

## Leveraging Learning by Doing in Online Psychology Courses: Replicating Engagement and Outcomes

Martha Hubertz  
Psychology Department  
University of Central Florida  
Orlando, USA  
email: martha.hubertz@ucf.edu

Rachel Van Campenhout  
Learning Science  
VitalSource Technologies  
Raleigh, USA  
email: rachel.vancampenhout@vitalsource.com

**Abstract**—For two upper-level online psychology courses—Psychology of Sex & Gender and Forensic Psychology—the instructor used adaptive courseware as the primary digital learning resource. This courseware was designed to integrate formative practice with textbook content in a learn-by-doing method known to generate the doer effect—a learning science principle shown to have six times the effect on learning than reading alone. The doer effect research provides confidence in the efficacy and generalizability of this learning method. However, it is imperative to investigate the practical implication of this learning method on student outcomes in natural learning contexts. How does doing practice while reading help students' exam scores? This instructor applied similar courseware implementation practices to both courses and was able to create similarly high engagement within the courseware as well as increased exam scores. These results show the importance of combining learning science methods with instructor expertise to provide an optimal learning environment for students, as well as the ability to replicate results across semesters and courses.

**Keywords**—online learning; learn by doing; courseware; learning outcomes; replication, teaching and learning.

### I. INTRODUCTION

Online learning has been a focus of higher education for decades with an increasing number of online courses offered, as well as entire online degree programs. However, the COVID-19 pandemic focused attention on the need for online learning options more than ever before. In the fall of 2020, 75% of all undergraduates (11.8 million) in the United States had at least one online class and 44% of undergraduates were exclusively enrolled in distance education (7 million) [1]. These numbers are especially impactful compared to those prior to the start of the pandemic in 2019. Undergraduates enrolled in distance education was 97% higher than in 2019 while those enrolled exclusively online rose 186% compared to 2019 [1]. This dramatic change in learning conditions has sparked significant conversation on effective teaching and learning online. Teaching models are not all equally effective [2] and neither are digital learning resources. In this paper, we describe online Psychology courses taught at the University of Central Florida (UCF) to illuminate both the teaching practices and the learning resources that produced positive learning outcomes. These examples of successful online teaching and learning practices are particularly beneficial to

share given that increased student outcomes were repeated in multiple semesters and across multiple courses.

Along with various forms of online and distance learning, there are many types of digital learning resources and environments. Learning science research has long investigated how best to learn in digital learning environments. Researchers at Carnegie Mellon University's Open Learning Initiative identified that courseware environments that combined frequent formative practice with expository text and media in objective-aligned lessons helped students learn the same or more information and retain it for longer than their peers in traditional courses [3]. This combination of formative practice with reading content became the focus of an area of learn-by-doing research. Koedinger et al. [4][5] found through correlational investigation that doing practice while reading had about six times the effect size on learning than just reading. Calling this learning by doing method the doer effect, they were also able to model this relationship to infer causality [5][6]. The doer effect research was replicated using similar courseware environments at different universities and found the same correlational and causal results [7][8][9]. In the field of education in particular, research on learning methods is all too often not replicated, and results are cited that could not be replicated [10]. The replication of the doer effect research confirms that this learning by doing method is generalizable across contexts and should be broadly applied.

The doer effect learning science principle proves doing practice is causal to learning, yet this research does not indicate the impact it will have to student learning outcomes in a practical sense (i.e., how much will doing practice increase exam scores?) [9]. Carvalho et al. [11] expressed the need to investigate this in natural learning contexts and, using data from MOOCs, identified that more doing did lead to higher quiz and exam scores. Their research also reported that students often overestimated the benefit of reading and underestimated that of doing practice. To further understand how the learn by doing method benefits students on course outcomes, this approach was applied in a large online Psychology course at UCF. The instructor found that student exam scores increased after they were assigned the practice as an integrated part of their core reading content [12].

In this paper, we extend the prior classroom research by investigating two different Psychology courses taught by the same instructor at UCF. The goal of this paper is to identify if the same learn by doing courseware environment with the

same implementation practices can produce similar improved exam scores for two different courses over multiple semesters.

## II. METHODS

The courseware learning environment employed in these courses was automatically generated through a process called SmartStart. The textbook was used as the primary learning text and natural language processing and machine learning tools were applied to identify learning objectives in the text, chunk content into shorter lessons, and generate formative practice [13]. Two types of practice questions were generated: matching and fill-in-the-blank. A large-scale analysis of these questions using student data from natural learning environments found that they performed as well as human-authored questions on engagement, difficulty, and persistence [14]. In addition to these automatically generated questions, the instructor added human-authored questions as additional formative practice to the lesson pages, and wrote adaptive activities for four of the most challenging topics in the courseware.

The students included in this analysis selected these courses as part of their coursework in Psychology. The majority of students were juniors and seniors and taking these courses as electives in their major. The student population at UCF consists of a high proportion of first generation (25%) and transfer students (70%). Both Psychology of Sex and Gender and Forensic Psychology were taught as entirely online courses with synchronous weekly class periods. The instructor used the learning management system to organize week by week instructions for students, post reminders, and link the learning resources for the course. The courseware was also linked in the learning management system to open to each week's chapter of the course.

The application of learning science in the classroom is a complex task for teaching and learning, as context matters when considering the needs of students and the most appropriate method of implementing technology for learning [15]. In this case, the instructor applied implementation practices augmented over several semesters that included a clear introduction of the courseware, frequent reminders in the learning management system, and assigned points for the completion of the practice questions. Incentivizing the practice holds students accountable for using their learning materials, but also *places value on the process of learning*, not just the outcomes. Each course had 20% of the assigned points for completing a minimum of 85% of the total formative practice questions. The points were only dependent on completion—not on first-attempt accuracy—to maintain the formative nature of the practice.

## III. RESULTS

### A. Psychology of Sex & Gender

To see how these courseware implementation practices impacted students, engagement data from the courseware platform was combined with student outcome data from the course exams. The platform engagement data was plotted on

graphs to provide a visual of how many students read and did practice on each page of the courseware. An engagement graph for the initial semester (2020) for Psychology of Sex and Gender is shown for comparison, as it had 2% points assigned for practice and shows a more typical student engagement and overall attrition [12]. In this example (Figure 1), the vertical space between the blue reading dots and red practice dots indicates that some students were reading and not doing any practice on those pages. This vertical space is called the reading-doing gap. Figure 1 also shows a very typical trend for overall engagement—attrition both within chapters and over the entire course.

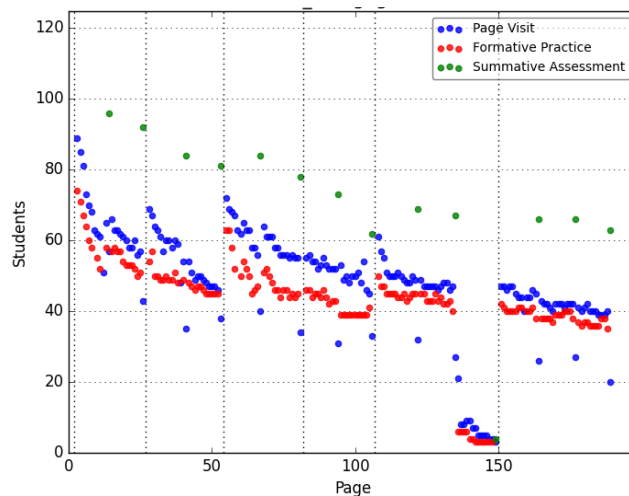


Figure 1. 2020 Psychology of Sex & Gender engagement.

The following two semesters of the Psychology of Sex and Gender course (Figure 2 and 3) had 20% of the grade assigned for completing 85% of the formative practice. By comparison, both of these graphs have no visible reading-doing gap, and there is very little attrition during the course. Compared to Figure 1, these two engagement graphs show a dramatically improved student engagement pattern in the courseware.

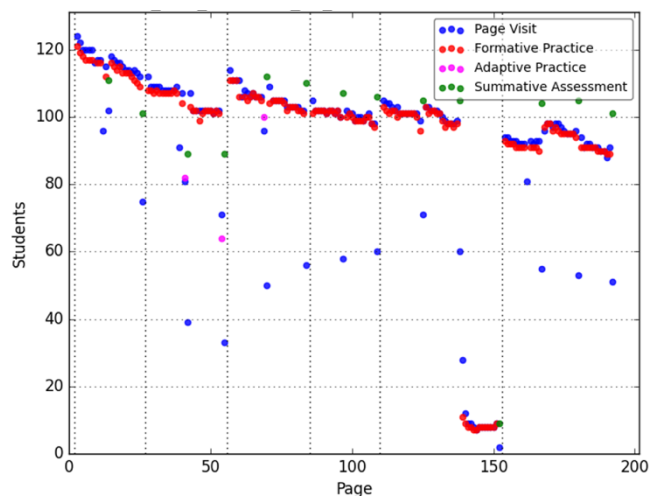


Figure 2. 2021 Psychology of Sex & Gender engagement.

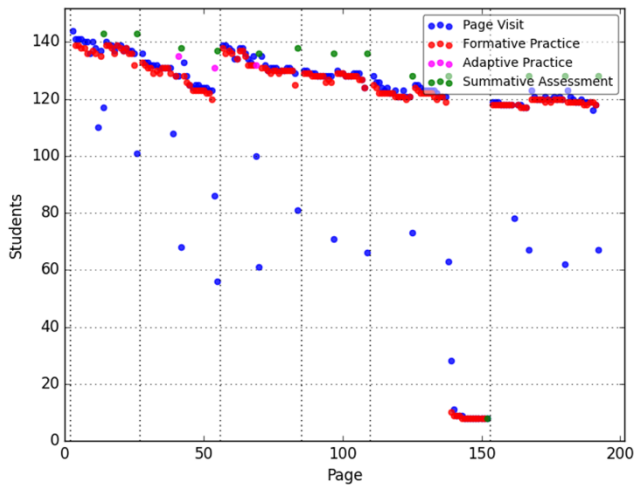


Figure 3. 2022 Psychology of Sex & Gender engagement.

Data from the three course exams is included in Table 1 to compare across several years. The exam data from 2019 was included as a comparison as this was a semester that used the same digital textbook without the learning by doing method. The spring 2020 semester was the first semester with courseware and the 2% for practice, and spring 2021 and 2022 had the 20% for practice. The mean exam scores increased between 5 and 10 points the first semester using courseware (2020) and increased again another 7 to 10 points in 2021 for exams 1 and 2. The mean exam scores for 2021 and 2022 remained close, showing a consistency in outcomes for each of these semesters.

TABLE I. EXAM SCORES FROM PSYCHOLOGY OF SEX AND GENDER ACROSS FOUR YEARS.

		Exam 1	Exam 2	Exam 3
Fall 2019	Mean Score	59%	66%	71%
	Score Range	39–101%	12–104%	19–103%
	n Students	64	66	66
Spring 2020	Mean Score	70%	68%	84%
	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
Spring 2022	Mean Score	75%	78%	80%
	Score Range	30–104%	39–104%	15–104%
	n Students	130	130	127

### B. Forensic Psychology

Forensic Psychology was selected for the SmartStart courseware process after the initial semester of Psychology of Sex and Gender was completed. For this course, both the 2021 and 2022 semesters had 20% of the course grade assigned to the formative practice. The engagement graphs for each semester (Figure 4 and 5) mirror those seen for the Psychology of Sex and Gender course, with a minimal reading-doing gap and very little attrition across the

course. It is also notable that while all courses presented here are large sections, the 2021 Forensic Psychology course had more than 250 students and still maintained consistent engagement.

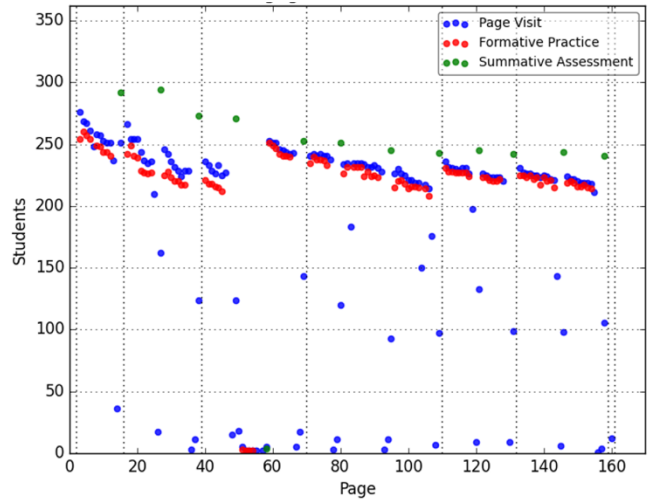


Figure 4. 2021 Forensic Psychology engagement.

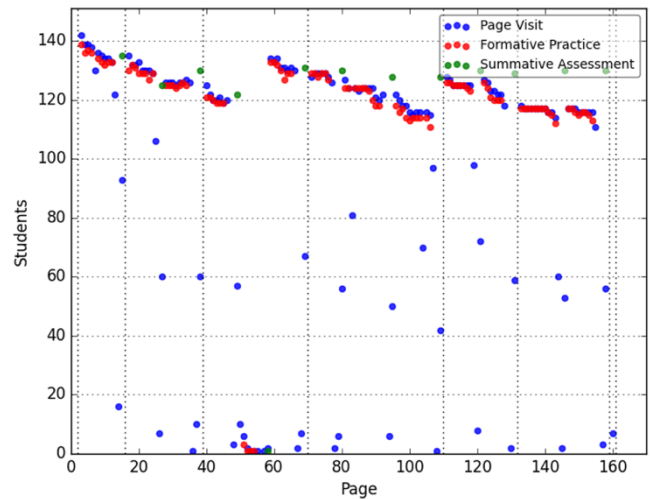


Figure 5. 2022 Forensic Psychology engagement.

The Forensic Psychology spring 2021 and 2022 mean exam scores are within a few points of each other and are within a similar range as Psychology of Sex and Gender. While there was no historical section for comparison, they are consistent in engagement and outcomes.

TABLE II. EXAM SCORES FROM FORENSIC PSYCHOLOGY.

		Exam 1	Exam 2	Exam 3
Spring 2021	Mean Score	75%	80%	83%
	Score Range	35–103%	26–102%	32–105%
	n Students	249	248	242
Spring 2022	Mean Score	77%	81%	82%
	Score Range	41–102%	39–104%	52–103%
	n Students	134	133	129

## IV. CONCLUSION

As online learning continues to become increasingly prevalent in a post-pandemic world, identifying effective teaching and learning strategies and tools that benefit students is key. Yet even more important is to ensure these methods and results are reproducible. This instructor applied the same teaching practices to two different online Psychology courses over multiple semesters, both of which utilized the same courseware platform as the primary learning by doing resource. The incentivization of the practice questions in the courseware increased student engagement dramatically. In both sections of both courses, attrition in the units and course was reduced to a minimal amount. The reading-doing gap was also nearly eliminated as students consistently did the practice as they read. Maximizing student engagement by incorporating formative practice into the course grade helped students take advantage of the doer effect's learning benefits. The practical outcome of maximizing student learning by doing was increased exam scores. That these outcomes were from multiple sections over different years, as well as from different courses, shows a particularly important outcome: these instructor practices and learning outcomes can be replicated. Research in teaching and learning needs more cases of learning interventions that can be replicated in natural learning contexts in order to recommend practices at scale.

Future research should work to extend the generalizability of these findings by investigating if similar implementation strategies can increase student engagement with formative practice and increase learning outcomes in a variety of subjects and institutions. Follow up research is also planned to investigate how completing formative practice in this courseware environment may benefit first generation and at-risk students specifically. As a school with a large first generation and transfer student population, approaches that benefit and support these student populations is important to UCF. Online learning will continue to grow at UCF and other higher education institutions and effective, reproducible teaching and learning methods will need to be researched and adopted.

## REFERENCES

- [1] National Center for Education Statistics. Undergraduate Enrollment. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved May 31, 2022, from <https://nces.ed.gov/programs/coe/indicator/cha>
- [2] L. E. Margulieux, W. M. McCracken, and R. Catrambone. Mixing in-class and online learning: Content meta-analysis of outcomes for hybrid, blended, and flipped courses. *Computer-Supported Collaborative Learning Conference*, CSCL, 1, 2015, pp. 220–227.
- [3] M. Lovett, O. Meyer, and C. Thille. The Open Learning Initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *Journal of Interactive Media in Education*, (1), 2008, pp. 1-16. <http://doi.org/10.5334/2008-14>
- [4] K. Koedinger, J. Kim, J. Jia, E. McLaughlin, and N. Bier, "Learning is not a spectator sport: doing is better than watching for learning from a MOOC." In: *Learning at Scale*, pp. 111–120, 2015. Vancouver, Canada. <http://dx.doi.org/10.1145/2724660.2724681>
- [5] K. Koedinger, E. McLaughlin, J. Jia, and N. Bier, "Is the doer effect a causal relationship? How can we tell and why it's important." *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, LAK 2016, pp. 388–397. <http://dx.doi.org/10.1145/2883851.2883957>
- [6] K. R. Koedinger, R. Scheines, and P. Schaldenbrand, "Is the doer effect robust across multiple data sets?" *Proceedings of the 11th International Conference on Educational Data Mining*, EDM 2018, pp. 369–375. <http://dx.doi.org/10.1145/2883851.2883957>
- [7] R. Van Campenhout, B. G. Johnson, and J. A. Olsen. The doer effect: replicating findings that doing causes learning. *Proceedings of eLmL 2021: The Thirteenth International Conference on Mobile, Hybrid, and On-line Learning*. ISSN 2308-4367, 2021, pp. 1–6. [https://www.thinkmind.org/index.php?view=article&articleid=elml\\_2021\\_1\\_10\\_58001](https://www.thinkmind.org/index.php?view=article&articleid=elml_2021_1_10_58001)
- [8] R. Van Campenhout, B. G. Johnson, and J. A. Olsen. The doer effect: Replication and comparison of correlational and causal analyses of learning. *International Journal on Advances in Systems and Measurements*, 15(1&2), pp. 48–59, 2022).
- [9] R. Van Campenhout, B. Jerome, and B. G. Johnson. The Doer Effect at Scale: Investigating Correlation and Causation Across Seven Courses. In *LAK23: 13th International Learning Analytics and Knowledge Conference (LAK 2023)*, pp. 357–365, 2023. <https://doi.org/10.1145/3576050.3576103>
- [10] M. Serra-Garcia, and U. Gneezy, "Nonreplicable publications are cited more than replicable ones." In *Science Advances*, vol. 7, pp. 1–7, 2021. <http://doi.org/10.1126/sciadv.abd1705>
- [11] P. F. Carvalho, E. A. McLaughlin, and K. R. Koedinger. Is there an explicit learning bias? Students beliefs, behaviors and learning outcomes. *Proceedings of the 39th Annual Conference of the Cognitive Science Society (Eds. Gunzelmann, G. et al.)*, pp. 204–209, 2017.
- [12] Hubertz, M. & Van Campenhout, R. (2022). Teaching and iterative improvement: The impact of instructor implementation of courseware on student outcomes. *The IAFOR International Conference on Education – Hawaii 2022 Official Conference Proceedings*. <https://doi.org/10.22492/issn.2189-1036.2022.19>
- [13] B. Jerome, R. Van Campenhout, and B. G. Johnson. Automatic Question Generation and the SmartStart Application. *Learning at Scale*, 2021. <https://doi.org/10.1145/3430895.3460878>
- [14] R. Van Campenhout, J. S. Dittel, B. Jerome, and B. G. Johnson, "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*, 2021. Retrieved June 30, 2022 from: [https://intextbooks.science.uu.nl/workshop2021/files/iTextbooks\\_2021\\_paper\\_6.pdf](https://intextbooks.science.uu.nl/workshop2021/files/iTextbooks_2021_paper_6.pdf)
- [15] R. Van Campenhout and A. Kessler. Developing Instructor Training for Diverse & Scaled Contexts: A Learning Engineering Challenge. *Proceedings of eLmL 2022: The Fourteenth International Conference on Mobile, Hybrid, and On-line Learning*, pp. 29–34, 2022. [https://www.thinkmind.org/index.php?view=article&articleid=elml\\_2022\\_2\\_40\\_58010](https://www.thinkmind.org/index.php?view=article&articleid=elml_2022_2_40_58010)