

Teaching and Learning with AI-Generated Courseware: Lessons from the Classroom

Kersten T. Schroeder

Martha Hubertz

University of Central Florida

Rachel Van Campenhout

Benny G. Johnson

VitalSource Technologies

Abstract

While research in the learning sciences has spurred advancements in educational technology, the implementation of those learning resources in natural learning contexts advances teaching and learning. In this paper, two faculty members at the University of Central Florida used courseware generated with artificial intelligence as the primary learning resource for their students. The selection and enhancement of this courseware is contextualized for each course. Instructor implementation practices over multiple semesters are described and related to resulting student engagement and exam scores. Finally, benefits of the adaptive courseware are discussed not only for student outcomes, but the qualitative changes faculty identified and the impact that iterative changes in teaching practice had on instructors as well as students.

Keywords: Courseware, Artificial Intelligence, Formative Practice, Learn by Doing, Automatic Question Generation, Automatically Generated Questions, Student Engagement, Learning Outcomes, Implementation, Teaching Practice

Schroeder, K., Hubertz, M., Van Campenhout, R., Johnson, B. G. (2022). Teaching and learning with AI-generated courseware: Lessons from the classroom. *Online Learning*, 26(3), 73-87.

Teaching and learning research are at the heart of moving educational practices forward, as it is knowledge of what works in natural learning contexts that benefits students. While there is enormous benefit in efficacy research that uses controlled experiments to produce results with high internal validity, effectiveness research that applies treatments to classroom settings provides external validity (O'Donnell, 2008). Both efficacy and effectiveness research have benefits, but effectiveness research is particularly beneficial for teaching and learning research as it is the application of a learning technology or pedagogical approach within a specific classroom that matters to the students learning in that instance. Effectiveness research in the classroom has generated new insights into teaching and learning with the rise of digital learning resources which generate a wealth of high-quality data (Goldstein & Katz, 2004; Singer & Bonvillian, 2013) that can provide valuable information to students, instructors, and researchers (Baker & Inventado, 2016). As noted by Koedinger et al. (2016), "the availability of process and outcome data from online courses makes it possible to investigate the generalizability of associations between learning method and outcomes" (p. 388). In this case, adaptive courseware was used as the primary learning resource, which generated engagement data that could be combined with exam grades and instructor observations to evaluate the effectiveness of this resource in their classrooms.

New learning technologies that are based in the learning sciences and able to generate new data insights are not a stand-alone solution for teaching and learning, but rather tools whose success depends on their implementation (Kessler et al., 2019; Sullivan et al., 2020). Implementation has long been an area of focus for the successful application of pedagogical interventions in the classroom (Fullan & Pomfret, 1977), and has been closely tied to effectiveness research (O'Donnell, 2008). Implementation in natural learning settings must also consider the contextual factors specific to each classroom setting, such as teaching model and modality, course subject, and student characteristics (Van Campenhout & Kessler, 2022). Effectiveness research in teaching and learning has the added benefit of identifying successful instructor strategies for implementation that can be put to practical use by other educators. In this paper, we present the implementation practices and student outcome results from two courses at the University of Central Florida wherein the faculty (and first authors) chose to use AI-generated courseware as the primary learning resource (Schroeder et al., 2021). The goals of this paper are to discuss the context for the use of the courseware, the implementation practices for each course across multiple semesters, and the resulting student courseware engagement and exam scores.

Literature Review

Learn by Doing

Courseware is a learning resource that integrates expository content with formative practice questions in short, objective-aligned lessons while also offering adaptivity and assessments within the flow of the student's learning path. The primary learning strategy employed in the courseware used in this study is the integration of formative practice into the text at frequent intervals in a learn by doing method, as seen in Figure 1. Students can answer the questions, receive immediate feedback, and continue attempts if they were incorrect at first. Formative practice is well known to benefit learning across many contexts and for all students but is especially beneficial to students who struggle (Black & William, 2010). By providing

students with practice that they can use while they study, these questions act as no-stakes practice testing, which was found to have high utility and broad applicability relative to other study methods (Dunlosky et al., 2013). This close integration of text with foundational practice engages students in a cognitive process to receive, organize, store, and retrieve information (Ertmer & Newby, 2013). From Piaget's theory of cognitive development (1926) to more recent cognitive and constructive theories such as generative learning theory (Fiorella & Mayer, 2016), the role of the student as an active participant in the learning process is foundational to this learn by doing approach.

Figure 1

Automatically Generated Matching and Fill-in-the-Blank Questions Used as Formative Practice.

The figure shows a screenshot of an online learning environment. At the top, a dark header bar contains the text "Learn by Doing". Below this, the first question is displayed with a lightbulb icon and the text: "Light _____ are advantageous for viewing living organisms, but since individual cells are generally transparent, their _____ are not distinguishable unless they are colored with special _____.
_____ components _____ microscopes _____ stains" Below the question are three buttons: "components", "microscopes", and "stains". A "Check My Answer" button is located below the question. In the bottom right corner of the question area, there is a small text box containing the word "microscopes". The second question at the bottom is partially visible with the text: "In order to gain a better understanding of cellular structure and function, scientists typically use _____ microscopes.
_____ Check My Answer" Below this question is another "Check My Answer" button.

The application of this *learn by doing* method in online learning resources, such as the courseware described here, was studied at Carnegie Mellon University's Open Learning Initiative. This learn by doing environment was found to accelerate learning, increase learning outcomes, and support learners in both asynchronous and instructor-led settings (Lovett et al., 2008). Koedinger et al. (2015) found that doing practice while reading had an average of six times the effect size on learning outcomes compared to solely reading. This learning science principle was called "the doer effect," and follow-up research identified a causal relationship between doing this practice and higher learning outcomes (Koedinger et al., 2016). Causal doer effect research was replicated using courseware data from other natural learning contexts, extending the generalizability of results (Van Campenhout, Olsen, & Johnson, 2021). The doer effect has also been found when accounting for student prior knowledge and demographic characteristics, showing its utility for all learners (Koedinger et al., 2015; Van Campenhout et al., in press).

Instructor Implementation. Yet even though spending more time completing activities has a larger impact on learning outcomes than spending more time reading, students often

underestimate the value of practice and overestimate the value of reading (Carvalho et al., 2017). In addition to student perception of practice, while instructors assign textbooks with the goal that students will read the assigned sections, it is known that students often do not use the textbook as intended (Fitzpatrick & McConnell, 2008). The learning science-based approach of the courseware itself does not guarantee that it—like traditional textbooks—will be used as intended by instructors. This is when the combination of platform data and instructor intervention is key. The combination of text and formative practice offers instructors a practical means by which to monitor student engagement and learning within courseware. Data dashboards with learning analytics on student engagement and performance act as a type of “course signal” for instructors that can be used for intervention and has been shown to increase retention in courses (Arnold & Pistilli, 2012; Baker, 2016). In this learning environment, the platform provides the data analytics but relies on the instructor to interpret and act appropriately based on the context of their teaching and learning environment (Baker, 2016; Van Campenhout & Kessler, 2022). The proper utilization of the educational environment and the learning technology together should produce better results than either could produce on its own (Ritter et al., 2016).

O’Donnell (2008) identified that implementation is a key aspect of effectiveness research to understand how well an intervention performs in natural settings, and the importance of this cannot be understated for educational technology. Research that compared instructors who received the same training and instructions and used the same courseware in their courses identified that differing instructor implementation policies had a large impact on overall student engagement with the courseware (Van Campenhout & Kimball, 2021). Given the influence of instructors over student use of their learning resources, it is valuable to investigate specific instructor implementation practices from naturally occurring learning settings that benefit student learning (Hubertz & Van Campenhout, 2022). In addition to identifying successful teaching practices that optimize student use of technology, it is critical to focus on the natural iterative improvement cycles that are part of this teaching and learning process (Sullivan et al., 2020; Van Campenhout & Kessler, 2022).

Methods

Courses and Participants

Two faculty members self-selected Acrobatiq SmartStart courseware (described below) for their courses, Microbial Metabolism (Microbe) and Psychology of Sex and Gender (Psychology). The Microbial Metabolism course is taught by a Burnet School of Biomedical Sciences faculty member and is typically populated by fourth-year students. Five years ago, the Microbial Metabolism course had been taught in a traditional face-to-face, lecture-style teaching model. Before the pandemic, the instructor then changed the course to a hybrid model (mix-mode) using some of the same traditional lecture format before finally adopting a flipped-blended hybrid model where all the lectures were prerecorded, and active learning exercises were implemented during face-to-face classroom sessions. After a year of using the flipped-blended hybrid model, the AI-generated courseware was adopted to help students learn the material better because moving from traditional lectures to the flipped-blended hybrid model without courseware only slightly increased the student’s learning outcomes.

The Psychology of Sex and Gender course had previously been called Psychology of Women and used a different text. The instructor’s choice to switch to Acrobatiq courseware was largely driven by the need to have a textbook with more contemporary content and research, and

to have adaptive courseware with that title. The UCF Center for Distributed Learning and Pegasus Innovation Lab aided in identifying the courseware options and the ability to turn the Psychology of Sex and Gender e-textbook into courseware through AcrobaTIQ's SmartStart process was an advantage. This type of niche subject—much like the Microbe textbook—does not typically come with custom courseware, so the ability to create it was ideal. The learn by doing approach of courseware would also be beneficial for this textbook in particular, as historically the psychology students struggled more with the biology-intensive content included in this new textbook compared to the previous title. This Psychology course was taught entirely online with synchronous class sessions, but in a flipped blended model. Students were primarily third- and fourth-year students but also nearly 70% of students were transfer students from other smaller schools.

Courseware. The courseware used for both courses was generated using the SmartStart process (Dittel et al., 2018) that applies artificial intelligence to an e-textbook to transform it into courseware (Jerome et al., 2021). The development of automatic question generation systems has become a popular research area, given the broad potential for application in education (Kurdi et al., 2020), including for use as formative practice. This automatic question generation system uses the e-textbook to create the volume of formative practice needed to engage students in the learn by doing method. Two different types of questions are generated (as seen in Figure 1): fill-in-the-blank (FITB), where students must type in a missing term, and matching, where students drag and drop three available terms to the correct locations in a sentence. The FITB questions are a recall cognitive process dimension on Bloom's Taxonomy while the matching are a recognition type (Anderson et al., 2001), and both of these types have been long researched for their learning benefits (Andrew & Bird, 1938). The Psychology of Sex and Gender textbook (Bosson et al., 2019) was used for the SmartStart process and produced over 600 formative practice questions. The Microbe textbook (Swanson et al., 2016) was used to produce over 400 questions. Research on these automatically generated questions compared them to human-authored questions and found no meaningful difference in engagement, difficulty, and persistence performance metrics (Van Campenhout, Dittel, et al., 2021) as well as question discrimination (Johnson et al., 2022).

To take full advantage of other features of the courseware platform (i.e., predictive learning estimates and adaptivity), the initial courseware produced with SmartStart was further enhanced by the instructors. For the Psychology course, additional human-authored multiple choice and true/false questions were taken from the textbook ancillary material and added as formative practice. For Microbe, some additional questions were added from ancillary materials and some were written by the instructor. In addition to the research that did not indicate a difference in performance of the automatically generated and human-authored questions (Van Campenhout, Dittel, et al., 2021), the instructor noted that it became difficult to distinguish which questions were generated and which were written (with the exception of the extraordinarily difficult questions written specifically to challenge students). Adaptive activities were written for the most challenging chapters of content in both courses. Designed to scaffold based on each student's predictive learning estimate, the adaptive activities have been shown to improve outcomes (Van Campenhout et al., 2020). The faculty and instructional designers wrote the adaptive activities to assist students with the most challenging content where they would most need scaffolded support. Creating the adaptive components was also made more feasible in a short time because the bulk of the formative practice was automatically generated. Instead of

spending copious amounts of time writing and implementing the foundational reading comprehension questions, the instructors and instructional designers could focus on the scaffolding and conceptualization of the adaptive activities.

Implementation and Results

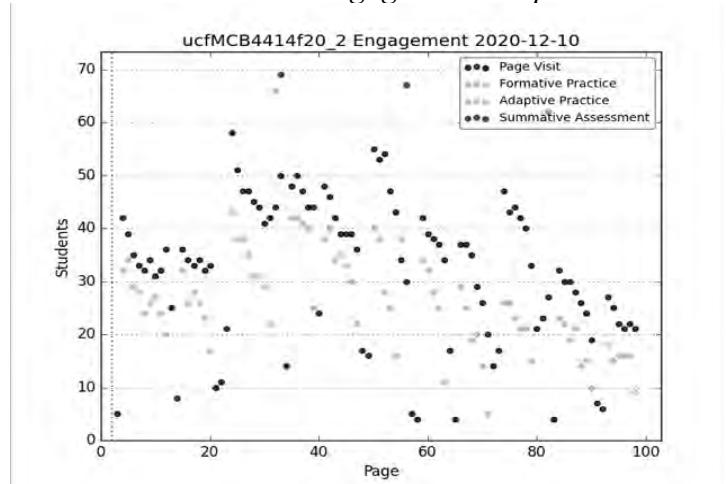
Microbe

The Microbe course had historically been taught as a face-to-face, lecture-style course attended by a large class of students as a part of their program of study. This approach is common and likely familiar to most faculty members. However, research has shown a significant benefit to student learning when a flipped-blended model is used for teaching and learning, where students use technology outside the class to learn the content and in-class time is led by instructors to expand student knowledge and provide feedback (Margulieux et al., 2015). Therefore, as the instructor changed to a hybrid (virtual and in-person) approach and flipped the classroom, the learning resource needed to provide more engagement for students and resources for the instructor.

The courseware was first used was the Fall of 2020. Students were assigned the courseware as they would have been assigned the e-textbook. During this first semester, the instructor did not provide any incentivization in the form of points or a grade for the courseware's "Learn by Doing" formative practice activities on the lesson pages (Figure 1), expecting that because students had the courseware, they would take advantage of it. Points were assigned in the first semester for completing the summative chapter quiz and the Personal Practice (adaptive activity) questions. Data gathered by the courseware platform were used to create engagement graphs that showed how many students (x-axis) read and did practice on each lesson page of the courseware (y-axis). In Figure 2, the fluctuation of the blue dots on the vertical axis means students chose to read some pages but not others in an inconsistent pattern. The lower red dots show that of the students who did read the lesson pages, only some of those students did the practice available. General attrition is seen over the course of the semester, which is typical as students use their learning resources less over time.

Figure 2

The Microbe Fall 2020 Engagement Graph.



Prior to the next semester, Spring of 2021, the instructor made some changes to how the courseware was implemented, with a goal of increasing student engagement. Thirty points were assigned to completing a minimum threshold of 80% of the formative practice in the courseware. This incentive had a visible change in student engagement, as seen in Figure 3. Student engagement was more consistent for both reading and doing throughout the courseware than the previous semester. While some students still did not do practice on pages, this proportion is visibly smaller than the previous semester.

Figure 3
The Microbe Spring 2021 Engagement Graph.

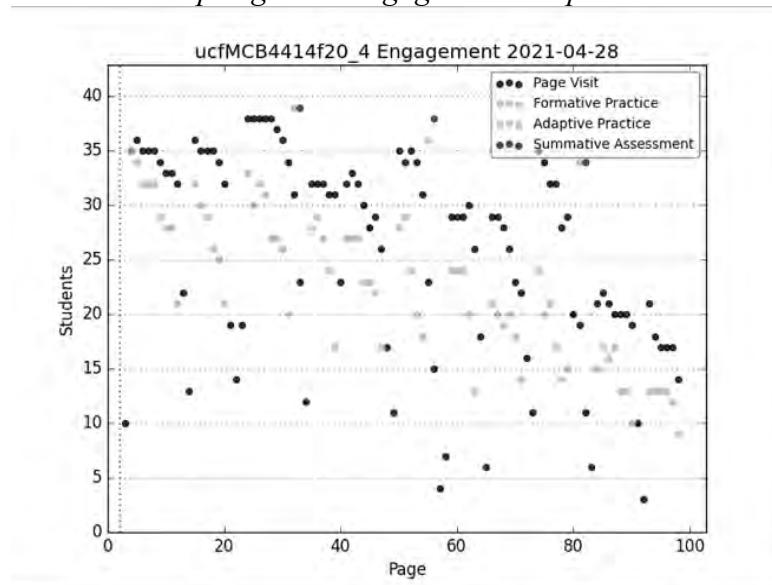


Table 1 presents the raw scores for the three exams and the final in the Microbes course. The Fall 2020 course was the largest of the semesters included, and first to use the courseware. The Spring 2021 course shows lower mean scores than the previous term. However, as a course that has been run every spring and fall term, the lower mean scores for the spring semester are consistent with prior spring courses and so therefore are not particularly unusual. The Fall 2021 course shows similar mean scores to those of Fall 2020 with the notable trend of higher low scores on ranges. When students took their time on the formative assessments, summative quizzes, and adaptive activities, students' exam scores and outcomes tended to improve. Unfortunately, not all students took their time and those who rushed through the assignments had exam scores that reflected this.

Table 1*Raw Exam Scores from the Microbial Metabolism Course from Fall 2020 to Fall 2021.*

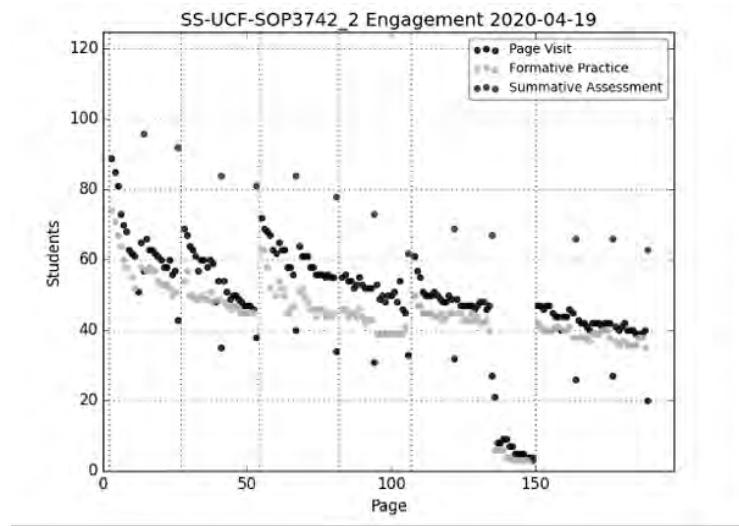
		Exam 1	Exam 2	Exam 3	Final
Fall 2020	Mean Score	82%	66%	69%	71%
	Score Range	48–96%	22–92%	28–96%	42–99%
	n Students	67	67	68	68
Spring 2021	Mean Score	77%	60%	60%	63%
	Score Range	52–94%	28–88%	26–86%	28–95%
	n Students	38	37	37	37
Fall 2021	Mean Score	81%	66%	70%	72%
	Score Range	60–98%	34–92%	32–96%	38–96%
	n Students	33	32	33	31

Students also verbalized to the instructor increased confidence on the exams after having done the courseware practice, a marked shift in attitude toward assessments from previous years. In addition, the quality of student engagement in class and online increased. Students asked more in-depth questions and responded to peers in discussion boards with more detail. Exams provide one measure of student learning, but the qualitative observation of student interactions both in virtual and in-person settings can also reveal a change in the depth of understanding students acquire.

Psychology

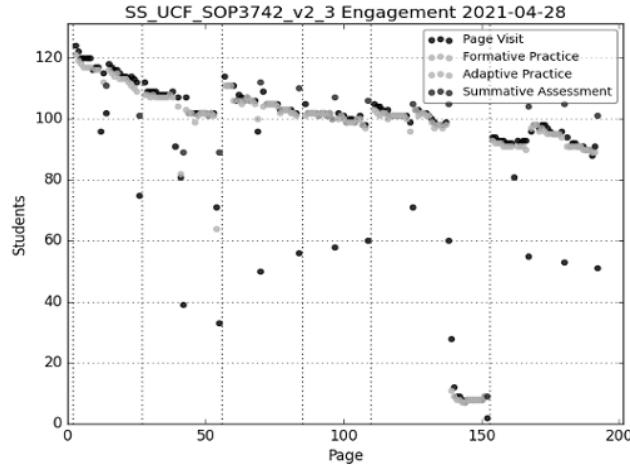
The Psychology of Sex and Gender course had been taught in an entirely online format with large sections, so the instructor was accustomed to teaching and learning with a flipped-blended format and interactive digital resources. The first section of the Psychology course taught with the courseware was in the Spring of 2020. The instructor assigned sections, posted reminders in the learning management system, and reminded students to do the courseware in class as well. In addition, two percentage points were assigned to completing a minimum of 85% of the practice in the course. The engagement graph (Figure 4) shows overall attrition in student engagement as the course progresses as well as within units, which, as noted previously, is typical engagement behavior. A small number of students read the pages but did not do the practice. The green summative assessment dots floating well above the nearest blue reading dots indicate a large portion of students entered the courseware only to take the assessments.

Figure 4
The Psychology Spring 2020 Engagement Graph.



The Psychology instructor similarly updated implementation practices for the following run of the course in the Spring of 2021. In addition to increased emphasis on the benefit of practice, 20% of the students' grade would be accounted for by completing a minimum of 85% of the practice. As seen in Figure 5, the engagement graph is closer to a horizontal line; the majority of students engaged with almost the entire courseware. The red practice dots are also next to or on top of the blue reading dots, which means nearly all students who read the pages also did the practice.

Figure 5
The Psychology Spring 2021 Engagement Graph.



In addition to increasing student engagement, exam scores also rose over the semesters. As seen in Table 2, the Fall 2019 semester—when only the e-textbook was used—shows the lowest mean scores for each exam. The Spring 2020 semester when the courseware was first used shows an increase in exam scores, while the Spring 2021 semester (when more points were

assigned and more students did the practice) shows even higher mean exam scores. As seen in the Microbe course, the lowest scores in the ranges also increased for each semester. These results strongly relate to the dramatic changes in student engagement visible in Figure 5.

Table 2

Student Exam Scores for Psychology from 2019 through 2021.

		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

Discussion

Students are best served when educators use learning technology as a tool. While advances in the learning sciences have led to more comprehensive and effective learning resources, they are best optimized by instructors for each specific learning context. The AI-generated courseware combines textbook content in niche subjects with large volumes of formative practice questions, enabling students to engage in learning by doing. The AI-generated courseware was beneficial to instructors first in that courseware for these niche titles did not exist previously, but second in that the automatically generated questions allowed them to spend their development time on targeted adaptivity. Years of teaching experience give faculty members unparalleled insight into where and how students struggle with content, and both instructors were able to apply their knowledge to tailor the courseware. Even the context of the teaching model and student characteristics were considered when instructors considered how to utilize the courseware in their teaching practice.

Iterative improvement is also a natural process of teaching and learning. Both faculty members believe in changing various components of teaching over time to better student outcomes—whether that is a new textbook, a new teaching model, or a new learning technology. The changes made in the implementation of the courseware over several semesters impacted student engagement and learning outcomes. Natural learning contexts differ greatly from controlled experiments; in these environments it is the instructor who is in the best position to identify the unique circumstances and provide the best conditions for learning. Iterative changes over time are an indication of student-centered teaching practices. It is also worth noting that though the addition of courseware as a learning resource and the changes in teaching practice were done to benefit students, faculty members also found that adding the courseware to a flipped-blended model had surprising advantages. Work-life balance improved and enjoyment of teaching increased. Research in teaching and learning often focuses more on learning, but future research should also focus on iterative improvements that can benefit instructor satisfaction as well. Given how impactful educators are on student engagement and outcomes, instructor perceptions and satisfaction deserve increased consideration.

Finally, this paper showcases the advances in teaching and learning research that can be identified when educational institutions and educational technology companies partner to share data and findings. The courseware platform provides enormous quantities of data for analysis that can advance what is known about the science of learning. Yet not all meaningful data can be collected via technology. Faculty members were able to identify qualitative changes in their students, such as increased engagement, preparedness, and satisfaction. Student-to-student interactions in online discussion boards improved, indicating that there are secondary benefits to a learn by doing approach outside of assessments. Instructors were also able to monitor data through dashboards and take steps to identify struggling students not previously available, which can have a meaningful impact on those specific students. Those same data dashboards identified where the entire class struggled with course content, which allowed the instructors to specifically target that content for additional explanation and practice. Combined data sources reveal how instructor implementation practices can change student engagement and scores. Further collaboration on future research could include investigating the impact of learning by doing for specific groups of students, the benefits of prerequisite adaptive testing, and methods of increasing instructor enjoyment and satisfaction during teaching and learning.

Acknowledgements

We would like to thank the UCF Pegasus Innovation Lab for their help identifying new teaching and learning solutions, the UCF instructional designers Joe Lloyd and Jessica Tojo for their work on enhancing the courseware, and Emily Svara for her assistance writing questions for Microbes.

Declarations

The authors declared no conflicts of interests.

The authors declared that they received no funding.

Ethics approval for this work was granted by the University of Central Florida.

References

- Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., May-er, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). Longman.
- Andrew, D. M., & Bird, C. (1938). A comparison of two new-type questions: recall and recognition. *Journal of Educational Psychology*, 29(3), 175–193.
<https://doi.org/10.1037/h0062394>
- 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. 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>
- Baker, R. S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. In G. Veletsianos (Ed.), *Emergence and innovation in digital learning* (pp. 83–98). Athabasca, Alberta: AB University Press. <https://doi.org/10.15215/aupress/9781771991490.01>
- Black, P., & William, D. (2010). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90.
<https://doi.org/10.1177/003172171009200119>
- Bosson, J. K., Vendello, J. A., & Buckner, C. V. (2018). *The psychology of sex and gender* (1st ed.). Thousand Oaks, CA: Sage Publications.
- Carvalho, P. F., McLaughlin, E. A., & Koedinger, K. R. (2017). Is there an explicit learning bias? Students beliefs, behaviors and learning outcomes. In G. Guzelmann, et al. (Eds.) *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 204–209). <https://cogsci.mindmodeling.org/2017/papers/0050/index.html>
- 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*. VitalSource Technologies.
- Dunlosky, J., Rawson, K., Marsh, E., Nathan, M., & Willingham, D. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58.
<https://doi.org/10.1177/1529100612453266>

- Ertmer, P. A., & Newby, T. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 24(3), 55–76. <https://doi.org/10.1002/piq.21143>
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 717–741. <https://doi.org/10.1007/s10648-015-9348-9>
- Fitzpatrick, L., & McConnell, C. (2008). Student reading strategies and textbook use: An inquiry into economics and accounting courses. *Research in Higher Education Journal*, 3, 1–10.
- Fullan, M., & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research*, 47(2), 335–397.
<https://doi.org/10.3102/00346543047002335>
- Goldstein, P. J., & Katz, R. N. (2005). Academic analytics: The uses of management information and technology in higher education. *Educause*. <https://library.educause.edu/-/media/files/library/2005/12/ers0508w-pdf.pdf>
- 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. ISSN: 2189-1036 (pp. 201–210). <https://doi.org/10.22492/issn.2189-1036.2022.19>
- Jerome, B., Van Campenhout, R., & Johnson, B. G. (2021). Automatic question generation and the SmartStart application. *Proceedings of the Eighth ACM Conference on Learning@Scale* (pp. 365–366). <https://doi.org/10.1145/3430895.3460878>
- Johnson, B. G., Dittel, J. S., Van Campenhout, R., & Jerome, B. (2022). Discrimination of automatically generated questions used as formative practice. *Proceedings of the Ninth ACM Conference on Learning@Scale* (pp. 325–329).
<https://doi.org/10.1145/3491140.3528323>
- 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., Kim, J., Jia, J., McLaughlin, E., & Bier, N. (2015, March). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. *Proceedings of the Second ACM Conference on Learning@Scale* (pp. 111–120).
<https://doi.org/10.1145/2724660.2724681>
- Koedinger, K. R., McLaughlin, E. A., Jia, J. Z., & Bier, N. L. (2016, April). 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 & Knowledge* (pp. 388–397).
<http://dx.doi.org/10.1145/2883851.2883957>

- Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30, 121–204. <https://doi.org/10.1007/s40593-019-00186-y>
- 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*, 2008(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), 33–84. <https://doi.org/10.3102/0034654307313793>
- Piaget, J. (1926). *The language and thought of the child*. Kegan Paul, Trench, Trubner and Company.
- 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, Spain.
- Schroeder, K., Hubertz, M., Johnson, B. G., & Van Campenhout, R. (2021). *Courseware at scale: Using artificial intelligence to create learning by doing from textbooks*. Presented at OLC Accelerate, Washington D.C., Oct. 6, 2021.
- Singer, S. R., & Bonvillian, W. B. (2013). Two revolutions in learning. *Science*, 339(6126), 1359. <https://doi.org/10.1126/science.1237223>
- Holiday, T., O'Sullivan, P., & Adams, S. (2020). *Adaptive courseware implementation guide* (2nd ed.). Every Learner Everywhere. <https://www.everylearnereverywhere.org/resources/adaptive-courseware-implementation-guide>
- Swanson, M., Reguera, G., Schaechter, M., & Neidhardt, F. (2016). *Microbe* (2nd ed.). ASM Press.

Van Campenhout, R., Dittel, J. S., Jerome, B., & Johnson, B. G. (2021). Transforming textbooks into learning by doing environments: An evaluation of textbook-based automatic question generation. *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., 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. (2021). 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* (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. (2022). The doer effect: Replication and comparison of correlational and causal analyses of Learning. *International Journal on Advances in Systems and Measurements, 15(1&2)*, 48–59. http://www.iarajournals.org/systems_and_measurements/tocv15n12.html

Van Campenhout, R., & Kessler, A. (2022). *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. ISSN: 2308-4367* (pp. 29–34). https://www.thinkmind.org/index.php?view=article&articleid=elml_2022_2_40_58010

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>