UpGrade: Scalable Digital Experimentation in Authentic Learning Settings

April Murphy, Stephen E. Fancsali, Steven Ritter

Carnegie Learning, Inc., Pittsburgh PA 15219, USA
{amurphy, sfancsali, sritter}@carnegielearning.com

Abstract. This interactive event demonstrates UpGrade, an open-source platform for scalable, digital experimentation (sometimes called A/B testing) in authentic learning environments like K-12 classrooms. Such learning contexts often present obstacles to the design and deployment of high-quality experiments that test the effectiveness, efficiency, or other characteristics of substantively different instructional approaches.

Keywords: Digital Experiments, A/B Testing, Educational Technology.

1 Challenges for Digital Experiments in Education

Learning technologies often use artificial intelligence (AI) for adaptive instruction. Technologies may, for example, implement mastery learning [1] to provide practice for students until the system estimates that the student has achieved mastery on a set of skills. Such student work is "self-paced," with students within, for example, the same middle school math classroom often working on disparate topics in a class session.

Consider the case in which an educational researcher seeks to rigorously test, via a large-scale experiment, two substantively different approaches to a math topic like fraction addition, calling them Experience A and Experience B. At any particular day or time in a particular class, some students may have already completed the relevant content covered by Experience A and Experience B, and others may have yet to cover important pre-requisites to be ready to learn the content. In addition, if Experience A and Experience B differ substantially, teachers may be burdened by having to support two substantively different approaches to learning this content. The UpGrade platform (https://www.upgradeplatform.org) is developed to help researchers deal with these and other complexities of real-world classrooms to deliver ecologically valid, large-scale experiments or A/B tests to rigorously evaluate different approaches to learning in AI-driven and related adaptive or self-paced learning technologies.

2 UpGrade

We demonstrate several features of the UpGrade platform [2], which substantially decreases the difficulty and lowers the costs of running large-scale, ecologically valid

experiments that test substantive instructional changes in settings like K-12 classrooms that use adaptive learning platforms like Carnegie Learning's MATHia [3]. Specifically, we consider UpGrade's management of group random assignment to experimental conditions and its consistency rules, which help to manage, as an experimental is designed and specified, decisions that need to be made to rigorously run experiments given the self-paced nature of AI-driven, adaptive learning technologies.

Experiments deployed using UpGrade may randomize condition assignments at either the individual or group level. Group-level random assignment in UpGrade [4] allows for experimental conditions to be assigned based on meaningful clusters or groupings of students in environments like K-12 classrooms, including grouping by class, teacher, or school, helping to ensure that, if desired, students within such clusters all receive consistent learning experiences.

UpGrade's consistency rules are intended to capture a researcher's intent in managing questions that often arise during experiments due to the self-paced nature of learning in AI-driven adaptive software. Consider the case in which a student in a class has already reached the content relevant to Experiences A and B, but the researcher has yet to actually deploy the experiment which contrasts Experience A with Experience B. If the researcher designs a group-randomized experiment at the class level with "group" consistency, then consistency of the learning experience for the student's group (i.e., her class) is paramount, so one student having reached this content excludes her class as a whole from the experiment contrasting Experience A with B when it is deployed. Individual consistency would merely exclude the student who reached this content earlier from the experiment contrasting Experience A with B, as this early-arriving student would not have been randomly assigned. The remainder of the early-arriving student's class would be assigned at the class-level to either work with Experience A or B. In an experiment using the individual consistency rule, a student who switches classes will continue to have the same experience to which they were originally assigned, regardless of the experimental condition assigned to the class to which they "switch."

This interactive event serves as a brief introduction to UpGrade and its power to facilitate rigorous, randomized digital experiments in authentic learning settings.

References

1. Bloom, B.S.: Learning for mastery. Evaluation Comment 1(2) (1968).
2. Ritter, S., Murphy, A., Fancsali, S.E., Fitkariwala, V., Patel, N., Lomas, J.D.: 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 (at Learning @ Scale 2020). (2020).
3. Ritter, S., Anderson, J.R., Koedinger, K.R., Corbett, A.T.: Cognitive Tutor: Applied research in mathematics education. Psychonomic Bulletin & Review 14, 249–255 (2007).
4. Ritter, S., Murphy, A., Fancsali, S.E.: Managing group random assignment in UpGrade. In: Proceedings of the First Workshop on Educational A/B Testing at Scale (at Learning @ Scale 2020). (2020).