ManyClasses: A model for abstracting generalizable research principles from different learning contexts
Benjamin A. Motz1*, Joshua R. de Leeuw2, Paulo F. Carvalho3, Emily R. Fyfe1, Robert L. Goldstone1

1 Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana
2 Cognitive Science Department, Vassar College, Poughkeepsie, New York
3 Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania
* Email: bmotz@indiana.edu

Abstract

Research on learning has unique strengths when conducted in authentic educational contexts, as compared with laboratory contexts. Principal among these, an experiment embedded in a real class will provide a more legitimate measure of a learning activity’s benefit for real educational outcomes than an experiment with a synthetic learning task and materials (Motz et al., in press). But this strength also presents a problem: the properties of the class context constrain the generalizability of the research conclusions. For example, there might be differences in the effectiveness of a learning activity for different disciplines (e.g., Psychology vs. Chemistry), class formats (e.g., MOOC vs. face-to-face), and student populations (e.g., urban community colleges vs. elite universities), among others (see Koedinger, Booth, & Klahr, 2013; Gašević et al., 2016). How can one abstract a generalizable research inference from the situated context of a classroom experiment?

We propose a research model for disentangling abstract learning principles from authentic learning contexts, and we call this model ManyClasses. As with similar efforts in psychology (ManyLabs, Klein et al., 2014; ManyBabies, Frank et al., 2017), the core feature of ManyClasses is that researchers would measure an experimental effect across many independent samples – in this case, across many classes. Rather than conducting an embedded learning study in just one educational context (e.g., Introductory Physics course at a small private college), a ManyClasses study could examine the same research question in dozens of contexts, spanning a range of courses, institutions, formats, and student populations. By drawing the same experimental contrast across a diversity of educational implementations, and then analyzing pooled results, researchers can assess the degree to which an experimental effect might yield benefits, untethered from specific contexts. In addition to contributing to an estimation of the generalizable effect size of manipulations beyond particular classroom implementations, ManyClasses could also systematically investigate how a manipulation might be more- or less-effective for different students, in different situations.

The challenge for this ManyClasses model is the development of an experimental design that is sufficiently well-structured to rigorously support specific research contrasts, while still providing the flexibility to customize the materials to each class's content and unique learning goals. Indeed, a researcher cannot simply present Introductory Physics material to a Psychology class and expect a legitimate measure of the learning activity’s benefit in these independent samples. In order to appropriately test the effects of a given learning strategy across a diverse range of realistic implementations, teachers should be able to prepare original materials that are authentic to their disciplines’ and their institutions’ norms. To enable this level of customization, while maintaining the rigor of a controlled experiment, we imagine that ManyClasses researchers would prepare a structured
template of an online homework assignment, but with empty fields for content: For example, (1) an empty introductory content page; (2) another content page with a short answer question that could take one of two forms; (3) another content page with additional instruction in one of two different forms; (4) a question set that could give feedback in one of two different forms, and so on. The researchers outline the desired differences between the forms according to the research question (e.g., different forms could contrast whether instructional examples are abstract or concrete, whether instructions are controlling or supportive of students’ autonomy, etc.), providing examples to illuminate the intended contrasts. Once the template is complete, the ManyClasses researchers should publicly preregister the template, as well as the teacher recruitment, sampling, design, and analysis plans, in order to promote transparency in the conduct of this collective effort.

Given the template for this homework assignment, participating teachers would customize it for their classes, populating empty fields with their own materials (e.g., in a WYSIWYG editor) as befitting each class’s learning goals and content. The teacher presents this customized variant to their students, who then complete the homework as if it were a normal class assignment. The online platform would automatically provision different forms (randomly assigning experimental conditions) to different students, and would record detailed web logs from student interactions with the assignment. Finally, teachers report relevant learning outcomes to the researchers for individual students, according to disciplinary standards (e.g., as assessed by a subsequent test), and researchers conduct analyses on the combined dataset. As a rule, all teachers who added content and arranged for their classes to participate in the ManyClasses study are authors on the ensuing research report.

As envisioned here, ManyClasses is a flexible model for evaluating the generalizability of learning principles across educational contexts, and is not limited to a specific online platform or even to a specific set of researchers. Teams might initiate a ManyClasses study using an online learning platform such as ASSISTments (Heffernan & Heffernan, 2014), an online survey platform such as Qualtrics, or by designing their own custom web application. Learning tool interoperability (LTI) standards would allow these kinds of web-based learning objects to be seamlessly embedded in teachers’ local learning management systems (LMSs). As a solution for examining learning in different contexts, ManyClasses shares common ground with data-sharing tools that pool educational research data across individual studies, such as DataShop (Koedinger et al., 2010) and LearnSphere (http://learnsphere.org/), but ManyClasses is more focused than these, adopting a single specific research question, and replicating a particular experimental design template in different classes.

A controlled educational experiment that brings together dozens of teachers in different disciplines at different institutions presents no shortage of challenges, but also represents an ideal model for drawing inferences about the generalizability of learning principles across educational contexts. In this proposal, we describe a preliminary vision for how ManyClasses might work in practice, and seek feedback from the community about this vision.
References


