Abstract Title Page
Not included in page count.

Title: How can digital online educational resources be used to bridge experimental research and practical applications? Embedding In Vivo Experiments in "MOOClets"

Authors and Affiliations:

Joseph Jay Williams
Harvard University
125 Mt. Auburn St Rm. 422
Cambridge, MA 02138 USA
joseph_jay_williams@harvard.edu

Samuel Maldonado
San Jose State University
5025 23rd st
Sacramento, CA, 95822 USA
sam.maldonado@cognitivescience.co

Betsy Anne Williams
Stanford University
520 Galvez Mall, Fifth Floor
Stanford, CA 94305 USA
betsyw@stanford.edu

Sara Rutherford-Quach
Stanford University
520 Galvez Mall, Fifth Floor
Stanford, CA 94305 USA
ruthersa@stanford.edu

Neil Heffernan
Worcester Polytechnic Institute
100 Institute Road, Computer Science Department
Worcester, MA, 01609 USA
nth@wpi.edu

Abstract Body

Background / Context:

Given the worldwide spread of digital technology and the Internet throughout modern life, students' everyday education increasingly includes components or activities that use digital software or online resources. This will be even more true for students in just two or five years. What opportunities and obligations does this carry for doing education research that rigorously identifies learning benefits, and linking these results to everyday practice?

One emphasis in the discussion of how online learning could support large scale research is that Massive Open Online Courses (MOOCs) allow for collecting fine-grained data, so that "big data" might allow for drastic improvements in learning, such as by identifying problematic questions, promoting the best videos, or adaptive/personalized instruction (Brusilovsky et al., 2007; Koedinger & Corbett, 2006; Sampson et al., 2010). While more extensive data collection is certainly a novel and valuable feature of online environments, we place our attention and effort on a less frequently cited but tremendously important feature of incorporating digital online resources alongside in-person instruction.

This is the capacity for conducting randomized experiments, like changing features of content and dynamics of exercises, or adding additional study resources or interactive tools. For example, students working with online videos and exercises are in an environment that merges real-world learning with a laboratory setting, because these *in vivo* randomized experiments can investigate how modifications to digital resources impact quantifiable measures of learning (Koedinger, 2011; Williams, 2013a). Far more easily than a classroom with a physical teacher or pen-and-paper exercise, an experiment embedded in a blended/online resource can assign hundreds of students to two (or ten) different versions of an online resource.

The ease with which this can be done in digital online environments *could* provide tremendous advantages*, if* existing education research and practice can appropriately leverage these benefits of digital Internet accessible resources in authentic learning contexts (Williams, Renkl, Koedinger, & Stamper, 2013). How can this actually be done? What are examples of experiments that have been implemented using online resources, and what are potential experimental paradigms to explore?

Objective & Focus of this Abstract:

To help readers in exploring the space of answers to these questions, this abstract describes the design and implementation of illustrative examples of three kinds of experimental approaches we have taken with collaborators. (Results & findings from each experiment are not reported here, since the design and detals of these *kinds* of experiments is the focus).

Adding Broad Impact Resources to a Course. The first experiment randomized which version of an introductory motivational video/webpage was included in a freely available online course on biology provided by Stanford to the general Internet population – as is typical for MOOCs. We label this kind of experiment "Adding Broad-Impact Resources to a Course" because the experimental manipulation is to add different versions of a resource like a video/webpage in the context of an entire course, and we can analyze the impact on a range of dependent measures of engagement and learning that are digitally logged throughout the course.

Expanded details of this experiment (and the two others) are in Table 1 below "Details of the Approaches for In Vivo Experiments in MOOClets", which presents: Setting, Participants, Kind of Experiment, Description, Example, Experimental Variable, Resources to be Compared, Educational Outcome Variable, Intervention, Data Collection & Analysis. The remaining bulk of this Abstract is presented in this table, facilitating direct comparison of the approaches.

Exercise Modifications. The second kind of experiment, "Exercise Modifications", is illustrated by an example experiment which added questions for reflection or prompts to explain to high school algebra exercises & worked examples on the website www.KhanAcademy.org, measuring the effect on accurately solving problems.

Add Interactive Online Tools plus Emails Reminders & Guidance on Use. The third kind of experiment is more novel and unusual. In the setting of a freely available online course helping teachers analyze and support student conversations as required by the Common Core standards, a "Conversation Analysis Tool" worksheet for teachers to analyze the learning effectiveness of student-to-student conversations (developed by Sara Rutherford-Quach, Jeff Zwiers, and Kenji Hakuta) was converted from a paper/PDF worksheet to an interactive tool in Qualtrics and presented as a resource that could be accessed at any time to remind or guide teachers through the process and a memorable hyperlink provided to it (tiny.cc/ctool, tiny.cc/ctooldirect is a demo without course registration).

The effect of providing the interactive tool is analyzed by comparing it to a condition where there is no interactive tool at all (default course) and to a condition where the tool is simply the static PDF. To examine the practical value of emphasizing this as a practical tool usable in the everyday activities of teaching (e.g. opening up the tiny.cc/ctool browser page on a computer or smartphone to remind oneself before or after class about the dimensions to analyze in student conversations) the dependent measures are usage of course materials, completion rate and quality of assignments related to the conversation analysis skill supported by the tool, the frequency and times of accessing the tool, and self-report by teachers of its utility.

In addition to investigating the benefits of the presence/absence of the tool, the experiment can also vary the extent to which the tool is used, by randomly assigning teachers to receive with differential frequency emails that remind them (and/or help them reflect on when) to use the tool in their actual practice. This further opens up a current area of experimental investigation – how to motivate, remind, and guide teachers and students use of various interactive online tools (regardless of the virtues of the particular kind of version discussed here).

It also shows how our work embedding such custom tools can be a valuable source of data as to when learners are implementing and applying skills, because use of the tool provides a digital trace of data for such activities, and because the embedding of custom online tool designed for research collects far better measures of behavior than are typically available in Learning Management Systems or MOOC platforms and makes them available instantaneously.

MOOClets. The term "MOOClet" (www.josephjaywilliams.com/mooclets) is used to describe modular educational components of online courses – lessons/videos, exercises, interactive tools, emails – and target modularity *within* courses, emphasizing how the conversation and technology around MOOCs could be more productively directed towards real-world application by focusing research and development on experiments comparing the effect of these MOOClets & modular components on quantitative measures of learning and engagement.

Conclusions

Three examples of experiments are presented that successfully navigated constraints in technology, practical implementation, and partnership with existing online educational resources to identify and implement opportunities to embed *in vivo* experiments testing the educational efficacy of particular lessons, exercises, interactive tools, and emails. While the extent to which the designs seem obvious and natural is a positive indicator of the quality of their design, it should be noted that these are selected from a wide range of other possible approaches that might have seemed reasonable but would not have been possible or successful in implementation. The contribution of this abstract will be realized if it guides future implementation of such in vivo experiments and identification of the appropriate granularity of other "MOOClets" for experimental research better than if readers had not been exposed to this work.

To further guide readers, Table 2 in the Appendix contains a preliminary list from our experience of "Examples of Instructional Components for Experimental Manipulation", and Table 3 a list of "Examples of Common Quantitative Educational or Learning Outcome Variables in Online Educational Resources" which will be greatly refined by future researchers.

Table 1: Details of the Approaches for In Vivo Experiments in MOOClets

Setting	**Setting**	**Setting**
Free online course on physiology in everyday life offered by Stanford and hosted as a MOOC at class.Stanford.edu using the MOOC platform OpenEdX	Algebra exercises on www.khanacademy.org that are part of both an exercise bank and sequence of videos	Free online course (with paid certification option) for K-12 teachers offered by Stanford Education faculty, using team-based platform www.NovoEd.com
Participants	**Participants**	**Participants**
5000 people with varied demographics: over 20 countries, 16-75 years	10 000 people (no demographics collected), likely range from students working independently to students doing in-class work or HW (~8^{th} grade level) to teachers, tutors, interested browsers	2000 U.S. K-12 teachers during the Spring semester, preparing to support new Common Core standards. Elementary & Secondary levels, multiple subjects, levels of experience. One group of Seattle teachers being paid as professional development, and meeting in person.
Kind of Experiment	**Kind of Experiment**	**Kind of Experiment**
Add Broad-Impact Resources to Course	Exercise Modifications	Add Interactive Online Tools plus Emails Reminders & Guidance on Use.
Description	**Description**	**Description**
Manipulating webpage/video resources that influence learning over multiple course components.	Changing assignments or practice exercises to improve learning of a specific topic.	Examining effect of adding interactive tool (& reminders) on the application of what is being taught, and examining how to encourage people to use the interactive tools.
Example	**Example**	**Example**
Presenting videos from instructor that were designed to increase motivation and resilience in the course.	Adding or varying prompts to reflect that appeared above mathematics exercises on Khan Academy.	Investigate educational benefits of motivational messages and reflection questions by collecting data from an interactive tool that supported real-world application of core ideas.
Experimental Variable	**Experimental Variable**	**Experimental Variable**
Presence and Type of encouraging video or text message.	Presence and Nature of prompt to explain that is inserted above math exercise.	Presence of interactive tool, presence of email reminders & encouragement to use the tool.
Resources to be compared	**Resources to be compared**	**Resources to be compared**

[A] No message (default course) [B] Encouraging Video [C] Growth Mindset video [D] Growth Mindset video by another instructor [E] Growth Mindset written text	Algebra Math exercise & worked examples on Khan Academy with [A] No explanation prompt [B] Prompt to paraphrase solution steps [C] Prompt to self-explain solution steps	[A] No interactive tool (default course). [B] Add course page with Interactive Tool for analyzing students' peer-to-peer conversations, has memorable hyperlink (tiny.cc/ctool) [C] Email with reminders to use tool
Educational Outcome Variable	**Educational Outcome Variable**	**Educational Outcome Variable**
Dropout Rate from MOOC Number of videos watched & Number of Exercises completed	Accuracy on First Attempt of Problems Time and number of hints needed to solve problems Accuracy, Time, Hints on subsequent problems without prompts	Completion Rate and Quality of Relevant Course Assignments Self-Reported Use & Judged Usefulness of Tool Frequency & Times at which Interactive Tool was accessed Responses made, in Tool, amount of time spent on each Tool component
Intervention	**Intervention**	**Intervention**
Until April-September www.EdX.org did not support any randomization of educational content, but does allow custom HTML, Javascript, iFrames, so Qualtrics survey containing video/text message was embedded into OpenEdX instead of a video/exercise. EdX & Qualtrics interface to pass participant's anonymous ID, then uses last digits of (randomly generated) ID to assign participants to condition. Because this anonymous ID is always linked to the person, this randomization method allows person to receive further treatments in that condition at other points in the course.	At the time this experiment was conducted, Khan Academy supported the GAE/Bingo A/B testing framework (a framework implemented in Google Apps Scripts). Participants were entered into a persistent condition upon attempting one of the target intervention exercises. Different versions of exercises were created using Khan Academy's open-source exercise building framework (which involved the modification of simple HTML files), and a custom solution for tracking participant input to prompts was created using the framework.	Provided access to an external tool (built in Qualtrics) via an LTI (Learning Tool Interoperability) request. NovoEd is one of the first MOOC platforms to provide widespread, simple LTI support. LTI components in NovoEd are the only way to provide an anonymized identifier to an external tool. A bridge tool was created to generate an LTI request and pass the appropriate information to Qualtrics, as Qualtrics does not currently support LTI (although they may in the future). Qualtrics is excellent for sophisticated rapid authoring.
Data Collection & Analysis	**Data Collection & Analysis**	**Data Collection & Analysis**
Because anonymous ID carries condition information, any information available to the instructor can be analyzed – overall grades, completion of quizzes, correctness of final answers on quizzes (repeat attempts allowed). Information about videos being watched, accuracy on first attempt doing problems, is only available from tracking logs. Very difficult to access – now possible at Stanford & Harvard.	Basic descriptives like mean accuracy are immediately available from Khan Academy's internal dashboard for A/B testing. After waiting upon internal mechanism at Khan Academy to get full data set, this provides a detailed by-problem log of events – accuracy on first attempt, all answers entered, number of hints requested, total time spent on problem, time logged in to solve problem, order in which problems completed. Data was provided for the 2 target problems over 2 week period, as well as all problems 2 months before and after.	Rich interaction log, including time at which resource is accessed. Possible but more taxing to get information from MOOC platform (www.NovoEd.com) or Learning Management System (e.g. Blackboard/Canvas) about interaction with videos and exercises.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Brusilovsky, P., Kobsa, A., & Nejdl, W. (Eds.). (2007). *The adaptive web: methods and strategies of web personalization* (Vol. 4321). Springer.

Koedinger, K.R. (2011). *Bridging the lab and the field with* in vivo *experimentation*. Paper presented at the Society for Research on Educational Effectiveness (SREE).

Koedinger, K. R. & Corbett, A. T. (2006). Cognitive Tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.) *The Cambridge Handbook of the Learning Sciences* (pp. 61-78). CambridgeUniversity Press.

Sampson, D., Karagiannidis, C., & Kinshuk, D. (2010). Personalised learning: educational, technological and standardisation perspective. *Digital Education Review*, (4), 24-39.

Williams, J. J. (2013a). Finding connections between basic experimental research and realistic online education contexts. In J. J. Williams (chair), *Online Learning and Psychological Science: Opportunities to integrate research and practice*. Symposium conducted at the annual convention of the Association for Psychological Science.

Williams, J.J. (2013b). *Improving Learning in MOOCs by Applying Cognitive Science*. Paper presented at the MOOCshop Workshop, International Conference on Artificial Intelligence in Education, Memphis, TN.

Williams, J. J., Renkl, A., Koedinger, K., Stamper, J. (2013). Online Education: A Unique Opportunity for Cognitive Scientists to Integrate Research and Practice. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Appendix B. Tables and Figures

Not included in page count.

Table 2: Examples of Instructional Components for Experimental Manipulation

Technical Specification	Pedagogical Description
Addition or Modification of video/webpage resources	Motivational messages & testimonials; Information about study skills; Bonus lessons on specific topics; Within-video editing that changes explanations or examples; Addition/Changes to within-video quizzes; Adding surveys;
Modification of Prompts & Messages in Exercises/Assignments	Motivational messages & feedback; Prompts to use learning strategies; Hints; Feedback; Additional explanations; Worked-out examples; Solutions; Changes from multiple-choice to open-response formats; Inclusion of hyperlinks to relevant topics or additional resources;
Emails to participants	Motivational messages & links to resources on learning strategies; Reminders to complete assignments; Explanations of why to use forums or other course resources;
Addition/Modification of assignments & exercises	Additional practice; Review of earlier concepts; Including challenging problems that make students aware of gaps in knowledge.
Changes to order and sequencing of videos and exercises	Reorganizing order in which topics are introduced; Inserting exercises before videos; Mixing together exercises practicing different concepts/skills;

Table 3: Examples of Common Quantitative Educational or Learning Outcome Variables in Online Educational Resources

Technical Description of Data Source	Educational Outcomes Computed
Logs of visits to course.	Dropout rates at different time periods; Frequency & Degree of engagement.
Quizzes embedded in videos; Practice problems; Weekly Problem Sets/Assignments; Projects; Midterm & Final Exams	Measures of accuracy in solving problems or answering questions. Time needed to respond accurately. Support needed – hints requested or resources viewed in order to respond accurately. Different **kinds** of measures vary in: a) Which specific topics or skill proficiencies they measure or target. b) Relationship to target knowledge: Measure reflects learning during *Study*, after *Delay*, that requires *Transfer*.
	Overall grades in course.