware) samples, the descriptive statistics across ACT Composite, cumulative GPA, female, non-white, and freshman showed that the one-to-one matching did a decent job and there is balance (i.e. good amount of overlap) between the two groups after matching, except for one demographic characteristic—female.

The descriptive statistics presented before and after matching only provide a visual examination of the differences between the two groups. Such visual examination involves multiple comparisons and considers each characteristic equally important. For a more rigorous analysis of balance between the two groups, Rubin's (2001) criteria were also used:

1. The difference in the mean propensity score in the two groups should be near zero. That is the difference in the mean of the propensity score should be less than half a standard deviation.
2. The ratio of the variance of the propensity score in the two groups should be near one, where less than 0.5 or larger than 2 are considered extremes.
3. The ratio of the variances of the (continuous) covariates after adjusting for the propensity scores should be close to 1. Ratios between 0.80 and 1.25 are acceptable. Ratios less than 0.50 or greater than 2.0 are considered extremes.

When Rubin's (2001) balance diagnostics was performed across both the Fall 2013 versus Fall 2015 and Fall 2013 versus Fall 2016 samples, there were eight diagnostics comparisons performed in total, with seven of the eight comparisons meeting the acceptable range, and only one comparison being out of the acceptable range (that is, out of the acceptable range of 0.8 to 1.25 for the ratio of variances). However, it is still not in the extremes (which is less than 0.5 or greater than 2). Therefore, we can conclude that the matched samples were relatively balanced after propensity score matching and can proceed with the analysis.

After matching, paired *t*-tests were conducted for the matched pairs separately for the Fall 2015 (versus Fall 2013) sample and the Fall 2016 (versus Fall 2013) sample. Table 1 shows the results, which were mixed. In most cases significant findings for Fall 2015 were not replicated in Fall 2016. For example, students received a significantly lower exam average score in Fall 2015 compared to Fall 2013 whereas the difference in exam average score between the Fall 2016 and the Fall 2013 samples was not significant. It could be that the instructor was learning and adjusting to a new teaching format, which could account for the significantly lower exam average in Fall 2015.

Table 1a

Paired t-test Results after One-to-One Propensity Score Matching: Fall 2015 and Fall 2013 Matched Sample

Outcome	Paired Mean Difference	Sample Size	*p*-value
Exam Average	-10.53	174	<.0001
System Log Data Variables			
Average Percent Correct	7.97	171	<.0001
Total Time (hours)	-6.55	174	<.0001
Total Items Attempted	-114.3	174	.0415

Table 1b

Paired t-test Results after One-to-One Propensity Score Matching: Fall 2016 and Fall 2013 Matched Sample

Outcome	Paired Mean Difference	Sample Size	*p*-value
Exam Average	-3.65	191	.0772
System Log Data Variables			
Average Percent Correct	-0.11	186	.9163
Total Time (hours)	-4.73	191	<.0001
Total Items Attempted	119.0	191	.0748

Notes for Tables 1a and 1b:

1. The system log data variable, average percent correct, for Fall 2013 (adaptive courseware) was calculated based on the scores that each student

received for post-tests averaged across all the post-test assessments that the student took.
2. Similarly, the average percent correct for Fall 2015 and Fall 2016 (interactive courseware) was calculated based on the scores that each student received for assessments average across all the assessments that the student took.
3. A negative difference indicates that overall, the Fall 2013 (adaptive courseware) class had higher outcome values, while a positive difference indicates that overall, the Fall 2015 or Fall 2016 (interactive courseware) class is had higher outcome values.

The only consistent finding across the two matched samples is that students appeared to spend less time in the interactive courseware than the adaptive courseware. There are at least two possible reasons for this result. The first is that in the adaptive program, once students complete the pre-test, they use links to other material that could assist in their learning. Doing so would ultimately result in longer amounts of time spent within the courseware. Another possible reason is that the learning design principle of the interactive courseware emphasized "read a little, do a little," thus resulting in shortened times within courseware. A limitation of this analysis is that the data we have on time spent did not differentiate between the time when students were actively logged into the courseware versus the time when they were idle while logged in. Hence, having a significant finding on time spent might not be indicative of time actively engaged with the courseware.

Discussion

Educational technology has opened up new ways for students to learn in class that are not possible with a print text. Assessing whether supplemental courseware systems are effective in improving learning is challenging. One of the goals of the current study was to use propensity score matching to compare two online courseware systems: one in which adaptive features were used to optimize studying and a second in which interactive features were embedded within an e-text with the intent on making the material more engaging. The use of propensity score matching enables more accurate comparisons to be made. A second goal of this research was to determine if students do better when they prepare using interactive tools or if they do better when using adaptive features to help them with their assignments. The current study did not seem to show clear cut evidence that one approach is better than the other. Some of the mixed results could be due to the fact that the instructor was using the interactive courseware for the first time in Fall 2015 and so, was still adjusting to it. Another possible explanation is that students' individual learning styles influenced how they interacted with the courseware; and, in turn, affected their learning and achievement.

Though differences in the two systems were not apparent, it is important to note that either or both of the courseware systems might be effective in improving learning over and above that which normally takes place within traditional course instruction. That is, the current study did not include a control group that did not use any digital learning software. Anecdotally, the instructor of the course noted that overall student performance improved with the use of either system compared to when no system was used (e.g., overall failure rates were reduced from 10-12% to fewer than 7%). There are several important messages that can be gleaned from this study. First, and perhaps foremost, we are reminded that not all students learn in the same manner. There is no ideal way of teaching whether in the classroom or on a multimedia course with specific courseware. Bikowski and Casal (2018) suggested that digital materials are most effective when they are customized, interactive, and useable for both faculty and students. Though one of the courseware products studied in this project was considered "interactive", Bikowski and Casal (2018) were more likely referring to the interactive nature of e-texts in that they can be used to present videos, demonstrations, and simulations. It remains possible that both of the courseware programs improve learning.

Though neither online courseware system seemed better than the other, the importance of using propensity score matching to examine or compare efficacy of a program should not be overlooked. The educational system lends itself to the study of non-randomized groups and propensity scoring enables comparison of those groups. This is especially true if the two groups do not have much overlap.

There are limitations to the current study that could be addressed in future research. One limitation is that there was not a control group that did not utilize courseware. Because of this, interpretation of some of the results needed to rely on anecdotal evidence. Future research is needed to confirm that any digital learning system is better than none at all. Another limitation is that because the two courseware systems were designed by publishers for pedagogical rather than empirical use, it is impossible to isolate and control for variables that differ in the adaptive and interactive courseware systems. This could be improved with further study regarding student engagement with each of the tools. Furthermore, the use of qualitative interviews with the students and professor would help answer some of the questions regarding why the students were online more in the adaptive courseware, and how the instructor's approach changed in the Fall 2015 class. Finally, it would be useful to know which type of system fits best for different learning styles. Though it is not practical to think that each student could select her/his own system, different learning institutions may determine that the majority of their students would benefit from a particular type of courseware. Such information may also be informative for textbook publishers as they continue to enhance their programs.

Technology is taking teaching and learning to new heights, once never imagined. The tools for teaching must be enhanced to teach the learners of today. As technology grows so must the teaching forum. Technology is not only opening doors, but is also giving educators an opportunity to enhance the learning experience. What is done with technology may shape how the educational arena of tomorrow is viewed. The findings of this study are valuable to the educator who wishes to adapt some type of courseware but has been hesitant to do so or has not known where to start. The authors of the current study have anecdotal evidence that regardless of which system one uses, overall performance in the class will improve over using no courseware system at all. Additionally, students independently commented that they found the systems beneficial to their learning.

Perhaps the courseware chosen is less important than simply providing students with an additional study tool. The question may no longer be whether or not to use online courseware, but rather which courseware to use.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

Bikowski, D. & Casal, E. (2018). Interactive digital textbooks and engagement: A learning strategies framework, *Pennsylvania State University Language Learning & Technology 22* (1), 119–136.

Bunch, J.M. (2009). An approach to reducing cognitive load in the teaching of introductory database concepts. *Journal of Information Systems Education, 20*(3), 269-275.

Clark, C. M. & Bjork, R. A. (2014). When and why introducing difficulties and errors can enhance instruction. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.). A*pplying science of learning in education: Infusing psychological science into the curriculum*. Retrieved from the Society for the Teaching of Psychology website: http://teachpsych.org/ebooks/asle2014/index.php

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294.

Ellman, M.S., Schwartz, M.L. (2016). Online Learning Tools as Supplements for Basic and Clinical Science Education, *Journal of Medical Education and Curricular Development, 3,* 109-114.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*(1-2)*,* 33-58.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hattie. J. (2012). *Visible learning for teachers: Maximizing impact on learning.* New York, NY: Routledge.

Jereb, E., & Šmitek, B. (2006). Applying multimedia instruction in e learning. *Innovations in Education and Teaching International, 43*(1), 15–27. https://draweb.njcu.edu:2103/10.1080/14703290500467335

Jiang, J. (2018). Millennials stand out for their technology use. Pew Research Center, http://www.pewresearch.org/fact-tank/2018/05/02/millennials-stand-out-for-their-technology-use-but-older-generations-also-embrace-digital-life/

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice, 41*(4), 226-232.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63(*2*)*, 81-97.

Ong, C. P. & Tasir, Z. (2015). Self-instructional module based on cognitive load theory: A study on information retention among trainee teachers. *Educational Technology Research and Development, 63*(4), 499-515.

Pociask, F.D., DiZazzo-Miller, R., & Samuel, P.S. (2013). Reducing cognitive load while teaching complex instruction to occupational therapy students. *The American Journal of Occupational Therapy, 67*(5), 92-98.

Reddy, S., Labutov, I. & Joachims, T. (2016). Latent skill embedding for personalized lesson sequence recommendation. *Under review by the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using sub-classification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516-524.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2*(3-4), 169-188.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology--enhanced learning environments. *Interactive Learning Environments, 15*(1), 27-46.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Van Merrienboer, J. J. G., Kirschner, P.A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5-13.

Virk, S., Clark, D. & Sengupta, P. (2015). Digital games as multirepresentational environments for science learning: Implications for theory, research, and design. *Educational Psychology, 50*(4), 284-312.

Xie, H., Mayer, R. E., Wang, F., & Zhou, Z. (2018, July 2). Coordinating visual and auditory cueing in multimedia learning. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000285

Yaman, M., Nerdel, C., & Bayrhuber, H. (2008). The effects of instructional support and learner interests when learning using computer simulations. *Computers & Education, 51*(4)*,* 1784-1794.