Rigor and Relevance in an Efficacy Study of an Online Mathematics Homework Intervention

Jeremy Roschelle, Robert Murphy, and Mingyu Feng (SRI International)
Craig Mason, Janet Fairman (University of Maine)

A randomized controlled trial of an online mathematics homework was conducted to evaluate efficacy while also providing insights relevant to school policies and practices. Key opportunities to connect rigor with relevance occurred in the conceptualization of the study, the research design and data analysis.

Background
Assigning, doing, and reviewing homework is a common practice in middle school mathematics instruction (Loveless, 2014), yet students, teachers and parents and the general public often question the value of homework (Eren & Henderson, 2011; Galloway & Pope, 2007). The relevance of our research is anchored in this instructional challenge. A research-based intervention supporting online mathematics homework called “ASSISTments” was ready for a rigorous experimental trial at the scale required to obtain an estimate of impact. ASSISTments was designed to fit in the daily work of students and teachers to leverage theoretically-driven principles of formative assessment (Black & Wiliam, 1998; Bennett, 2011) and mastery-oriented practice (Paschler et al, 2007) to enhance learning and had shown promise in previous small scale experiments (Mendicino, 2009; Kelly et al, 2013).

Purpose
The study was designed to span issues of rigor and relevance. We planned to:
- evaluate impact on student learning and moderators of impact
- investigate changes in teacher practice and how such changes mediated student outcomes
- examine factors that enabled or impeded adoption and implementation, as well as the program costs.

Setting
Maine provided a strong setting because this state provides a laptop to every middle school student to take home, and because whether the one-to-one technology initiative improves learning is a topic of current public scrutiny in Maine (Herrold & Kazi, 2016).

Population
The study population consisted of 43 middle schools in the state of Maine, who volunteered for the study, along with 2850 7th grade mathematics students and their teachers.
**Intervention**  
ASSISTments comprises both a technology and teacher professional development. The technology implements formative assessment by giving immediate feedback on homework to students and reporting on homework performance to teachers. Further, a feature called “Skill Builders” organizes practice problems for students based on principles such as spaced practice and mastery learning. The professional development helps teachers to use the system reports to adapt their review of homework based on student difficulties.

**Research Design**  
To evaluate the impact of ASSISTments, we designed a randomized controlled trial with two conditions and assigned schools to conditions. Teachers participated in the experiment for two years. In year one, teachers in treatment schools learned to use the intervention. In year two, treatment teachers used the system with a new cohort of students while data was collected. Control teachers proceeded with a business-as-usual approach to homework for two years, but then had full access to ASSISTments. Complementary aspects of study included interviews of school administrators to understand their policies and concerns about homework, which contributed to relevance. An embedded study of how teaching practice changed because of the intervention had both elements of rigor (use of multiple measures) and relevance (both observing and interviewing the teachers to better understand their perspective). The overall research design was balanced to both yield a rigorous estimate of impact but also to illuminate questions of policy and practice and thus to increase relevance.

**Data Collection**  
The student achievement outcome measure was the TerraNova Common Core Edition. We collected students’ prior achievement scores and demographic variables from a statewide database. We interviewed both administrators and teachers. A sample of teachers was directly observed. All teachers were asked to take surveys and complete logs on specific teaching days. We also collected system data in ASSISTments to examine how the system was used, and collected cost of implementation data through interviews.

**Data Analysis**  
Efficacy was analyzed using a hierarchical linear model with student-, classroom-, and school-level covariates and the outcome measure. Interviews were analyzed systematically, as were observations. Surveys and logs were analyzed both descriptively and then in analytic models that extended from the base HLM model. System data was also analyzed descriptively and will be used in analytic models in ongoing work.

**Findings**  
After controlling for student prior achievement and other covariates, the experiment found a main effect of condition; students in the treatment schools had higher TerraNova scores.
(t(20)=2.992, p=.007) compared to students in control schools (Roschelle et al, in press). The effect size was $g=0.18$ and was stronger for students with low prior math achievement (see Figure 1). Complementary analyses have yielded a consistent account of change in teaching practice. Where control teachers tended to review all homework problems each day, treatment teachers used the same amount of time, but reviewed fewer problems. They used ASSISTMents reports to target their review to focus on student difficulties and errors. We are currently analyzing how this change in practice mediated outcomes.

Additional analyses helped frame the relevance of the findings. We learned from school principals that school policy required homework, and they wanted teachers to use data to improve learning, hence the intervention fit an existing policy and need. We learned from teachers that the tool reduced their teaching load (for example, they spent less time grading homework), which was a boon to adoption, and that some students were initially demotivated by immediate feedback after answering a problem incorrectly and teachers had to develop ways to counter this. We learned that the costs to schools as implemented were modest and that teachers viewed the intervention favorably. Hence the efficacy finding (rigor) was supported by complementary policy, practice, and cost findings (relevance).

**Conclusion**
The ASSISTments intervention had a positive impact on student outcomes over the course of a full school year. Many tools can provide feedback to students and can report student data to teachers, but ASSISTments was both teacher-friendly and teacher-practice-oriented. It saved teachers’ time and enabled them to target homework reviews to student difficulties. It implemented research-based theoretical principles. It fit to typical school policies and administrator concerns and had reasonable costs. Overall, we foresee enhanced opportunities to communicate this research and its findings to practitioners because it addressed both rigor and relevance.

**References**


Figure 1. Students were divided into a low and high groups based on prior achievement scores. The results shown are the observed TerraNova scores. In both groups, the students in Treatment schools obtained higher TerraNova scores. The difference between treatment and control groups was greater for students with lower prior achievement (t(2770)=2.432, p=0.015). The TerraNova is scaled from 0-1000 from kindergarten through the end of high school. However, this experiment concerned only the 7th-grade test, which uses a smaller range of possible scale scores. We scaled the vertical axis to show the minimum possible score on the 7th-grade test (487) at the low end and the largest observed results on the high end of the axis. (reprinted from Roschelle et al, in press).