UNIVERSITY OF BATH

Citation for published version:
St-Hilaire, F, Burns, N, Belfer, R, Shayan, M, Smofsky, A, Vu, DD, Frau, A, Potochny, J, Faraji, F, Pavero, V, Ko, N, Ching, AO, Elkins, S, Stepanyan, A, Matajova, A, Charlin, L, Bengio, Y, Serban, IV & Kochmar, E 2021, 'A Comparative Study of Learning Outcomes for Online Learning Platforms', Paper presented at The 2021 conference on Artificial Intelligence in Education, Utrecht, Netherlands, 14/06/21 - 18/06/21.

Publication date:
2021

Document Version
Peer reviewed version

Link to publication

Publisher Rights
CC BY

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

A Comparative Study of Learning Outcomes for Online Learning Platforms

Francois St-Hilaire[1], Nathan Burns[1], Robert Belfer[1], Muhammad Shayan[1], Ariella Smofsky[1], Dung Do Vu[1], Antoine Frau[1], Joseph Potochny[1], Farid Faraji[1], Vincent Pavero[1], Neroli Ko[1], Ansona Onyi Ching[1], Sabina Elkins[1], Anush Stepanyan[1], Adela Matajova[1], Laurent Charlin[1,2], Yoshua Bengio[1,2], Iulian Vlad Serban[1], and Ekaterina Kochmar[1,3]

[1] Korbit Technologies Inc., Canada
[2] Quebec Artificial Intelligence Institute (Mila), Canada
[3] University of Bath, United Kingdom

Abstract. Personalization and active learning help educational systems to close the gap between students with varying abilities. We run a comparative head-to-head study of learning outcomes for two popular online platforms: `Platform A`, which delivers content over lecture videos and multiple-choice quizzes, and `Platform B`, which provides interactive problem-solving exercises and personalized feedback. We observe a statistically significant increase in the learning outcomes on `Platform B`. Further, the results of the self-assessment questionnaire suggest that participants using `Platform B` improve their metacognition.

Keywords: Online and distance learning · Models of Teaching and Learning · Intelligent and Interactive Technologies · Data Science

1 Introduction

We investigate the learning outcomes induced by two popular online learning platforms in a comparative head-to-head study. `Platform A` is a widely-used platform that follows a traditional model, where students learn by watching lecture videos, reading, and testing their knowledge with multiple choice quizzes. In contrast, `Platform B`[4] focuses on personalized, active learning approach with problem-solving exercises [36]. `Platform B` is powered by an AI tutor, which alternates between lecture videos and interactive problem-solving exercises. The AI tutor shows students problem statements and students attempt to solve them. Each incorrect attempt is addressed with personalized pedagogical interventions tailored to student's needs and misconceptions (see Figure 1).

In this study, we formulate and test the following hypothesis:

> **Hypothesis**: Participants studying with `Platform B` have higher learning gains than those studying with `Platform A`, because `Platform B` employs personalized, active learning and problem-based learning and provides a wider and more personalized set of pedagogical elements to its students.

[4] `Platform B` is the Korbit learning platform available at www.korbit.ai.

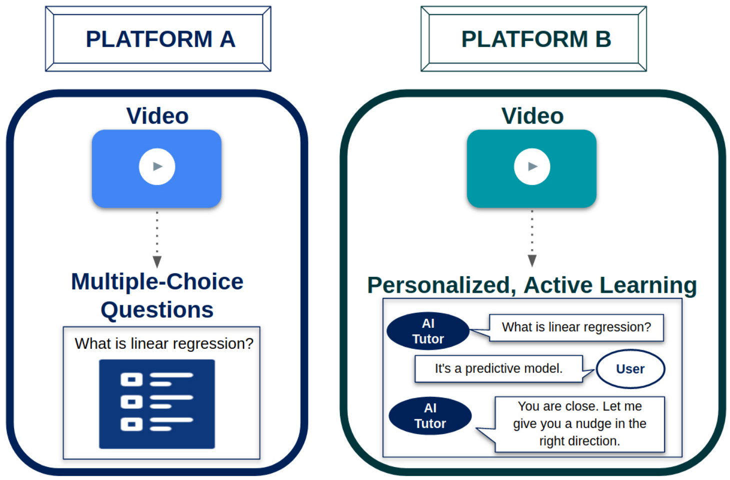

Fig. 1. `Platform A` follows a traditional learning approach utilizing videos and multiple choice quizzes, while `Platform B` uses a personalized, active learning approach with problem-solving exercises.

2 Related Work

Online learning platforms have the capability of bridging the gap and addressing inequalities in society caused by uneven access to in-person teaching [13, 16, 18, 32, 41, 45]. The current `COVID-19` pandemic further exacerbates the need for high quality online education being accessible to a wide variety of students [1, 4, 30].

Nevertheless, the efficacy of online and distance learning has been challenged by researchers: specifically, it may be hard to address the differences in students' learning needs, styles and aptitudes on such platforms [9, 15, 39, 42]. This calls for approaches that can be adapted and personalized to the needs of each particular student. Studies confirm that personalization is key to successful online learning [28, 35], as it can maximize the learning benefits for each individual student [48]. In addition, problem-solving has been shown to be a highly effective approach for learning in various domains [12, 19, 20, 46, 47]. Such problem-solving and active learning activities can be addressed by intelligent tutoring systems, which are also capable of giving personalized feedback and explanations and incorporating conversational scaffolding [2, 7, 8, 12, 21, 23, 26, 27, 29, 33, 34].

In contrast to previous studies investigating learning outcomes with intelligent tutoring systems, in this study the AI-powered learning platform, `Platform B`, is a fully-automated system based on machine learning models [36]. The system is trained from scratch on educational content to generate automated, personalized feedback for students and has the ability to automatically generalize to new subjects and improve as it interacts with new students [37, 38].

To evaluate the impact of educational technology and online learning platforms on student learning outcomes, we follow previous research [3, 11, 17, 24, 25, 31, 40, 43]. We adopt the well-established pre-/post-assessment framework, where students are split into intervention groups and their knowledge of the subject is evaluated before and after their assigned intervention. Further, we measure student's metacognition. Students' ability to self-assess and develop self-regulation skills plays a crucial role in online learning [17, 27], though studies show that students struggle to evaluate their own knowledge and skills level [5, 6, 10].

3 Experimental Setup

48 participants were randomly divided between the two platforms, where the first group was asked to study the course from `Platform A` and the second from `Platform B`. Each group completed a 3-hour long course on *linear regression*. The majority fall into our target audience of undergraduates (89.6%) studying disciplines not centered around mathematics (e.g. health sciences).

Linear regression was selected as the topic of study since it is one of the most fundamental topics, that is covered early on in any course on machine learning and data science, and the material covering this topic on both platforms is comparable. To ensure a fair comparison, extra care was taken to ensure that the courses and the subtopics they covered were as similar as possible.

The study ran over a 4-day period with strict deadlines and detailed instructions set for the participants. All participants were required to take an assessment quiz on linear regression before the course (*pre-quiz*) and another one after the course (*post-quiz*). The quizzes contained 20 multiple-choice questions each and were equally adapted to both courses, with questions in pre- and post-quizzes isomorphically paired. Using pre- and post-quiz scores, we measure *learning gains* to quantify how effectively each participant has learned. A student's learning gain g is estimated as the difference between their pre-quiz (pre_score) and post-quiz ($post\_score$) scores. Further, a student's normalized learning gain g_{norm} is calculated by:

$$g_{norm} = \frac{post\_score - pre\_score}{100\% - pre\_score} \tag{1}$$

4 Results and Discussion

25 participants completed the course on `Platform A` and 23 on `Platform B`. Average learning gains are shown in Figure 2 for the two platforms. The average normalized learning gains for `Platform B` participants are 49.24% higher than for `Platform A` participants, with the difference being statistically significant at a 90% confidence level (p=0.068 w.r.t. one-sided t-test). Average raw learning gains for `Platform B` participants are 70.43% higher than for `Platform A` participants, with the difference being statistically significant at a 95% confidence level (p=0.038 w.r.t. one-sided t-test). Overall, our hypothesis that learning outcomes are higher for participants on `Platform B` than on `Platform A` is confirmed.

We estimate that participants on `Platform B` spent at least twice as much time doing active learning (problem-solving exercises) compared to participants on `Platform A`, although the total average study times on the two platforms

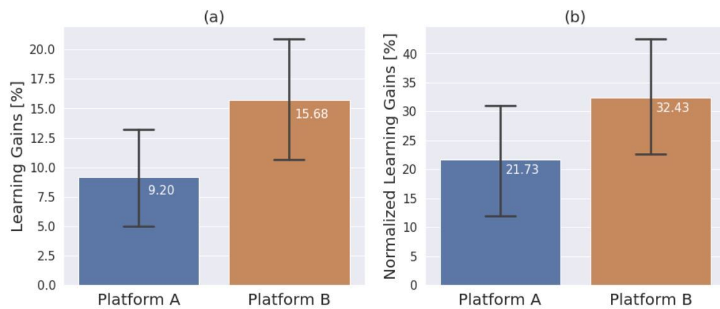

Fig. 2. (a) Average learning gains g with 95% confidence intervals.* (b) Average normalized learning gains g_{norm} with 95% confidence intervals.** Here * and ** indicate a statistically significant difference at 95% and 90% confidence level respectively.

were equivalent. We further observed that the rate of correct answers on the first try positively correlates with both learning gains (r=0.44) and post-quiz results (r=0.46), and the number of exercises completed positively correlates with the post-quiz score (r=0.28), suggesting that participants who spent more time on active learning performed better and, as a result, obtained higher post-quiz scores and learning gains.

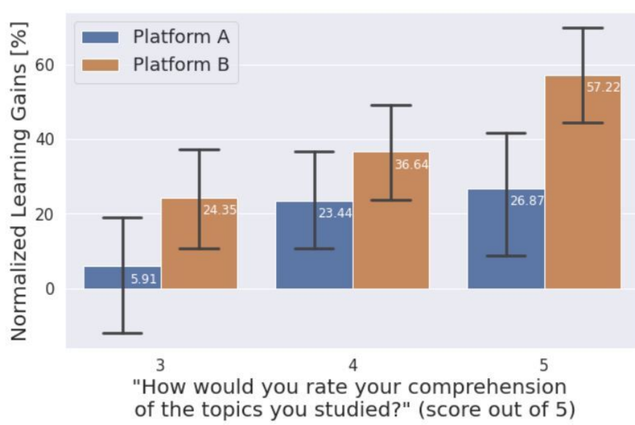

Fig. 3. Normalized learning gains for each self-assessed comprehension rating with 95% confidence intervals. Only 1 participant gave a score lower than 3 (not shown here).

Finally, we evaluated meta-cognitive aspects related to the students' learning experience with the two platforms using a questionnaire. In particular, students were asked the question *"How would you rate your comprehension of the topics you studied?"*. As shown in Figure 3, it appears that `Platform B` not only induced overall higher learning gains, but also gave participants a more accurate understanding of their knowledge level and helped improve their meta-cognition.

References

1. Adedoyin, O. B., Soykan, E.: Covid-19 pandemic and online learning: the challenges and opportunities. Interactive Learning Environments, 1–13 (2020)
2. Albacete, P., Jordan, P., Katz, S., Chounta, I.A., McLaren, B.M.: The Impact of Student Model Updates on Contingent Scaffolding in a Natural-Language Tutoring System. In: International Conference on Artificial Intelligence in Education, pp. 37–47. Springer (2019)
3. Barokas, J., Ketterl, M., Brooks, C., Greer, J.: Lecture capture: Student perceptions, expectations, and behaviors. In: J. Sanchez & K. Zhang (Eds.), World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Educatio, pp. 424–431 (2010)
4. Basilaia, G., Kvavadze, D.: Transition to online education in schools during a SARS-CoV-2 coronavirus (COVID-19) pandemic in Georgia. Pedagogical Research, **5**(4) (2020)
5. Brown, G. T. L., Harris, L. R.: Student self-assessment. In J. H. McMillan (Ed.). The SAGE handbook of research on classroom assessment (pp. 367–393). Thousand Oaks, CA: Sage (2013)
6. Brown, G. T., Andrade, H. L., Chen, F.: Accuracy in student self-assessment: directions and cautions for research. Assessment in Education: Principles, Policy & Practice, **22**(4), 444–457 (2015)
7. Büdenbender, J. and Frischauf, A. and Goguadze, G. and Melis, E. and Libbrecht, P. and Ullrich, C.: Using computer algebra systems as cognitive tools. International Conference on Intelligent Tutoring Systems, 802–810 (2002)
8. Chi, M., Koedinger, K., Gordon, G., Jordan, P., Vanlehn, K.: Instructional Factors Analysis: A Cognitive Model For Multiple Instructional Interventions. In: EDM 2011 - Proceedings of the 4th International Conference on Educational Data Mining, pp. 61–70 (2011)
9. Coffield, F. J., Moseley, D. V., Hall, E., Ecclestone, K.: Learning Styles for Post 16 Learners: What Do We Know? London: Learning and Skills Research Centre/University of Newcastle upon Tyne (2004)
10. Crowell, T. L.: Student Self Grading: Perception vs. Reality. American Journal of Educational Research **3**(4), 450–455 (2015). https://doi.org/10.12691/education-3-4-10
11. Demmans Epp, C., Phirangee, K., Hewitt, J., Perfetti, C. A.: Learning management system and course influences on student actions and learning experiences. Educational Technology, Research and Development (ETRD), **68**(6), 3263–3297 (2020)
12. Fossati, D., Di Eugenio, B., Ohlsson, S., Brown, C., Chen, L.: Data driven automatic feedback generation in the iList intelligent tutoring system. Technology, Instruction, Cognition and Learning, **10**(1), 5–26 (2015)
13. Graesser, A., VanLehn, K., Rose, C., Jordan, P., Harter, D.: Intelligent tutoring systems with conversational dialogue. AI Magazine, **22**(4), 39–51 (2001)
14. Hake, R. R.: Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American journal of Physics **66**(1), 64–74 (1998)
15. Honey, P., Mumford A.: The Manual of Learning Styles. Peter Honey Publications, Maidenhead (1992)
16. Hrastinski, S., Stenbom, S., Benjaminsson, S., Jansson, M.: Identifying and exploring the effects of different types of tutor questions in individual online synchronous tutoring in mathematics. Interactive Learning Environments, 1–13 (2019)

17. Kashihara, A., Hasegawa, S.: A model of meta-learning for web-based navigational learning. International Journal of Advanced Technology for Learning, **2**(4), 198–206 (2005)
18. Koedinger, K., Corbett, A.: Cognitive tutors: Technology bringing learning sciences to the classroom. R.K. Sawyer (Ed.), The Cambridge handbook of the learning sciences, Cambridge University Press, New York, NY, pp. 61–78 (2006)
19. Kolb, D. A.: Experiential learning experience as the source of learning and development, New Jersey, Prentice Hall (1984)
20. Kumar, A. N.: Results from the evaluation of the effectiveness of an online tutor on expression evaluation. ACM SIGCSE Bulletin **37**(1), 216–220 (2005)
21. Kumar, A.N.: Generation of problems, answers, grade, and feedback—case study of a fully automated tutor. Journal on Educational Resources in Computing (JERIC), **5**(3), 3–es (2005)
22. Lehman, B., D'Mello, S., Graesser, A.: Confusion and complex learning during interactions with computer learning environments. The Internet and Higher Education **15**(3), 184–194 (2012)
23. Lin, C. F., Yeh, Y. C., Hung, Y. H., Chang, R. I.: Data mining for providing a personalized learning path in creativity: An application of decision trees. Computers & Education **68**, 199–210 (2013). https://doi.org/https://doi.org/10.1016/j.compedu.2013.05.009.
24. Ma, W., Adesope, O. O., Nesbit, J. C., Liu, Q.: Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis, Journal of Educational Psychology (2014)
25. Mark, M.A. and Greer, J.E.: Evaluation methodologies for intelligent tutoring systems. Journal of Artificial Intelligence in Education, **4**, 129–129 (1993)
26. Melis, E., Siekmann, J.: ActiveMath: An Intelligent Tutoring System for Mathematics. Artificial Intelligence and Soft Computing, 91–101 (2004)
27. Munshi, A., Biswas, G.: Personalization in OELEs: Developing a Data-Driven Framework to Model and Scaffold SRL Processes. In: International Conference on Artificial Intelligence in Education, pp. 354–358. Springer (2019)
28. Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., Melis, E.: Exploring feedback and student characteristics relevant for personalizing feedback strategies. Computers & Education, 71, 56–76 (2014)
29. Nye, B. D., Graesser, A. C., Hu, X.: AutoTutor and family: A review of 17 years of natural language tutoring. International Journal of Artificial Intelligence in Education, **24**(4), 427–469 (2014)
30. Onyema, E. M., Eucheria, N. C., Obafemi, F. A., Sen, S., Atonye, F. G., Sharma, A., Alsayed, A. O.: Impact of Coronavirus pandemic on education. Journal of Education and Practice, **11**(13), 108–121 (2020)
31. Penstein, C. R., Moore, J., VanLehn, K., Allbritton, D.: A Comparative Evaluation of Socratic versus Didactic Tutoring. Annual Meeting of the Cognitive Science Society, 23 (2001)
32. Psotka, J., Massey, D., Mutter, S.: Intelligent tutoring systems: Lessons learned. Lawrence Erlbaum Associates, Hillsdale, NJ (1988)
33. Rus, V., Stefanescu, D., Baggett, W., Niraula, N., Franceschetti, D., Graesser, A.C.: Macro-adaptation in conversational intelligent tutoring matters. In: International Conference on Intelligent Tutoring Systems, pp. 242–247. Springer (2014)
34. Rus, V., Stefanescu, D., Niraula, N., Graesser, A.C.: DeepTutor: towards macro- and micro-adaptive conversational intelligent tutoring at scale. In: Proceedings of the first ACM conference on Learning@ Scale conference, pp. 209–210 (2014)
35. Sampson, D., Karagiannidis, C.: Personalised learning: educational, technological and standarisation perspective. Digital Education Review, (4), 24–39 (2002)

36. Serban, I. V., Gupta, V., Kochmar, E., Dung, Vu, D. D., Belfer, R., Pineau, J., Courville, A., Charlin, L., Bengio, Y.: A Large-Scale, Open-Domain, Mixed-Interface Dialogue-Based ITS for STEM. In: International Conference on Artificial Intelligence in Education, pp. 387–392. Springer, Cham (2020)
37. Kochmar, E., Do Vu, D., Belfer, R., Gupta, V., Serban, I. V., Pineau, J. Automated Personalized Feedback Improves Learning Gains in An Intelligent Tutoring System. In International Conference on Artificial Intelligence in Education (pp. 140-146). Springer, Cham.
38. Grenander, M., Belfer, R., Kochmar, E., Serban, I. V., St-Hilaire, F., Cheung, J. C. (2021). Deep Discourse Analysis for Generating Personalized Feedback in Intelligent Tutor Systems. In The 11th Symposium on Educational Advances in Artificial Intelligence (2021)
39. Stash, N. V., Cristea, A. I., De Bra, P. M.: Authoring of learning styles in adaptive hypermedia: problems and solutions. In: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, pp. 114–123 (2004)
40. Tan, Y., Quintana, R. M.: What can we learn about learner interaction when one course is hosted on two MOOC platforms? In: Companion Proceedings to the International Conference on Learning Analytics and Knowledge (LAK), pp. 155–156 (2019)
41. Tomkins, S., Ramesh, A., Getoor, L.: Predicting Post-Test Performance from Online Student Behavior: A High School MOOC Case Study. International Educational Data Mining Society (2016)
42. VanLehn, K., Graesser, A.C., Jackson, G., Jordan, P., Olney, A., Rose, C.P.: When are tutorial dialogues more effective than reading? Cognitive Science, **31**(1), 3–62 (2007)
43. VanLehn, K.: The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems, Educational Psychologist (2011)
44. Vigentini, L., Clayphan, A.: Pacing through MOOCs: course design or teaching effect? In: EDM, pp. 572–573 (2015)
45. Wang, Y., Paquette, L., Baker, R.: A longitudinal study on learner career advancement in MOOCs. Journal of Learning Analytics, **1**(3), 203–206 (2014)
46. Wood, D., Wood, H.: Vygotsky, Tutoring and Learning. Oxford review of Education **22**(1), 5–16 (1996)
47. Woolf, B.: Building intelligent interactive tutors Morgan Kaufmann Publishers, Burlington, MA (2009)
48. Yin, B., Patikorn, T., Botelho, A. F., Heffernan, N. T.: Observing personalizations in learning: Identifying heterogeneous treatment effects using causal trees. In Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale, pp. 299–302 (2017)
49. Zhang, Y., Paquette, L., Baker, R. S., Ocumpaugh, J., Bosch, N., Munshi, A., Biswas, G.: The relationship between confusion and metacognitive strategies in Betty's Brain. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, pp. 276–284 (2020)