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

Ritter, Steven; Heffernan, Neil; Williams, Joseph Jay; Lomas, Derek; Motz, Ben; Basu Mallick, Debshila; Bicknell, Klinton; McNamara, Danielle; Kizilcec, Rene F.; More Authors

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

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

Steve Ritter
sritter@carnegielearning.com
Carnegie Learning

Neil Heffernan
neiltheffernan@gmail.com
Worcester Polytechnic Institute

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

Derek Lomas
derek@playpowerlabs.com
Delft University of Technology

Ben Motz
bmotz@indiana.edu
Indiana University

Debshila Basu Mallick
debshila@rice.edu
Rice University

Klinton Bicknell
klinton@duolingo.com
Duolingo

Danielle McNamara
dsmcnamara1@gmail.com
Arizona State University

Rene F. Kizilcec
kizilcec@cornell.edu
Cornell University

Jeremy Roschelle
jroschelle@digitalpromise.org
Digital Promise

Richard Baraniuk
richb@rice.edu
Rice University

Ryan Baker
ryanshaunbaker@gmail.com
University of Pennsylvania

ABSTRACT

The Third Annual Workshop on A/B Testing and Platform-Enabled Learning Research will be held in June 2022 at the 2022 ACM Conference on Learning at Scale in New York, NY, USA. We present the workshop's theme, goals, organizers, and structure. The workshop builds on two previous virtual workshops held in conjunction with the 2020 and 2021 ACM Conferences on Learning at Scale. The present workshop's title is a variation on the titles of previous workshops, which were the First and Second Annual Workshop on Educational A/B Testing at Scale, respectively.

CCS CONCEPTS

• **Applied computing** → **Computer-assisted instruction**; **E-learning**; **Computer-managed instruction**.

KEYWORDS

A/B testing, platform-enabled learning research, digital experimentation, experimental design, large-scale field trials, learning engineering

1 INTRODUCTION

> 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 [4].

Learning engineering adds tools and processes to learning platforms to support improvement research [6]. One kind of tool is A/B testing [3], 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 [5] and E-TRIALS [2]. 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 with open and generalizable science

The 2020 and 2021 workshops on this topic were very successful, with some of 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.

2 ORGANIZERS

- Steve Ritter, Carnegie Learning
- Neil Heffernan, WPI
- Joseph Jay Williams, University of Toronto
- Klinton Bicknell, Duolingo
- Derek Lomas, Delft University of Technology
- Jeremy Roschelle, Digital Promise
- Ben Motz, Indiana University
- Danielle McNamara, Arizona State University
- Richard Baraniuk, Rice University
- Debshila Basu Mallick, Rice University
- Rene Kizilcec, Cornell University
- Ryan Baker, University of Pennsylvania

3 PRE-WORKSHOP PLANS

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.

4 WORKSHOP STRUCTURE

This will be a full-day workshop. The morning session will be devoted to presentations and discussions of accepted papers. We will organize presenters into sessions 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.

Each presenter will have 15 minutes to present, followed by 5 minutes for questions. At the end of each theme session, a discussant will lead a panel discussion including the presenters and structured around discussion related to the general theme.

The afternoon session will consist of two sets of breakout sessions, organized by key questions. Attendees can choose which session they wish to attend. One session will focus more on A/B testing within one's own platform; the other on how a researcher might use other platforms that are open to research from third parties. They collaborate on a Google Doc, which serves to guide a presentation to the full workshop, along with notes from the session. In the 2022 workshop, the breakout groups addressed the following questions:

- How can we promote the idea to the public that (a) A/B testing and RCTs and (b) opening commonly used platforms to improvement research are paths to improving educational outcomes?
- Different communities use terms like "A/B Testing," randomized field trials, RCTs, experiments, etc. Are these the same things? What language choices will create desirable engagement and avoid resistance?
- How do we communicate about the idea of incremental improvement in education? Does the public understand this? How do we contextualize A/B testing as part of this process?
- How can we ensure that the results of A/B tests are applicable to vulnerable and historically underserved populations and address equity?
- How do we talk about the "winner" of an A/B test in the educational context? How do we communicate about certainty of outcomes? How should we think about generalization to other populations and contexts?
- How do we talk to the education policy community about the value of A/B tests? How can we communicate about a balance between "doing what works" and "finding out what works better"?
- We sometimes fear that there will be objections to random assignment, which leads to A/B testing being done without public awareness. How can we be more open about continuous improvement and encourage the idea that A/B testing is a best practice that should be seen as a benefit to schools, teachers and students?
- How do we think about informed consent in educational A/B testing? In what cases should we expect to get approval from teachers, parents and the students themselves?
- How can we better include educators, administrators, students and teachers in the design and coordination of A/B tests?

5 POST-WORKSHOP PLANS

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://join.slack.com/t/edu-ab-testing/shared_invite/zt-lwo0fas6-eV7GyaOb3ArdVEUp0b_

s3Q), and we expect to continue discussions on this forum. We expect this workshop to be repeated and become part of the basis for a community of researchers who are conducting A/B tests at scale.

6 CALL FOR PARTICIPATION

> 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 [4].

Learning engineering adds tools and processes to learning platforms to support improvement research [6]. One kind of tool is A/B testing [3], 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 [5] and E-TRIALS [2]. 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.

We invite papers (up to 4 pages in CHI Proceedings format) addressing issues with conducting A/B testing and learning engineering platforms, including those addressing:

- 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 with open and generalizable science

REFERENCES

[1] Emily 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] Neil T. Heffernan Korrin S. Ostrow and Joseph J. 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.

[3] Brian Frasca Toby Walker Ya Xu Ron Kohavi, Alex Deng 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* (Chicago, Illinois, USA) *(KDD '13)*. Association for Computing Machinery, New York, NY, USA, 1168–1176. https://doi.org/10.1145/2487575.2488217

[4] Herbert A. Simon. 1967. The job of a college president. *Educational Record* 48 (1967), 68–78.

[5] Stephen E. Fancsali J. Derek Lomas Vivek Fitkariwala Steve Ritter, April Murphy and Nirmal Patel. 2020. UpGrade: An open source tool to support A/B testing in educational software. In *Proceedings of the First Workshop on Educational A/B Testing at Scale*, April Murphy Stephen E. Fancsali and Steve Ritter (Eds.). EdTechBooks, –. https://edtechbooks.org/ab_testing_2020

[6] Melina R. Uncapher. 2019. From the science of learning (and development) to learning engineering. *Applied Developmental Science* 23, 4 (2019), 349–352. https://doi.org/10.1145/1219092.1219093