



Delft University of Technology

## Workshop Proposal

### Educational A/B Testing at Scale

Ritter, Steven; Heffernan, Neil; Williams, Joseph Jay; Settles, Burr; Grimaldi, Phillip; Lomas, Derek

#### DOI

[10.1145/3386527.3405933](https://doi.org/10.1145/3386527.3405933)

#### Publication date

2020

#### Document Version

Final published version

#### Published in

L@S 2020 - Proceedings of the 7th ACM Conference on Learning @ Scale

#### Citation (APA)

Ritter, S., Heffernan, N., Williams, J. J., Settles, B., Grimaldi, P., & Lomas, D. (2020). Workshop Proposal: Educational A/B Testing at Scale. In *L@S 2020 - Proceedings of the 7th ACM Conference on Learning @ Scale* (pp. 219-220). (L@S 2020 - Proceedings of the 7th ACM Conference on Learning @ Scale). Association for Computing Machinery (ACM). <https://doi.org/10.1145/3386527.3405933>

#### 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.

# Workshop Proposal: Educational A/B Testing at Scale

**Steve Ritter**  
Carnegie Learning  
sritter@carnegielearning.com

**Neil Heffernan**  
Worcester Polytechnic Institute  
neilheffernaniii@gmail.com

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

**Burr Settles**  
Duolingo  
burr@burrsettles.com

**Phillip Grimaldi**  
Kahn Academy  
phillip.grimaldi@gmail.com

**Derek Lomas**  
Delft University of Technology  
derek@playpowerlabs.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. [2]

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 [3]. 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 [1]. 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

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the Owner/Author.

L@S '20, August 12–14, 2020, Virtual Event, USA  
© 2020 Copyright is held by the owner/author(s).  
ACM ISBN 978-1-4503-7951-9/20/08.  
<https://doi.org/10.1145/3386527.3405933>

- 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

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
- Burr Settles, Duolingo
- Phil Grimaldi, Kahn Academy
- Derek Lomas, Delft University of Technology

## 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.

**WORKSHOP STRUCTURE**

This will be a four-hour virtual workshop. 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.

**POST-WORKSHOP PLANS**

We will publish papers and continue to develop and deploy systems in this area. 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. [2]

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 [3]. A/B testing can be an important part of this approach [1], 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 its use in other domains and discuss 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 at scale, including:

- 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

**REFERENCES**

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