Workshop: Design and Application of Collaborative, Dynamic, Personalized Experimentation

Williams, J.J. [1], Heffernan, N.[2] , Poquet, O. [1]

[1] National University of Singapore, Singapore
[2] Worcester Polytechnic Institute, USA

Abstract. The proposed workshop will focus on the design and application of randomized experimental comparisons, that investigate how components of digital problems impact students' learning and motivation. The workshop will demonstrate how randomized experiments powered by artificial intelligence can enhance personalized components of widely-used online problems, such as prompts for students to reflect, hints, explanations, motivational messages, and feedback. The participants will be introduced to dynamic experiments that reweight randomization to be proportional to the evidence that conditions are beneficial for future students and will consider the pros and cons of using such more advanced statistical methods to ensure research studies lead to practical improvement. The focus will be on real-world online problems that afford the application of randomized experiments; examples include middle school math problems (www.assistments.org), quizzes in on-campus university courses, activities in Massive Open Online Courses (MOOCs). The attendees will have the opportunity to collaboratively develop hypotheses and design experiments that could then be deployed, such as investigating the effects of different self-explanation prompts on students with varying levels of knowledge, verbal fluency, and motivation. This workshop aims to identify concrete, actionable ways for researchers to collect data and design evidence-based educational resources in more ecologically valid contexts.

1 Introduction

The adoption of digital technologies in education offers novel opportunities for bridging research and practice, as it lowers the barriers to conduct randomized comparisons in real-world settings. Currently many studies take place in laboratories, as classroom field experiments in physical environments are challenging to randomize at the student level. To address this challenge, the workshop aims to connect theories and approaches from the learning sciences with the improvement of ecologically valid educational resources, by designing randomized comparisons that can be deployed with students.

The workshop will investigate how to enhance components of digital problems to increase students' learning and motivation. The goal is to focus on components of widely used online problems, like prompts for students to reflect [1, 2], hints, explanations, motivational messages, and feedback. The focus of the workshop will be real-world digital problems or activities for which it is actually possible to conduct the proposed experiments. These include problems on the www.assistments.org platform for middle school math, quizzes in on-campus university courses, and Massive Open Online Courses (MOOCs).

Components of online problems are especially germane because: (1) They are ubiquitous in a wide range of educational settings, topics, age groups. (2) There are immediate dependent measures of engagement (time spent on problems, repeated attempts) and learning (accuracy and time needed to solve future near and far transfer problems). (3) A wide range of variables can be experimentally investigated in enhancing online problems, through appropriate design of hints [3], explanations [4], and learning tips [5]. Despite the extant research that demonstrates that quality support in problems can benefit learning, there are many open questions about how to provide the best instructional support and feedback in interactive problems (see [6] for a review).

2 Dynamic Experimentation

A challenge that arises in conducting randomized comparisons is minimizing the chances that students are disadvantaged by receiving conditions that are bad for learning and maximizing the chances that data from experiments leads to practical improvements for future students. The workshop will introduce how to dynamically adapt experiments, by analyzing data in real-time and weighting randomization, so that the probability of assigning a student to a condition is proportional to the probability that the condition is best for them (leads to highest learning or engagement).

Algorithms that provide statistically principled trade-offs between experimenting and maximizing outcomes have been extensively studied in machine learning [7], website testing [8], and medical applications [9]. Workshop co-organizer Williams has implemented one of these algorithms, i.e. Thompson Sampling [10], into a system for experimentation on explanations in online problems [11], freely available at the URL www.josephjaywilliams.com/dynamicproblem. The Thompson Sampling algorithm used an adaptively weighted random policy that changed the probability of assigning different explanations by using data about students' ratings of how helpful explanations were for their learning. For example, the probability of receiving any one of four explanations was 25% for the first learner. But if the first twenty learners rated explanations three and four more highly than explanation one or two, the probability of the twenty-first learner receiving each explanation might be changed to 20%, 16%, 34%, 30%. This preserves random assignment and allows causal conclusions, while dynamically using the data collected to increase how many people receive higher rated explanations. A separate study with pre- and post- test measures of learning showed that using the system to identify highly rated explanations increased learning and transfer to accurately solving new math problems. Learning gains from these adaptive explanations did not differ significantly from high quality explanations written by a teacher. This is one example of the kinds of studies we hope workshop participants can design during the workshop, and then collaborate after the workshop.

3 Target Audience and Organizers

The workshop targets a broad group of researchers interested in randomized comparisons in digital educational environments, as well as university instructors and K12 teachers interested in collaborating on such studies.

Organizers of the workshop: **Joseph Jay Williams**, Assistant Professor, School of Computing, National University of Singapore (www.josephjaywilliams.com). **Neil Heffernan**, Professor, Learning Sciences Program & Computer Science, Worcester Polytechnic Institute (www.neilheffernan.net), Co-founder of www.assistments.org platform. **Oleksandra Poquet**, Postdoctoral Fellow, Institute for Applied Learning Sciences and Educational Technology, National University of Singapore.

References

1. J. J. Williams, T. Lombrozo, The role of explanation in discovery and generalization: evidence from category learning. *Cogn. Sci.* **34**, 776–806 (2010).
2. M. T. H. Chi, N. De Leeuw, M.-H. Chiu, C. Lavancher, Eliciting Self-Explanations Improves Understanding. *Cogn. Sci.* **18**, 439–477 (1994).
3. S. Kardan, C. Conati, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (ACM, 2015; https://dl.acm.org/citation.cfm?doid=2702123.2702424), pp. 3671–3680.
4. J. Wittwer, A. Renkl, Why Instructional Explanations Often Do Not Work: A Framework for Understanding the Effectiveness of Instructional Explanations. *Educ. Psychol.* **43**, 49–64 (2008).
5. K. Berthold, M. Nückles, A. Renkl, Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*. **17**, 564–577 (2007).
6. Valerie J. Shute, Focus on Formative Feedback. *Rev. Educ. Res.* **78**, 153–189 (2008).
7. J. Langford, T. Zhang, The Epoch-Greedy Algorithm for Contextual Multi-armed Bandits (available at http://hunch.net/~jl/projects/interactive/sidebandits/bandit.pdf).
8. L. Li, W. Chu, J. Langford, R. E. Schapire, in *Proceedings of the 19th International Conference on World Wide Web* (ACM, New York, NY, USA, 2010; http://doi.acm.org/10.1145/1772690.1772758), *WWW '10*, pp. 661–670.
9. S. A. Murphy, An experimental design for the development of adaptive treatment strategies. *Stat. Med.* **24**, 1455–1481 (2005).
10. O. Chapelle, L. Li, An Empirical Evaluation of Thompson Sampling in *Proceedings of the 24th International Conference on Neural Information Processing Systems* (Curran Associates Inc., USA, 2011; http://dl.acm.org/citation.cfm?id=2986459.2986710), *NIPS'11*, pp. 2249–2257.
11. J. J. Williams *et al.*, AXIS: Generating Explanations at Scale with Learnersourcing and Machine Learning in *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (ACM, New York, NY, USA, 2016; http://doi.acm.org/10.1145/2876034.2876042), *L@S '16*, pp. 379–388.
12. Williams, J., Rafferty, A., Tingley, D., Ang, A., Lasecki, W., & Kim, J. Enhancing Online Problems Through Instructor-Centered Tools for Randomized Experiments. in *Proceedings of CHI (2018), 36th Annual ACM Conference on Human Factors in Computing Systems. (2018)*
13. National Science Foundation: SI2-SSE. Adding Research Accounts to the ASSISTments Platform: Helping Researchers Do Randomized Controlled Studies with Thousands of Students. (1440753) $486,000. 2014 - 2016.