Platform-based Adaptive Experimental Research in Education: Lessons Learned from Digital Learning Challenge

Ilya Musabirov[1], Mohi Reza[1], Steven Moore[2], Pan Chen[1], Harsh Kumar[1], Tong Li[1], Haochen Song[1], Jiakai Shi[1], Koby Choy[1], Thomas Price[3], John Stamper[2], Norman Bier[2], Nina Deliu[4], Sofia Villar[5], Anna Rafferty[6], Audrey Durand[7], Joseph Jay Williams[1]

[1] University of Toronto, [2] Carnegie Mellon University, [3] North Carolina State University, [4] Sapienza Università di Roma, [5] University of Cambridge, [6] Carleton College, [7] Université Laval

ABSTRACT: We report on our experience with a real-world, multi-experimental evaluation of an adaptive experimentation platform within the XPRIZE Digital Learning Challenge framework. We showcase how EASI (Experiment as a Service) cross-platform software supports quick integration and deployment of adaptive experiments as well as five systematic replications within a 30-day timeframe. The outline the key scenarios of the applicability of platform-supported experiments and reflect on lessons learned from this two-year project that can help researchers and practitioners to integrate adaptive experiments in real-world courses.

Keywords: adaptive experiments, posterior sampling, experimentation platforms

1 INTRODUCTION

In this paper, we report on our experience with a multi-experiment field deployment for evaluating an adaptive experimentation platform within the XPRIZE Digital Learning Challenge (DLC) framework, which took place between March 2022 and 2023. We show how we used our Experiment as a Service (EASI) cross-platform software infrastructure for experimentation to conduct and systematically replicate five experiments within 30 days. EASI has been used for a diverse range of over 250 traditional experiments to date. Unique strengths include a range of existing random assignment methods and usage across many different settings, whereas most tools are overly specific to the context/platform of creation. Designed for interoperability, EASI has been used in different digital learning platforms (edX, Coursera, Moodle, Canvas, ASSISTments), and can be integrated with any LTI (Learning Tools Interoperability) compliant LMS. EASI provides access to a library of machine learning algorithms and statistical methods (such as Bayesian inference) for analyzing data in real time. This approach offers flexibility in changing how conditions are assigned to future students and how instructors and students can be involved in classroom experimentation (Reza et al., 2021). We report the key lessons on platform-supported adaptive experiments learned from our work in the XPRIZE DLC[1].

[1] https://www.xprize.org/challenge/digitallearning

2 ADAPTING EXPERIMENTS PROPORTIONAL TO UNCERTAINTY

Instructors have reasonable concerns about experiments being fair and assigning students to a worse condition. When one experiment provides some evidence for a difference in conditions, instructors may be even more reluctant to do replications to better understand the design choices that led to an intervention's effectiveness or how underrepresented learner groups benefited. However, it is practically and scientifically essential to become increasingly certain that our intervention is effective, or test out potentially better ideas. We believe the apparent dichotomy—to experiment or not—is better framed as a fundamental statistics and Machine Learning (ML) tradeoff between 'exploring' (collecting data to converge on the best action) and 'exploiting' (exploiting/using the data to change which actions we choose).

Traditional experiments assign conditions uniformly, while adaptive experiments adjust assignment probabilities based on the latest evidence. EASI turns any single deployment into a rapid sequence of replications since data from even the past one, five, or ten students can gradually adjust the design of the experiment. EASI uses a Posterior Sampling Algorithm (Chapelle & Li, 2011) to analyze individual participant data, updating the likelihood of assigning conditions based on student data collected from a course. When dealing with distinct student groups, Contextual Posterior Sampling is used to determine the optimal condition for each group of students. Over time, as more student data accumulates, the approach shifts from a standard experimental split (i.e. 50/50) toward a more personalized one for each student group. This algorithm enables customized interventions for each group, enhancing their learning results.

3 FIELD DEPLOYMENT: XPRIZE DIGITAL LEARNING CHALLENGE (DLC)

During the IES-sponsored DLC, we aimed to demonstrate the adaptive approach to experimentation and our platform capabilities in the rapid multi-replication of educational interventions for different student demographics. To achieve this, we integrated EASI with the Open Learning Initiative (OLI) (Bier et al., 2023), which is designed to support robust experimentation at scale, in collaboration with institutions that are already using OLI courseware. The institutions include R1 universities and community colleges. Instructors of these courses had previously used the OLI platform for one to three semesters prior, without EASI integrated into it. We reached out to them for their consent and, with IRB approval, integrated EASI into these courses. We performed one pilot and five replication studies in distinct courses. These courses are offered across five diverse institutions, serving 2295 enrolled students in several domains, from Anatomy to Statistics.

Our iterative design approach involved an interdisciplinary working group of EASI and OLI developers, machine learning, learning science, and engineering researchers in collaboration on all stages of the DLC, focusing on eliciting critical scenarios of using adaptive experimentation for continuous improvement in the set of diverse courses. Three team members were also course instructors with extensive experience using OLI, while another two had expertise in other LMS.

Tu support deployment in a diverse set of courses, we chose intervention cases which, with the help of EASI and OLI, were designed as loosely coupled with the course content and easily portable, allowing us as well as researchers and instructors to rapidly replicate them in any new course, using existing course content. These interventions were designed for a common formal education context where students independently work through an online textbook containing short passages and videos with content knowledge. At the end of each section of the textbook, students engage in a variety of activities to promote learning. We aimed the first intervention case at the motivational domain, encouraging students to participate in optional course activities, randomizing Growth Mindset, and Self/Peer-Focused Framing, and using engagement outcomes. The second intervention aimed to provide students with retrieval practice prompts tied to course activities and used accuracy on the following problem as a proximal outcome (algorithm reward). We used these designs to discuss how adaptive interventions can help to explore and replicate the impact of novel variations of existing interventions, precisely targeting various outcomes of interest (e.g., assessments and participation in learning activities).

4 LESSONS LEARNED

In the iterative design of experiment and replications, we elicited **key adaptive scenarios** that can accelerate analysis and action using Bayesian statistics and ML algorithms: **(S1)** dynamically adding a new condition to expand from a two- into a three-condition experiment; **(S2)** how a three-condition study can assign more students to the most effective conditions while increasing statistical power to identify which ones are the best; and **(S3)** balancing practical impact with scientific insight, by helping a majority group get conditions that are better on average (for that group), while still collecting data to personalize so that a statistical minority is not unfairly receiving a condition that is worse for them. These features help researchers and instructors interested in course improvement to conduct a broad range of adaptive experiments. They can target micro-level objectives, such as enabling specific actions or improving a particular course element, or macro-level goals, such as improving overall course outcomes. Moreover, the dynamic assignment of students to better conditions decreases the decision-making burden on the course team, while allowing for improved student outcomes. It also has the potential to increase statistical power while better discriminating between conditions in multi-armed cases.

Another critical focus should be showing the consequences of particular adaptive experimental patterns in the early design stages using **statistical simulations**. During the DLC, we prototyped modules aimed to let instructors/researchers take data from one experiment and use it to specify alternative Scenarios for what the effects of conditions might be in different replications. That helps researchers simulate collecting and analyzing data from thousands of repeated runs of an experiment under different scenarios for what the effects could be and what might mediate these effects, such as student characteristics. This allows researchers and instructors to specify precisely and explore different kinds of effects they could discover in future replications of their experiment and understand what impact the particular adaptive experimental design can achieve compared to traditional approaches. Another related requirement is providing a set of custom **data visualizations and data analysis workflows** tied to the experimental design. This avoids potential issues arising from the application of unsuitable or suboptimal analytical methods. It allows us to understand not only what we have learned – causal effects, but also how we did – the impact of adaptation and personalization on students.

We illustrate these two connected tools in the simulated example, based on **S3**, with two groups of students showing higher and lower accuracy. In this scenario, students are encouraged to contribute their own questions to the course bank using two approaches: Self-Focused condition (focused on utility for student's learning) and Peer-Focused condition (encouraging sharing knowledge with others). Overall, it looks like the Peer-Focused condition is better than the Self-Focused condition (z= 4.58. p<0.0001, Cohen's w 0.12). However, a closer analysis would reveal that for a statistical minority (20% of students with Lower Accuracy), the opposite is true: the Peer-Focused condition is worse than the Self-Focused condition, with a larger effect size (z= 3.52. p=0.0004, Cohen's w 0.2). The Self-Focused message being better for Lower Accuracy students is obscured by the fact that 80% of the students are in the Higher Accuracy group, where the opposite is true: The Peer-Focused condition is better than the Self-Focused condition (z=6.91. p<0.0001, Cohen's w 0.2). This effect in the 80% Higher Accuracy students drives the overall average positive effect, although giving everyone the Peer-Focused prompts is harmful to Lower Accuracy students. A suitable adaptive intervention template can automatically account for this crossover interaction. In many similar scenarios the potential of adaptivity and/or personalization needs to be well-communicated, and experimental platforms need to support their users in exploring and visualizing them early on to make informed decisions about intervention designs, especially from the equity perspective.

The last direction s providing researchers and course teams with **pre-defined templates for adaptive experimentation**, capturing potential decision points to intervene in the course, meaningful outcomes to use in the adaptive experiment, a template and examples for the content part of the intervention, and potential bandit designs (e.g. three-arm non-contextual, contextual based on previous performance) with their impact.

5 CONCLUSION

Our work during the DLC has emphasized the challenges of experimentation in post-secondary settings, particularly as educators adapt and modify their planned instructional activities. The ability to flexibly adjust to these types of changes is an essential characteristic of any educational experimentation platform. We highlight the advantages of adaptive experiments: the ability to adjust experiments on the fly, based on real-time data, can lead to more efficient and effective research outcomes, ultimately helping to accelerate progress in the field of education. By leveraging platforms like EASI, researchers, and educators can gain new insights into the most effective teaching and learning strategies, paving the way for improved student outcomes and a brighter future for all. This work was partially supported by the National Science Foundation (#2209819).

REFERENCES

Bier, N., Stamper, J. , Moore S., Siegel, D., & Anbar, A. (2023). OLI Torus: a next-generation, open platform for adaptive courseware development, delivery, and research. Companion Proceedings of the 13th International Learning Analytics and Knowledge Conference.

Chapelle, O., & Li, L. (2011). An Empirical Evaluation of Thompson Sampling. Advances in Neural Information Processing Systems, Vol. 24. Curran Associates, Inc.

Koedinger, K. R., Stamper, J. C., Leber, B., & Skogsholm, A. (2013). LearnLab's DataShop: A Data Repository and Analytics Tool Set for Cognitive Science. Top. Cogn. Sci., 5(3), 668-669.

Reza, M., Kim, J., Bhattacharjee, A., Rafferty, A., & Williams, J. (2021). The MOOClet Framework: Unifying Experimentation, Dynamic Improvement, and Personalization in Online Courses. Proceedings of the Eighth ACM Conference on Learning @ Scale.