Teaching and Iterative Improvement: The Impact of Instructor Implementation of Courseware on Student Outcomes

Martha Hubertz, University of Central Florida, United States
Rachel Van Campenhout, VitalSource, United States

The IAFOR International Conference on Education in Hawaii 2022
Official Conference Proceedings

Abstract
Students need digital learning resources that will benefit their learning process, and educators need tools that provide meaningful data insights and can be integrated into their teaching practice. Courseware as a learning resource is designed based on learning science principles to optimize the learning process for students. However, it is also well-known that digital resources do not get optimum engagement on their own. The instructor's approach for implementation can have a sizable impact on student engagement—and ultimately—on outcomes. In this paper, we will compare two Psychology courses run in the Spring of 2020 and Spring of 2021 at the University of Central Florida. The courseware used was initially created by artificial intelligence and further enhanced by the instructor and instructional designer. The instructor taught both classes online using the same courseware, but made changes to how she implemented it. We will compare data from both sections to understand how these implementation changes impacted students—from platform engagement and learning data to student exam data. Results show that the instructor's implementation changes increased student use of the courseware throughout the semester, and also increased exam scores. This direct comparison showcases the importance of instructor choices when incorporating digital resources into the classroom and provides a set of successful implementation practices for other educators to model in the future. In a time of significant change in education, it is more important than ever to better understand how technology and teaching practice can work together to help students be successful.

Keywords: Courseware, Formative Practice, Learn by Doing, Student Engagement, Learning Outcomes, Implementation, Teaching Practice

Introduction

Digital learning resources are changing educational ecosystems for students and teachers alike, offering new ways of teaching and learning and new insights into how learning works. Digital learning resources have also been the focus of research to determine how they can best support learning. For example, Carnegie Mellon's Open Learning Initiative developed and studied best practices for online learning (Lovett et al., 2008), providing the learning science foundation for the courseware used in this study. Yet for as valuable as the focus has been on developing high quality, effective digital resources for teaching and learning, there is still a great misconception that the resource itself could, or should, stand alone as a solution to teaching and learning. Kessler et al. (2019) noted, "research consistently indicates that instructional innovations are only as effective as their implementation." Implementation has been identified as a key component in effectiveness studies to understand how well an intervention performs in naturally occurring contexts (O'Donnell, 2008). Previous research on a courseware learning resource found that different instructor implementation policies for the same courseware greatly influenced student engagement (Van Campenhout and Kimball, 2021). The goal of this research paper is to analyze the impact instructor implementation practices had on student engagement in courseware used in an online Psychology course at the University of Central Florida, and the subsequent difference in student exam scores and course grades.

The digital learning resource used in this study is courseware—a comprehensive learning environment that combines textbook expository text, formative practice, adaptive activities, and summative assessments. Content is chunked into short lessons that are aligned to learning objectives and grouped into units. Previous research found that similarly designed courseware environments helped students learn more efficiently than traditional methods (Lovett et al., 2008). Additional features such as the adaptive activities have also been found to help struggling students increase their learning outcomes (Van Campenhout et al., 2020).

The primary learning method employed in the courseware is learn by doing: the inclusion of formative practice for each lesson of content. This approach creates the doer effect—the learning science principle that doing practice while reading leads to higher learning gains than just reading (Koedinger et al., 2016). Studied extensively at Carnegie Mellon, the doer effect was found to be causal to learning (Koedinger et al., 2016; 2018). The causal doer effect analysis was replicated in courseware on the Acrobatiq platform, confirming external validity of this learn by doing method (Van Campenhout et al., 2021a). The doer effect was also found to remain nearly unchanged even when accounting for prior knowledge and student demographics (Van Campenhout et al., 2021b), indicating its usefulness to all students. With the learn by doing method of courseware proven to generate the doer effect and found to cause better outcomes in a variety of natural learning contexts, we are confident in recommending that all students should take advantage of this learning approach.

The formative practice, adaptive activities, and summative assessments provide learning opportunities for students, but these features also generate data that is surfaced to the instructor in dashboards. Clickstream data are used to generate predictive learning estimates that drive the adaptivity for students and are surfaced in the dashboard to help instructors identify struggling students or challenging content. The data available to instructors makes it possible for instructors to closely monitor their students' progress, tailor in-class content, and evaluate learning concerns for individual students. The availability of this type of data for instructor use is a type of Course Signal, which has been shown to help improve retention in

courses and institutions (Arnold and Pistilli, 2012; Baker, 2016). Researchers have argued that proper utilization of both the educational environment and the learning technology should produce better results than either could produce on its own (Ritter et al., 2016). Providing instructors with tools to manage their class is second only to providing students with proven learning features.

Given the proven learning benefits of courseware, do all students take advantage of this resource when available? No. Student access to learning resources does not also mean they will use those resources as intended. This is not a novel concept. Research into textbook usage confirmed what many instructors experience: students did not use textbooks as intended (Fitzpatrick and McConnell, 2008). That same is true of digital learning resources; just because students have them doesn't mean they will all use them as intended without motivation. In a previous research study on engagement in courseware (Van Campenhout and Kimball, 2021), data were collected from 20 sections of a probability and statistics course across 8 institutions. The goal was to use the same courseware across a state-wide system of schools to ensure credit could transfer between institutions. Engagement in those course sections varied dramatically, even between instructors in the same department. The best predictor of student engagement was how the instructor chose to implement the courseware in their class.

In higher education, there are many models for teaching and learning currently being employed, even within the same institution. Whether face-to-face, online, or hybrid, in the majority of cases the instructor is still the architect of the scope, content, assessment, and grading policies for the course. The use of technology and learning resources in a classroom will always be context-specific. Therefore, the most direct and effective method for influencing student use of learning resources is the instructor and their approach to implementation.

Understanding instructor implementation is also critical in educational research. In a literature review updating and expanding previous research by Fullan and Pomfret (1977), O'Donnell (2008) outlines the state of implementation research in education:

Although seemingly well defined in the health literature (cf. Hansen, Graham, Wolkenstein, & Rohrbach, 1991; Kolbe & Iverson, 1981), fidelity of implementation is rarely reported in large-scale education studies that examine the effectiveness of K–12 core curriculum interventions, especially with regard to how fidelity enhances or constrains the effects of the intervention on outcomes (L. D. Dobson & Cook, 1980; NRC, 2004; U.S. Department of Education, 2006). Moreover, according to the NRC (2004), even less seldom is such a measure of fidelity to K–12 curriculum interventions used to adjust for or interpret outcome measures, (p. 34).

While O'Donnell's focus was on K-12 education research, the same is true in higher education. When considering how an intervention or new educational technology could impact student outcomes in any classroom, implementation matters. Fidelity of implementation is the extent to which "an intervention is implemented in comparison with the original program design during an efficacy and/or effectiveness study" (O'Donnell, 2008, p. 33). Implementation is therefore a natural component of effectiveness studies, where effectiveness is "the ability of an intervention to produce the desired beneficial effect in actual use" (Dorland, 1994, p. 531). In effectiveness studies, "variations in fidelity are measured in natural settings and then related to student outcomes" (O'Donnell, 2008, p. 42).

Given this relationship between effectiveness studies and implementation, we can see why these constructs are so important for research on educational technology. The courseware itself was designed using established learning science principles; the intention of the learn by doing method of integrating practice with learning content is to elicit the doer effect, which has been proven to benefit student learning outcomes. However, students will not benefit from the doer effect if they do not do the practice. By studying how one instructor changed implementation of the courseware between semesters, we can evaluate how the implementation can impact student engagement and outcomes, thereby evaluating the effectiveness of the courseware when implemented under different conditions. Our research questions for this study are:

- How did changing implementation practices of the psychology courseware impact student engagement between semesters?
- How did the change of student engagement with the courseware impact learning outcomes?

Methods

The Courseware

The courseware used in this study was initially generated by an artificial intelligence-based process called SmartStart (Dittel et al., 2019). This process uses an e-textbook as the corpus and applies natural language processing and machine learning techniques to identify learning objectives, divide the content into lessons aligned to learning objectives, and apply an automatic question generation process to create formative practice and feedback for each lesson (Jerome et al., 2020). For this psychology course, a Psychology of Sex and Gender (Bosson et al., 2019) textbook was used for the SmartStart process. The result was a base courseware learning environment that divided the textbook into objective-aligned lessons and included over 600 automatically generated (AG) practice questions. Previous research on SmartStart AG questions from six different courses used in natural learning contexts, including this Psychology course, found that students did not treat these questions any differently than the human-authored counterparts (Van Campenhout et al., 2021c). This large-scale AG question evaluation studied three performance metrics—engagement, difficulty, and persistence—and found no significant differences between AG and human-authored questions on any metric.

This generated courseware was then further enhanced by the instructor and instructional designer. Additional human-authored questions were taken from ancillary materials and added to the lesson pages as formative practice. This was primarily to increase the number of formative practice questions where necessary to be able to generate a predictive learning estimate within the platform. Summative assessments were also added at the unit level. Prior to the Spring 2021 course, the instructor also wrote questions to create scaffolded adaptive activities for the three most challenging chapters within the courseware. The adaptive activities had been shown to help students increase their learning estimate and benefited lower-performing students the most (Van Campenhout et al., 2020). The adaptive activities were also formative for students, providing immediate feedback and the opportunity to continue to try again.

The Implementation

The Psychology of Sex and Gender course was offered online to students (majors and non-majors) at UCF. Although there were students from all years present in each semester, more than 80% were juniors and seniors. About 60% of students were transfer students at UCF and 30% were first-time college students. It is also important to note that the student population at UCF is generally experienced with digital learning resources, including a variety of adaptive learning platforms.

Taught entirely online by the first author, an experienced faculty lecturer, this course was taught in a flipped-blended model where the courseware was the primary learning resource assigned to provide instruction as homework. The synchronous class sessions were designated for activities, discussion, and feedback. This flipped-blended approach has been found to be the most effective type of instructional model (Margulieux et al., 2015). This class was also a High-Impact (HIP) designated Service-Learning course. The first section of this course using this courseware ran in the Spring of 2020 (n = 119), meaning that this section was running when the COVID-19 outbreak first disrupted education globally. However, as this course was originally delivered online, no changes had to be made to its implementation. Students progressed through the remainder of the semester as expected. The second section of this course was run in the Spring of 2021 (n = 125).

There were many similarities in implementation between semesters. The course was delivered through a learning management system. The course was organized into weeks with specific instructions for assignments in each week. The courseware was linked to the learning management system through single sign-on so students only had to click the link to be taken to the assigned section of the courseware. Each week had reminders to do the assigned courseware section, and reminders were sent via email and during synchronous class as well. The course was specifically set up to provide as many opportunities to keep students on track as possible.

However, there were also differences in how the courseware was implemented between semesters. For the Spring of 2020 section, the formative practice completion was given a 2% contribution to students' overall grade. Students had to complete more than 85% of practice to receive their points, and this was evaluated at the end of the semester. Note this was for completion of practice and not first attempt accuracy of practice to retain the formative nature of the practice. The Spring 2021 section was still deeply impacted by the pandemic and this class was not a HIP designated Service-Learning course. The instructor continued to post three announcements per week but increased the focus on the adaptive courseware as a self-monitoring learning tool. Also, for the Spring 2021 section, changes were made to the scoring policy for the practice questions. For this semester, the practice questions were worth 20% of student grades, still with a minimum threshold of 85%.

Results and Discussion

The first type of data to look at is engagement data—how students used the courseware in each semester. In aggregate, students in the Spring 2020 semester completed 48% of practice, while students in the Spring 2021 semester completed 76% of practice. These practice totals include all students on the roster (even those who may have dropped) and all questions (even those on pages students may not have visited). A data visualization called an engagement graph can provide a more holistic view of how each class read and practiced in the course.

The engagement graph shows the number of students on the y-axis and the pages of the course on the x-axis. Attrition as the semester progresses is a natural and expected pattern. A vertical gap between the reading dot (blue) and the practice dot (red) is also typical, as there will usually be some students who read the page but choose not to do the practice available.

The engagement graph for Spring 2020 is very standard, if not above average compared to many other historical course sections (see Van Campenhout & Kimball, 2021). There is some general level of attrition over the entire course, as well as within units. Some students stop using the courseware over time. There is a gap between the reading and doing dots, indicating roughly five to ten students read the content but do not do practice. The assessment dots (green) are higher than the rest by about 20 students, indicating those students went directly to the assessment without reading the material at all.

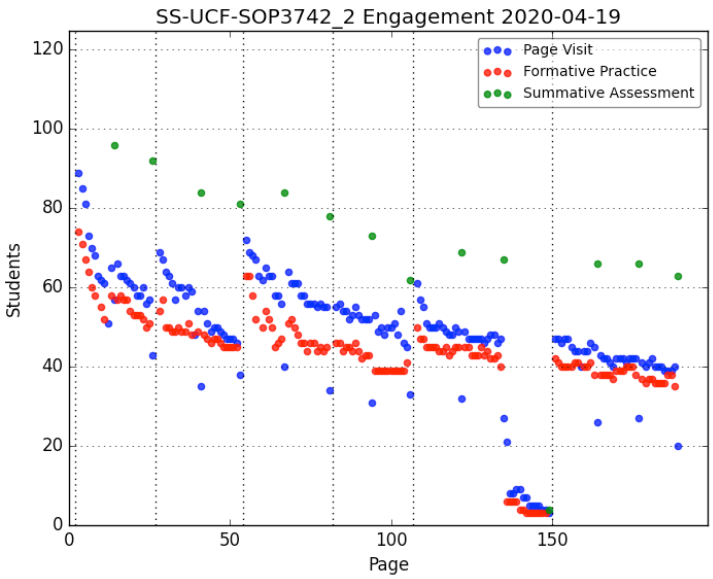

Figure 1: The engagement graph for Spring 2020.

The engagement graph for Spring 2021 shows a different pattern. While there is still some attrition, it is minimal over the duration of the courseware. Not only did nearly all students stay in the course, but the reading-doing gap is nonexistent. The reading and doing dots are together throughout the entire course. If we consider that an ideal engagement graph would be a horizontal line—showing all students reading and doing practice on every page—this Spring 2021 graph is very close.

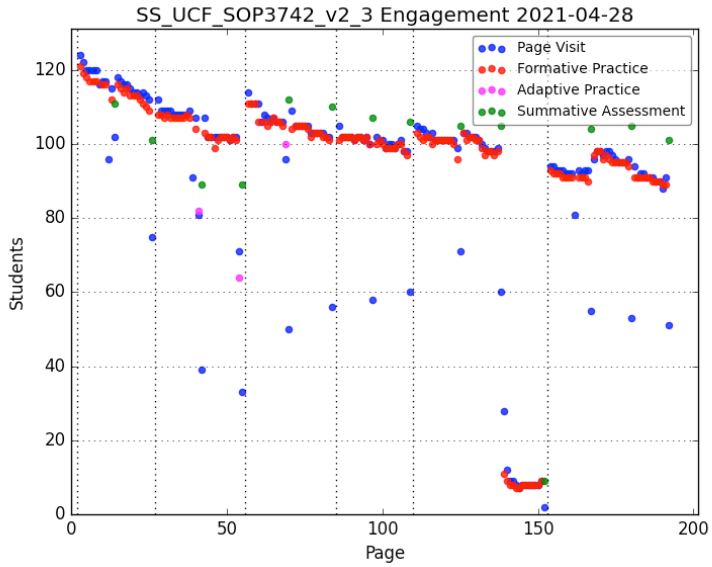

Figure 2: The engagement graph for Spring 2021.

Next, student exam scores can be compared between semesters. As seen in Table 1, student mean scores increased on each exam when using the adaptive version of the text. The Spring 2020 mean exam scores were several points higher than the Fall 2019 e-text baseline, yet the Spring 2021 mean scores increased even further. In Spring 2021, when the courseware was further incentivized and the personal practice was added to the most difficult chapters, test scores increased significantly.

		Exam 1	Exam 2	Exam 3
Fall 2019	Mean Score	60%	63%	71%
	Score Range	39–101%	12–104%	19–104%
	n Students	97	78	71
Spring 2020	Mean Score	70%	68%	78%
	Score Range	23–98%	24–104%	39–104%
	n Students	98	86	71
Spring 2021	Mean Score	77%	78%	79%
	Score Range	43–102%	42–102%	42–99%
	n Students	106	105	104

Table 1. Mean scores on the three course exams, 2019–2021.

These data revealed additional trends of note. First are the score ranges for the exams. With the exception of Exam 1, the lowest scores on the 2020 exams were twice as high as the previous year. Even more interesting is that the lowest scores for the Spring 2021 exams are even higher still, and nearly consistent at 42–43%. In addition to higher mean scores, the incentivized practice could have benefited struggling students and raised the lower bound of exam scores. This trend is worth additional investigation in future research. Another notable result from Table 1 is the number of students who took each exam across the years. The Fall 2019 and Spring 2020 courses have fewer students taking Exam 3 compared to Exam 1 (26 and 27 students, respectively). However, the number of students taking the Spring 2021 exams only decreases by two students. The high retention of students in the Spring 2021 course is also worth further investigation, as student retention is an important issue in higher education along with student learning.

Conclusion

There is unquestionable value in learning science research, as it reveals critical insights into how students learn and directs future research and development. The doer effect research, for example, quantified the benefits of doing practice while reading and revealed the causal nature of this relationship. Yet equally important is research into the application of these learning science principles to understand how they can be successfully implemented into classrooms and impact student learning outcomes. While the courseware used in this study was designed to engage students in learn by doing to generate the doer effect, this method will not benefit students if it is not used. Instructors have enormous sway over student engagement through their implementation of learning resources within the course. By comparing different implementation choices and the impact they had on student engagement and outcomes, we can add to the effectiveness research on how the learn by doing method in courseware can practically help students.

In this study, we can clearly see the effects of instructor implementation on student engagement with Psychology courseware between semesters. In Spring 2020, the courseware was the primary learning resource linked through the LMS and engaging with the practice was assigned 2% of the course grade. The engagement graph showed typical patterns of attrition over the course and a moderate percentage of students not reading or doing the practice. In Spring 2021, the instructor changed the grade percentage to 20% and the engagement graph shows very high reading and doing with very little attrition over the length of the courseware. While the instructor uses other strategies such as frequent LMS reminders, email updates, and in-class instructions, the change in participation points drove engagement up to near perfect levels. Students take course grades seriously and assigning points to the formative practice places a value on the learning content and process of learning while maintaining the purpose of the formative practice as just that—low-stakes practice.

What's even more meaningful for students was the shift in exam scores. The Fall 2019 scores were included as a control to provide scores using only the e-text. The Spring 2020 course using the courseware (the same e-text as courseware with the formative practice) saw an increase in mean scores for two of three exams over the previous year. The Spring 2021 course—in addition to the incentivized practice increasing engagement—had even higher mean exam scores than the previous year. These results confirm that implementation policies can increase student engagement in the courseware and ultimately exam scores. The learn by doing method has practical and meaningful implications for student learning and outcomes, and instructors can help maximize this benefit for students through their implementation choices.

Acknowledgements

We would like to acknowledge and thank Joe Lloyd and Linda Winer for their efforts on the development and implementation of this course. We also thank Benny Johnson for his feedback on this manuscript.

References

Arnold, K. E., & Pistilli, M. D. (2012) Course Signals at Purdue: Using learning analytics to increase student success. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. New York, NY: ACM, pp. 267–270.

Baker, R. S. (2016). Stupid Tutoring Systems, Intelligent Humans. International Journal of Artificial Intelligence in Education, 26(2), 600–614. https://doi.org/10.1007/s40593-016-0105-0

Bosson, J. K., Vendello, J. A., & Buckner, C. V. (2018). The psychology of sex and gender (1st ed.). Thousand Oaks, California: SAGE Publications.

Dittel, J. S., Jerome, B., Brown, N., Benton, R., Van Campenhout, R., Kimball, M. M., Profitko, C., & Johnson, B. G. (2019). SmartStart: Artificial Intelligence Technology for Automated Textbook-to-Courseware Transformation, Version 1.0. Raleigh, NC: VitalSource Technologies.

Fitzpatrick, L., & McConnell, C. (2008). Student reading strategies and textbook use: an inquiry into economics and accounting courses. Research in Higher Education Journal, 1–10.

Fullan, M., & Pomfret, A. (1977). Research on curriculum instruction implementation. Review of Educational Research, 47, 335–397.

Jerome, B., Van Campenhout, R., & Johnson, B. G. (2021). Automatic Question Generation and the SmartStart Application. Learning at Scale. https://doi.org/10.1145/3430895.3460878

Kessler, A., Boston, M., & Stein, M. K. (2019). Exploring how teachers support students' mathematical learning in computer-directed learning environments. Information and Learning Science, 121(1–2), 52–78. https://doi.org/10.1108/ILS-07-2019-0075

Koedinger, K., McLaughlin, E., Jia, J., & Bier, N. (2016). Is the doer effect a causal relationship? How can we tell and why it's important. Learning Analytics and Knowledge. Edinburgh, United Kingdom. http://dx.doi.org/10.1145/2883851.2883957

Koedinger, K. R., Scheines, R., & Schaldenbrand, P. (2018). Is the doer effect robust across multiple data sets? Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, 369–375.

Lovett, M., Meyer, O., & Thille, C. (2008). The Open Learning Initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. Journal of Interactive Media in Education, (1), 1-16. http://doi.org/10.5334/2008-14

Margulieux, L. E., McCracken, W. M., & Catrambone, R. (2015). Mixing in-class and online learning: Content meta-analysis of outcomes for hybrid, blended, and flipped courses. In O. Lindwall, P. Hakkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.) Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference (pp. 220-227), 2. Gothenburg, Sweden: The International Society of the Learning Sciences.

O'Donnell, C. L. (2008). Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K-12 Curriculum Intervention. (2008). Review of Educational Research. 78(1). Pp. 33–84. https://doi.org/10.3102/0034654307313793

Ritter, S., Fancsali, S., Yudelson, M., Rus, V., & Berman, S. (2016). Toward intelligent instructional handoffs between humans and machines. Workshop on Machine Learning for Education, The Thirtieth Conference on Neural Information Processing Systems (NIPS), Barcelona.

Van Campenhout, R., Dittel, J. S., Jerome, B., & Johnson, B. G. (2021c). Transforming textbooks into learning by doing environments: an evaluation of textbook-based automatic question generation. In: Third Workshop on Intelligent Textbooks at the 22nd International Conference on Artificial Intelligence in Education. CEUR Workshop Proceedings, ISSN 1613-0073, pp. 1–12. http://ceur-ws.org/Vol-2895/paper06.pdf

Van Campenhout, R. & Kimball, M. (2021). At the intersection of technology and teaching: The critical role of educators in implementing technology solutions. IICE 2021: The 6th IAFOR International Conference on Education. ISSN 2189-1036, pp. 151–161. https://doi.org/10.22492/issn.2189-1036.2021.11

Van Campenhout, R., Jerome, B., & Johnson, B. G. (2020). The impact of adaptive activities in Acrobatiq courseware: Investigating the efficacy of formative adaptive activities on learning estimates and summative assessment scores. In: Sottilare R., Schwarz J. (eds) Adaptive Instructional Systems. HCII 2020. LNCS, vol 12214. Springer. pp 543–554. https://doi.org/10.1007/978-3-030-50788-6_40

Van Campenhout, R. Johnson, B. G., & Olsen, J. A. (2021a). The Doer Effect: Replicating Findings that Doing Causes Learning. Presented at eLmL 2021: The Thirteenth International Conference on Mobile, Hybrid, and Online Learning. ISSN 2308-4367, pp. 1–6. https://www.thinkmind.org/index.php?view=article&articleid=elml_2021_1_10_58001

Van Campenhout, R. Johnson, B. G., & Olsen, J. A. (2021b). The Doer Effect: Replication and Comparison of Correlational and Causal Analyses of Learning. International Journal on Advances in Systems and Measurements, v14, n 1&2 2021

Contact email: rachel.vancampenhout@vitalsource.com