

Fifth Annual Workshop on A/B Testing and Platform-Enabled Learning Research

Steve Ritter
Carnegie Learning
Pittsburgh, United States
sritter@carnegielearning.com

Stephen E. Fancsali
Carnegie Learning
Pittsburgh, United States
sfancsali@carnegielearning.com

April Murphy
Carnegie Learning
Pittsburgh, United States
amurphy@carnegielearning.com

Neil Heffernan
Worcester Polytechnic Institute
Worcester, United States
neilheffernaniii@gmail.com

Ben Motz
Indiana University
Bloomington, United States
bmotz@indiana.edu

Debshila Basu Mallick
Rice University
Houston, United States
debshila@rice.edu

Jeremy Roschelle
Digital Promise
Redwood City, United States
jroschelle@digitalpromise.org

Danielle McNamara
Arizona State University
Tempe, United States
dsmcnamara1@gmail.com

Joseph Jay Williams
University of Toronto
Toronto, Canada
williams@cs.toronto.edu

ABSTRACT

Learning engineering adds tools and processes to learning platforms to support improvement research. One kind of tool is A/B testing—common in large software companies and also represented academically at conferences like the Annual Conference on Digital Experimentation (CODE), and the International Consortium for Innovation and Collaboration in Learning Engineering (IEEE ICICLE). Recently, several A/B testing systems have arisen that focus on conducting research in educational environments, including UpGrade, Terracotta, and E-TRIALS. A/B testing can help improve educational platforms, yet there are challenging issues unique to conducting such work in these contexts. In response, a number of digital learning platforms have opened their systems to learning-improvement research by instructors and/or third-party researchers, with specific supports necessary for education-specific research designs. This workshop will explore challenges of A/B testing in educational contexts, how learning platforms are accelerating education research, and how empirical approaches can be used to drive powerful gains in student learning. It will also discuss opportunities for funding to conduct platform-enabled learning research.

CCS CONCEPTS

• **General and reference** → **Empirical studies; Experimentation**; • **Applied computing** → **Education; Computer-assisted**

instruction; *Interactive learning environments; E-learning; Computer-managed instruction*; • **Human-centered computing** → **Empirical studies in HCI**; *HCI design and evaluation methods; Field studies*.

KEYWORDS

A/B Testing, Learning Engineering, Educational Technology, Digital Experimentation

ACM Reference Format:

Steve Ritter, Stephen E. Fancsali, April Murphy, Neil Heffernan, Ben Motz, Debshila Basu Mallick, Jeremy Roschelle, Danielle McNamara, and Joseph Jay Williams. 2024. Fifth Annual Workshop on A/B Testing and Platform-Enabled Learning Research. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale (L@S '24)*, July 18–20, 2024, Atlanta, GA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3657604.3664642>

1 THEME AND GOALS

The conference organizers all have deep practical experience with building learning engineering platforms for educational software. We will solicit presentations through the call for participation and, upon acceptance, organize those presentations into themes, which will form the basis of the workshop.

This will be a full-day workshop. The morning session will be devoted to the first keynote, as well as presentations and discussions of accepted papers. The first keynote will speak for 50 minutes with 10 minutes devoted to questions. We will organize presenters addressing major themes (e.g., “communicating to the public about random-assignment experiments”), with the expectation that we will have 4–6 themes addressed during the workshop. We will accept both long papers (8–10 pages) and short submissions for papers or demos (up to 4 pages). Accepted presenters for long papers will have 25 minutes to present, followed by 5 minutes for questions. Accepted short submissions will present their work at a poster and demo session following the long presentations.

The afternoon session will consist of the second keynote speaker will present, again with 50 minutes for the talk, and 10 minutes devoted to questions. After the second keynote will be additional

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

L@S '24, July 18–20, 2024, Atlanta, GA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0633-2/24/07

<https://doi.org/10.1145/3657604.3664642>

paper presentations followed by interactive activity such as "speed dating", matching researchers interested in large-scale educational research with digital learning platforms that are best suited to their needs.

2 BACKGROUND AND RELEVANCE TO L@S

There is no simple path that will take us immediately from the contemporary amateurism of the college to the professional design of learning environments and learning experiences. The most important step is to find a place on campus for a team of individuals who are professionals in the design of learning environments — learning engineers, if you will [5].

Learning engineering adds tools and processes to learning platforms to support improvement research [6]. One kind of tool is A/B testing [2], which is common in large software companies and also represented academically at conferences like the Annual Conference on Digital Experimentation (CODE). A number of A/B testing systems focused on educational applications have arisen recently, including UpGrade [4] and E-TRIALS [3]. A/B testing can be part of the puzzle of how to improve educational platforms, and yet challenging issues in education go beyond the generic paradigm. For example, the importance of teachers and instructors to learning means that students are not only connecting with software as individuals, but also as part of a shared classroom experience. Further, learning in topics like mathematics can be highly dependent on prior learning, and thus A or B may not be better overall, but only in interaction with prior knowledge [1]. In response, a set of learning platforms is opening their systems to improvement research by instructors and/or third-party researchers, with specific supports necessary for education-specific research designs. This workshop will explore how A/B testing in educational contexts is different, how learning platforms are opening up new possibilities, and how these empirical approaches can be used to drive powerful gains in student learning. It will also discuss forthcoming opportunities for funding to conduct platform-enabled learning research.

Issues to be addressed are expected to include:

- The role of A/B testing systems in complying with SEER principles (<https://ies.ed.gov/seer/>), which set a high bar for the goals of empirical studies of educational improvement
- Awareness of opportunities to use existing learning platforms to conduct research (<http://seernet.org>)
- Managing unit of assignment issues, as arise when students are in classrooms with a shared teacher
- Practical considerations related to experimenting in school settings, MOOCs, other contexts

- Ethical, data security, and privacy issues
- Relating experimental results to learning-science principles
- Understanding use cases (core, supplemental, in-school, out-of-school, etc.)
- Accounting for interactions between the intended contrast (A vs. B) and learners' prior knowledge, aptitudes, background, or other important variables
- A/B testing within adaptive software
- Adaptive experimentation
- Attrition and dropout
- Stopping criteria
- User experience issues
- Educator involvement and public perceptions of experimentation
- Balancing practical improvements

Four previous instances of this workshop (2020-2023) at the Learning at Scale conference were very successful, some with the highest registrations of any workshops at the conference. We welcome participation from researchers and practitioners who have either practical or theoretical experience related to running A/B tests and/or randomized trials as well as platform-enabled learning research. This may include researchers with backgrounds in learning science, computer science, economics and/or statistics.

3 EXPECTED OUTCOMES AND CONTRIBUTIONS

We will publish papers and continue to develop and deploy systems in this area. Following the earlier workshops, we created a Slack channel for continuing discussions on this issue (<https://edu-ab-testing.slack.com/>), and plan to continue discussions on this forum. We expect to conduct this workshop yearly, as it becomes part of the basis for a community of researchers conducting A/B tests at scale.

balance

REFERENCES

- [1] E.R. Fyfe. 2016. Providing feedback on computer-based homework in middle-school classrooms. *Computers in Human Behavior*, 63 (2016), 568–574. <https://doi.org/10.1016/j.chb.2016.05.082>
- [2] Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. 2013. Online controlled experiments at large scale. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1168–1176.
- [3] Korinn Ostrow, Neil T. Heffernan, and Joseph Jay Williams. 2017. Tomorrow's EdTech Today: Establishing a Learning Platform as a Collaborative Research Tool for Sound Science. *Teachers College Record*, 119, 3 (2017), 1–36.
- [4] Steve Ritter, April Murphy, Stephen Fancsali, Derek Lomas, Vivek Fitkariwala, and Nirmal Patel. 2020. UpGrade: An open source tool to support A/B testing in educational software. In *Proceedings of Learning@Scale 2020, L@S Workshop on A/B Testing at Scale*.
- [5] Herbert A. Simon. 1967. The job of a college president. *Educational Record* 48 (1967), 68–78.
- [6] Melina R. Uncapher. 2019. From the science of learning (and development) to learning engineering. *Applied Developmental Science* 23, 4 (2019), 349–352.