Toward a Sound Environment for Robust Learning Analytics

Korinn S. Ostrow
Worcester Polytechnic Institute
Worcester, MA 01609
ksostrow@wpi.edu

ABSTRACT
My dissertation is rooted in the promotion of ‘Big Experimentation’ rather than simply ‘Big Data.’ Equipping pre-existing learning platforms with the requirements necessary for researchers to conduct sound science is critical in the age of technology driven learning. My work enlists a dynamic approach to establish ASSISTments, a popular online learning platform, as a reputable tool for learning analytics research. Much of my work involves testing hypotheses regarding student learning through randomized controlled experiments. At scale however, my research also considers how to enrich the research environment within ASSISTments. This issue is two fold: increasing the capabilities of ASSISTments as a tool for researchers on both small and large scales (i.e., enhancing design capabilities and designing a universal data reporting infrastructure), and working closely with researchers external to the platform to better understand their needs. The primary end goal of my work will be to define a theoretical framework and a set of policies that will allow similar platforms to establish themselves as sound environments for robust learning analytics.

Categories and Subject Descriptors
K: Applications to Education. K.3: Computers and Education.

General Terms
Experimentation, Measurement, Standardization.

Keywords
Big Experimentation, Sound Research Environments, Online Learning Platforms, RCTs, Data Mining, Open Data Reporting

1. BACKGROUND
The field of learning analytics is increasingly driven by big data in education. Many researchers working in this space approach analysis and student modeling strictly by data mining large datasets that have been made public through the KDD Cup or the PSLC Datashop [5]. However, it is also important that researchers in this realm develop and test specific hypotheses regarding issues like pedagogy, classroom behaviors, enhancements to curriculum and course content, and elements of human computer interaction between students and the technologies meant to drive and improve learning. In essence, ‘Big Data’ offers a top down approach to learning about learning, while ‘Big Experimentation’ could offer a bottom up approach.

Rigorously controlled experimentation in the form of randomized controlled trials (RCTs) is in demand in K-12 learning [17, 14]. Scalable technologies like online learning platforms will soon play a key role in allowing educational researchers to conduct RCTs, the “gold standard” for pinpointing causal relationships, efficiently and cost effectively in authentic learning environments. Online learning platforms used for rigorously designed hypothesis testing have the potential to benefit the users of the system (i.e., teachers and students), stakeholders in the learning process (i.e., administrators and parents), other researchers and the quality of research in the field, the platform itself through iterative improvements to design and content, and the general knowledge of the learning analytics community.

The value of online learning platforms as tools to conduct real time experiments that passively tap into rich student data from authentic learning environments is a relatively new notion. Few platforms offer researchers the ability to design and implement rigorously controlled experiments. Those that do permit research fail to market the capability to researchers, as the processes involved require insider knowledge, considerable programming skills, or a similarly steep learning curve (i.e., Khan Academy, edX) [19]. Yet, when designed with the flexibility to modify content and instill specific learning interventions, online learning platforms can provide researchers with access to extensive and diverse pre-existing subject pools that log hundreds of thousands of data points on a daily basis. This logged data includes fine-grained information about each student’s learning process that can easily be used to examine how proposed interventions differ from control treatments while drastically reducing the overhead of such research. By overlaying a learning platform with the capability to conduct sound science, the resulting ‘Big Data’ suddenly carries the added value of ‘Big Experimentation.’ Data that may otherwise be skimmed across in a predictive student model, or that may otherwise go untouched, now carries a wealth of experimental results that help educational researchers answer the question “What works best for whom?”

1.1 Research Question
The specific research question driving my work is “Can a theoretical framework be defined by which preexisting learning platforms evolve into shared scientific tools for robust learning analytics research?” This question allows for broad, yet marketable, research in the learning sciences, as the importance of big data analytics and rigorously controlled experimentation within education are expanding in parallel with the growing popularity of online learning platforms. Through research
intended to establish ASSISTments, a popular platform for mathematics education, as a viable environment for other researchers to conduct sound cognitive science, my end goal is to bring ‘Big Experimentation’ to light in fields that currently place more value on ‘Big Data.’

1.2 Current Solutions

Few online learning platforms have embraced controlled experimentation, despite the somewhat obvious direction in which technology is propelling learning analytics. Large corporations value RCE designs and A/B testing for the iterative improvement of user interfaces and to discover how to prolong user interactions. Through minimally invasive changes to the user experience, software analysts are able to glean insights regarding what works best to retain users [6]. Although randomized controlled environments are growing consistent in marketing, few large-scale education platforms promote this type of research, instead mining data of natural user interactions to determine the aspects of the platform that require further attention. Platforms like Coursera, EdX, Udacity, Google’s “Course Builder,” and Khan Academy, focus on delivering quality educational content to users at scale but often lack the simple, user-friendly backend that would allow researchers to interact with the system and implement experiments designed to test the system’s content or features.

Opening popular platforms to this type of experimentation would allow cognitive scientists, psychometricians, educational researchers, and K-12 educators to tap into pre-existing user populations to pose questions that could otherwise not be considered at such scale. Take for instance the IES Efficacy and Effectiveness trials that seek to isolate and scale best practices in education. The IES can only simultaneously fund a handful of large Effectiveness Trials that span five years and each cost at least $6 million, with few actually resulting in positive effects [3].

Online learning platforms could be used to run hundreds of smaller-scale studies within a year or less, at little to no cost above typical system maintenance. Research that was once restricted by the availability of funding can now be considered through the use of online learning platforms. Findings from such studies can then provide important insight regarding the risks of larger projects considered by the IES and the NSF.

The ASSISTments platform is unique in that it was designed with the content flexibility and collaborative environment necessary for educational research [2]. This allows my research to take the basic position of using the system to analyze hypotheses pertaining to student learning, while simultaneously working at scale to better the platform’s research environment for other users. ASSISTments, commonly used for both coursework and homework, presents students with immediate feedback and a variety of rich tutorial strategies. The platform is also a powerful assessment tool, providing teachers with a variety of student and class reports that pinpoint where students are struggling and enhance classroom techniques using real time data. A recent NSF grant has allowed ASSISTments to better serve researchers; any external researcher is able to use the platform to conduct educational research at virtually no cost and in a fraction of the time previously required to run experiments within K-12 environments. From the researcher’s point of view, ASSISTments offers a simple interface in which pre-existing certified content and feedback can easily be manipulated and fashioned into an RCT. Studies embedded within ASSISTments problem sets are then assigned to students by their teachers following regular usage protocols. ASSISTments conducts student-level randomization, allowing for students in the same class to receive different treatments within the same assignment, thereby strengthening the power of analysis. At scale, my work simplifies and strengthens the research environment within ASSISTments, refining best practices for researcher outreach and interaction.

The learning analytics community already promotes shared datasets for data mining and analysis. For instance, the NSF funded Pittsburgh Science of Learning Center’s Datashop [5], has been a key player in promoting open science through an extensive database of anonymized student data logged from various types of learning platforms. Data sharing increases the internal validity of research by holding researchers accountable for their findings. A great deal of attention has recently been brought to the fields of psychology and medicine due to the overwhelming inability to replicate significant findings [4]. However, the downfall of open data shops is that analysis can only be considered post hoc. Researchers are not able to implement their own hypotheses and observe the resulting effects through these previously logged files. External researchers working within the ASSISTments platform are able to do just that. After implementing their study, researchers are provided weekly communications with stable links to three archived file types that are intended to suit a number of use cases (i.e., data logged at the action level, data logged at the problem level, and data logged at the student level; see Figure 2, and reference [13] for additional information). These files are also supported by historical student information that incorporates student, class, and school level characteristics collected prior to the study in question (i.e., a student’s average accuracy on every problem she has ever experienced within ASSISTments). The process of data reporting and communication is fully automated, with reports sent weekly if at least three new students have participated in a study. Researchers are also able to request immediate data reports at any time. The interactive nature of this approach extends beyond the norms set by educational datashops, serving to increase the pace and power of learning analytics.

Perhaps the remaining obstacle in the conversion of many popular online learning platforms into environments for robust learning analytics research is the requirement for IRB approval. ASSISTments was able to work through this obstacle by establishing IRB approval for educational research at the system level [9]. By building an environment in which researchers work with the system rather directly with students, and requiring that manipulations to content never intentionally disadvantage student learning, students are able to remain anonymous to the researcher and studies can be conducted passively within pre-existing coursework and homework assignments. Working under the umbrella of the system’s IRB approval, external researchers typically must only seek expedited approval for work with ‘non-human subjects’ because they are ultimately provided anonymized log files. This approach is critical to the successful implementation of research within online learning platforms because it establishes a user-friendly environment for sound research that leaves few hurdles for the researcher.

2. MY RESEARCH APPROACH

Working to refine a preexisting learning platform into a shared scientific tool for robust learning analytics research is a novel solution for the learning sciences. While the exponential expansion of educational data parallels the growth and validity of online learning platforms as valid tools for education, few (if any) research teams have provided a voice for the development of these technologies as sound environments for research. My approach to define a framework by which similar platforms can make this
transition is dynamic in nature. A large portion of my research is driven by testing simple hypotheses regarding learning on ASSISTments’ pre-existing user population of over 50,000 students around the world [2]. At scale, my research also considers how to make learning analytics research within the platform easier for other researchers, especially those that are unfamiliar with ASSISTments. This issue is two fold: increasing the capabilities of ASSISTments as a tool for researchers on both small and large scales (i.e., enhancing design capabilities and refining a universal data reporting infrastructure that promotes sound science), and working closely with researchers external to the platform to better understand their research needs.

The future of learning analytics should be driven by ‘Big Experimentation’ within existing technologies, taking popular learning platforms and turning them into shared scientific tools that advance the collective understanding of best practices in education. Thus, an important end goal of my research will be to define how the methods used in my work with ASSISTments can be applied to other popular learning systems through a theoretical framework to revolutionize learning about learning.

2.1 Methodology
With the overall goal of establishing a sound environment for robust learning analytics research, my work thus far has included three primary directions:

1) Conducting RCTs within ASSISTments and data mining logged files to answer research questions regarding best practices within technology based learning.

2) Using the findings of these studies to work closely with ASSISTments programmers to improve the platform’s research infrastructure and to establish best practices for the universal reporting and analysis of logged data.

3) Working with researchers external to the ASSISTments team to iteratively improve the platform’s usability as a research tool.

Together, these directions allow me to answer my own personal research questions while continuously working to improve the platform’s capabilities for sound research at scale.

2.2 Current Status of Work
Thus far, my work has lead to ten peer reviewed publications, two journal articles currently in press, and five projects currently under review. The following sections highlight some of my pinnacle work thus far, emphasizing the dynamic approach taken to enhance ASSISTments as a research platform.

2.2.1 RCTs & Data Mining
Much of my research entails conducting RCTs using student-level randomization within ASSISTments to isolate best practices for learning outcomes while enriching the user experience. This work allows me to provide the ASSISTments programming team with recommendations for new features or improvements to the platform’s capabilities that are fueled by evidence of student learning or user retention.

A substantial issue observed within ASSISTments has been student disengagement with and/or maladaptive usage of system feedback. Some of my original work involved the design and implementation of an RCT comparing traditional text feedback with enriched feedback presented using short (15-30 second) video snippets. Designed in consideration of Mayer’s multimedia principles for the optimal design of e-Learning environments [1], this study observed significantly enhanced next problem performance following video tutoring, and generally positive student perceptions of video [10]. Motivation and learning outcomes can also be improved by altering content delivery patterns, as shown through my work detailing the benefits of interleaving (or mixing) skill content within authentic homework settings [12]. Students can also benefit from a sense of autonomy while learning. Choice is an intrinsically motivating factor [15] that has the potential to boost subjective control, or a student’s perception of their causal influence over their learning outcomes [16]. My work regarding the effects of student choice within ASSISTments has suggested higher performance for students asked to choose the style of feedback they would receive during an assignment (video vs. text), with an effect observed even in those students that did not ultimately utilize feedback [11].

While much of my work promotes the use of RCTs to detect when interventions promote significant learning gains, a portion of my work has fallen back on simple data mining techniques to establish the groundwork for improvements to ASSISTments that can then be examined further through controlled experimentation. Following RCT results that emphasized the importance of considering hints and attempts in student performance [12], it was theorized that partial credit assessment could be used to better define student models. By considering non-experimental log files from the 2012-2013 school year, a simple tabling technique using maximum likelihood probability was designed to observe the reduction in modeling error when comparing algorithmically defined partial credit to traditional binary correctness [7]. This work confirmed the benefits of considering rich aspects of student performance (i.e., hint usage, attempt count, and first action) when modeling the likelihood of next problem correctness. Further data mining was then conducted to consider refinements to the partial credit algorithm used in [8] through a grid search of per hint and per attempt penalization models. Findings suggested an optimal range for score penalizations [8] and a revised partial credit algorithm has since been implemented within ASSISTments.

2.2.2 Enhancing Research Capabilities
My research goals for running RCTs within ASSISTments and data mining log files are twofold: isolating and presenting significant findings, while also harnessing outcomes to improve ASSISTments as an environment for sound educational research.

Findings from my work comparing video feedback with traditional text feedback [10] informed an influx of video content into the system, as well as the design of efficient video recording and uploading tools for the ASSISTments iPad application. Additionally, 72 studies currently running within ASSISTments (57%) consider video interventions or use video to examine other learning effects. There is currently a push for the ASSISTments team to consider a new problem type that will allow pre-service teachers to record automatic responses when students are confused, thereby crowdsourcing explanations while assisting the training of this subpopulation.

My pilot study on student choice [11] (a feature not previously considered possible within ASSISTments due to the platform’s design restrictions) led to a collaboration with the team’s programmers to design an If-Then routing structure to strengthen research designs. Among many other possible uses, the implementation of this structure allows researchers to provide students with a fair learning experience if they have technical difficulties accessing hypermedia elements like video tutoring. As shown in Figure 1, the original design of the RCT on student choice was not able to
considerately route students that were unable to access video content away from the study. In the second version of the design, using the newly implemented If-Then structure, each student must first pass a “Video Check” that enhances the internal validity of the research design. If-Then path routing has proven useful in a variety of research designs created by external researchers, serving to broaden the platform’s ability to ask and answer complex research questions.

Through my RCT considering the effects of interleaved skill practice [12] and my data mining findings [7, 8], the ASSISTments team discovered the importance of partial credit assessment. By using information that extends beyond binary correctness, researchers are better able to glean the effects of interventions. Partial credit scoring has since been implemented within ASSISTments, providing researchers with a richer metric of student performance useful for reliable group differentiation [18]. The presence of partial credit scoring should also increase the platform’s usability for students and teachers by increasing interaction with tutoring strategies and improving assessment reports, respectively.

2.2.3 Working with External Researchers
There are currently 126 research studies being conducted within ASSISTments to solve practical problems within education and to isolate and promote best practices within technology driven learning. Many of these studies were designed by researchers external to the ASSISTments team. Through my research, I have collaborated with eight external research teams to work toward improving the platform’s research environment.

From these collaborations, it became clear that for an online learning platform to successfully evolve into a tool for sound science, intense focus must be paid to the system’s usability, the
accessibility of data, and the simplification of interactivity. Working with the ASSISTments programming team, my findings have helped to build the Assessment of Learning Infrastructure (ALI) [13]. Until the spring of 2015, researchers were able to easily author and manipulate RCTs within ASSISTments, but often were often stuck when trying to retrieve their data, as the process required database access and knowledge of SQL (or access to an ASSISTments team member). In order to make a user-friendly environment for all researchers, ALI was built to automate the process of data retrieval. Within the past six months, the infrastructure has evolved from an automated process meant to present researchers with a single clean, organized, anonymized data file, to an approach that incorporates a variety of universal files types (see Figure 2) in an attempt to accommodate all researchers. Guided by collaborations with external researchers, ALI is also learning to provide automated data analysis, offering researchers a selection of preliminary findings and warning of threats to internal validity. In this sense, our team hopes that ALI will flag ‘Big Experimentation’ and promote learning analytics research within online learning platforms like ASSISTments.

Further work with ALI, and similar approaches to make ASSISTments a sound environment for robust learning analytics, will form much of my dissertation. As I continue to incorporate my own research questions with the feedback obtained through collaboration with external researchers, my work will lead to synergistic improvements for the platform and its users. The immediate impact of my research has been evident through continued improvements to the ASSISTments platform. My work thus far has inspired content expansion [10], infrastructure modifications [11], and the development of a revolutionary tool for researchers working at scale [13]. As such, I hope to craft a dissertation that defines a theoretical framework and a set of policies to aid the transition of similar platforms into tools for the betterment of learning analytics.

3. ACKNOWLEDGMENTS

Funding for my work has been granted by a U.S. Department of Education GAANN fellowship (P200A120238), an Office of Naval Research Grant (N00014-13-C-0127), and a PIMSE fellowship (NSF DGE-0742503). Thanks to S.O. & L.P.B.O.

4. REFERENCES


