



Delft University of Technology

Second Workshop on Educational A/B Testing at Scale

Ritter, Steve; Heffernan, Neil; Williams, Joseph Jay; Lomas, Derek; Bicknell, Kinton

DOI

[10.1145/3430895.3460876](https://doi.org/10.1145/3430895.3460876)

Publication date

2021

Document Version

Final published version

Published in

L@S 2021 - Proceedings of the 8th ACM Conference on Learning @ Scale

Citation (APA)

Ritter, S., Heffernan, N., Williams, J. J., Lomas, D., & Bicknell, K. (2021). Second Workshop on Educational A/B Testing at Scale. In *L@S 2021 - Proceedings of the 8th ACM Conference on Learning @ Scale* (pp. 311-314). (L@S 2021 - Proceedings of the 8th ACM Conference on Learning @ Scale). Association for Computing Machinery (ACM). <https://doi.org/10.1145/3430895.3460876>

Important note

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

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Second Workshop on Educational A/B Testing at Scale

Steve Ritter
Carnegie Learning
sritter@carnegielearning.com

Neil Heffernan
Worcester Polytechnic Institute
neiltheffernaniii@gmail.com

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

Derek Lomas
Delft University of Technology
derek@playpowerlabs.com

Klinton Bicknell
Duolingo
klinton@duolingo.com

BACKGROUND

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. [1]

The emerging discipline of Learning Engineering is focused on putting into place tools and processes that use the science of learning as a basis for improving educational outcomes [2]. An important part of Learning Engineering focuses on improving the effectiveness of educational software. In many software domains, A/B testing has become a prominent technique to achieve the software's goals [3]. Many large companies (Amazon, Google, Facebook, etc.) run thousands of AB tests and present at the Annual Conference on Digital Experimentation (CODE), but that venue is too broad to address AB testing issues specific to EdTech platforms. We see a need to address issues with running large-scale A/B tests within the educational context, where the use of A/B testing lags other industries. This workshop will explore ways in which A/B testing in educational contexts differs from other domains and proposals to overcome current challenges so that this approach can become a more useful tool in the learning engineer's toolbox. Issues to be addressed are expected to include:

- managing unit of assignment issues
- measurement, including both short and long-term outcomes
- practical considerations related to experimenting in school settings, MOOCs, & other contexts

- ethical and privacy issues
- relating experimental results to learning-science principles
- understanding use cases (core, supplemental, in-school, out-of-school, etc.)
- accounting for aptitude-treatment interactions
- 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 generalizable science

The 2020 workshop on this topic was very successful, with over 100 registrations. We welcome participation from researchers and practitioners who have either practical or theoretical experience related to running A/B tests and/or randomized trials. This may include researchers with backgrounds in learning science, computer science, economics and/or statistics.

ORGANIZERS

- Steve Ritter, Carnegie Learning
- Neil Heffernan, WPI
- Joseph Jay Williams, University of Toronto
- Derek Lomas, Delft University of Technology
- Klinton Bicknell, Duolingo

PRE-WORKSHOP PLANS

The conference organizers all have deep practical experience with running A/B tests within 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 work is licensed under a Creative Commons Attribution International 4.0 License.

L@S'21, June 22–25, 2021, Virtual Event, Germany.

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8215-1/21/06.

<https://doi.org/10.1145/3430895.3460876>

WORKSHOP STRUCTURE

This will be a half-day workshop. The 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”). 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.

We will end the session with breakout groups, organized by key questions. Attendees can choose which session they wish to attend. They collaborate on a Google Doc, which serves to guide a presentation to the full workshop, along with notes from the session. In the 2020 workshop, the breakout groups addressed the following questions:

- How can we promote the idea to the public that A/B testing (or randomized field trials) is a path to improving educational outcomes?
- Different communities use terms like “A/B Testing,” randomized field trials, RCTs, experiments, etc. Are these the same things? Is there a better name to use in the educational context?
- 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?
- 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?

POST-WORKSHOP PLANS

We will publish papers and continue to develop and deploy systems in this area. Following the 2020 workshop, 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.

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. [1]

The emerging discipline of Learning Engineering is focused on putting into place tools and processes that use the science of learning as a basis for improving educational outcomes [2]. An important part of Learning Engineering focuses on improving the effectiveness of educational software. In many software domains, A/B testing has become a prominent technique to achieve the software’s goals [3], but educational software tends to lag other fields in the use of A/B testing, particularly at scale. This workshop will explore ways in which A/B testing in educational contexts differs from other domains and proposals to overcome these challenges so that A/B testing can become a more useful tool in the learning engineer’s toolbox.

We invite papers (up to 4 pages in CHI Proceedings format) addressing issues with conducting A/B tests and random assignment experiments at scale, including those addressing:

- managing unit of assignment issues
- measurement, including both short and long-term outcomes
- practical considerations related to experimenting in school settings, MOOCs, & other contexts
- ethical and privacy issues
- relating experimental results to learning-science principles
- understanding use cases (core, supplemental, in-school, out-of-school, etc.)
- accounting for aptitude-treatment interactions
- A/B testing within adaptive software
- adaptive experimentation
- attrition and dropout
- stopping criteria
- User experience issues

Workshop

L@S'21, June 22–25, 2021, Virtual Event, Germany

- Educator involvement and public perceptions of experimentation
- Balancing practical improvements with generalizable science

REFERENCES

- [1] Herbert A. Simon. 1967. The job of a college president. *Educational Record*, 48, 68-78.
- [2] Melina R. Uncapher (2018): From the science of learning (and development) to learning engineering, *Applied Developmental Science*, <https://doi.org/10.1080/10888691.2017.1421437>
- [3] Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., & Pohlmann, N. (2013, August). Online controlled experiments at large scale. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1168-1176).