

## Equity in Scaling Up SimCalc: Investigating Differences in Student Learning and Classroom Implementation

Jeremy Roschelle, SRI International, [jeremy.roschelle@sri.com](mailto:jeremy.roschelle@sri.com)  
 Jessica Pierson, San Diego State University, [jpierson@mail.sdsu.edu](mailto:jpierson@mail.sdsu.edu)  
 Susan Empson, University of Texas at Austin, [empson@mail.utexas.edu](mailto:empson@mail.utexas.edu)  
 Nicole Shechtman, SRI International, [nicole.shechtman@sri.com](mailto:nicole.shechtman@sri.com)  
 Margie Dunn, Rutgers University, Newark, [dunn@cs.rutgers.edu](mailto:dunn@cs.rutgers.edu)  
 Deborah Tatar, Virginia Tech, [tatar@vt.edu](mailto:tatar@vt.edu)

**Abstract:** The Scaling Up SimCalc project implemented three large-scale studies designed to evaluate the impact of a Learning Sciences-based replacement unit targeting student learning of advanced middle school mathematics. Strong main effects in each study consistently showed the approach to be effective in enabling a wide variety of teachers in a diversity of settings to extend student learning to more advanced mathematics. In this paper, we take a closer look at *equity* of learning, the extent to which students within and across subgroups and classrooms had comparable learning gains. We describe four different patterns of equity and inequity that have emerged in our hierarchical linear modelling and in-depth case studies. We then present a number of conjectures about possible mechanisms by which opportunity to learn may be unevenly distributed between and within classrooms engaged in the same replacement unit. We discuss implications for instructional design, research methodology, and Learning Sciences theory.

### Introduction

In an on-going research program, we have been studying large-scale implementations of the Learning Sciences-based SimCalc approach to integrating representational technology, curriculum workbooks, and teacher professional development (see Roschelle et al., in press). The SimCalc program's slogan, "Democratizing Access to the Mathematics of Change and Variation," suggests a significant concern with equity (Kaput, 1994). Indeed, the design experiment research preceding our current large-scale work focused deliberately on giving diverse students in traditionally underperforming settings an opportunity to learn more advanced mathematics concepts. Hence, design experiments were conducted in diverse and economically challenged places such as Fall River, Massachusetts; Syracuse, New York; Newark, New Jersey; and San Diego, California (Roschelle, Tatar, & Kaput, 2008). This concern with equity was also reflected in the subsequent large-scale experiments, which occurred in Texas. We deliberately recruited schools in multiple regions of Texas to reflect that state's considerable economic, linguistic, and ethnic diversity (Roschelle et al., in press).

Equity has been an elusive goal in the reform of mathematics education. For example, prior research found that middle- to higher-achieving students benefitted from instruction that focused on integrating concepts and procedures, but lower-achieving students did not (Baxter, Woodward, & Olson, 2001); and achievement gaps in mathematics between white students and non-white students persist (Neal, 2005). Furthermore, there are competing definitions of equity (Lynch, 2000). Does it involve providing high-quality resources for all classrooms, ensuring that learning activities have ways for all students to participate meaningfully, closing achievement gaps, or perhaps some combination of these things? In this paper, we explore equity in the context of a large-scale implementation of a Learning Sciences-style approach, one integrating dynamic representations and visualizations to address learning of more advanced mathematical constructs.

After providing a brief overview of the program of research, we examine equity in the Scaling Up SimCalc Project. First, we describe four patterns of equity and inequity we have observed related to student achievement. We use a variety of indicators across multiple units of analysis. Second, we consider a variety of conjectured explanations for inequitable classrooms that are emerging in ongoing data analyses. Two highlighted conjectures are blocked access to learning resources and teacher talk that is not responsive to students and does not place a high level of intellectual demand on students. Other possible explanations for inequity are also considered.

### Overview of the Scaling Up SimCalc Project

The Scaling Up SimCalc Project (Roschelle et al., in press) implemented two randomized controlled experiments (with one embedded quasi-experiment) designed to address the broad research question, "Can a wide variety of teachers use an integration of technology, curriculum, and professional development to increase student learning of complex and conceptually difficult mathematics?" There were two interventions, one for seventh grade and one for eighth grade. Each intervention integrated the representational technology SimCalc

MathWorlds, curriculum workbooks, and teacher professional development organized around a 2-3 week replacement unit on rate, proportionality, and linear function. The replacement units incorporated the following hallmarks of the SimCalc approach to the mathematics of change and variation:

1. Anchoring students' efforts to make sense of conceptually rich mathematics in their experience of familiar motions, which are portrayed as computer animations.
2. Engaging students in activities to make and analyze graphs that control animations.
3. Introducing piecewise linear functions as models of everyday situations with changing rates.
4. Connecting students' mathematical understanding of rate and proportionality across key mathematical representations (algebraic expressions, tables, graphs) and familiar representations (narrative stories and animations of motion).
5. Structuring pedagogy around a cycle that asks students to make and compare their predictions.

The SimCalc MathWorlds software provides a “representational infrastructure” (Kaput, Hegedus, & Lesh, 2007) that is central to enabling this approach. Most distinctively, the software presents animations of motion. Students can control the motions of animated characters by building and editing mathematical functions in either graphical or algebraic forms. The program developers view student use of the software and teacher explanations and teacher-led discussions as complementary activities.

In the scaling research, across over 100 seventh- and eighth-grade classrooms throughout the state of Texas, the main effects of the treatment in the three main studies were positive, with student-level effect sizes of .63, .50 and .56; classrooms that used a SimCalc replacement unit had students who learned more advanced mathematics.

## Four Patterns of Equity and Inequity

Programs can be more or less equitable at a variety of levels. While overall there were robust learning gains, here we drill down more deeply at demographic and regional levels, as well as the classroom level, to examine three patterns of equity and inequity.

### 1. Equity at the Demographic and Regional Levels

As Figure 1 shows, at the highest level of aggregation, we found learning gains to be equitable across major demographic and regional groups. In the treatment group, we found no statistically significant differences in gains between gender or ethnic groups, or among students in schools serving different levels of socioeconomic populations. We also found no differences among geographic regions. This is particularly notable for “Region 1” of Texas, which covers the Rio Grande Valley along the Mexican border and has a high-poverty population mainly of Mexican descent. The effects of SimCalc were similar in this region to those in other regions sampled in the study (e.g., the more cosmopolitan regions near the state capital, Austin). These findings of equity were replicated in the Eighth-Grade study (see Roschelle et al., in press).

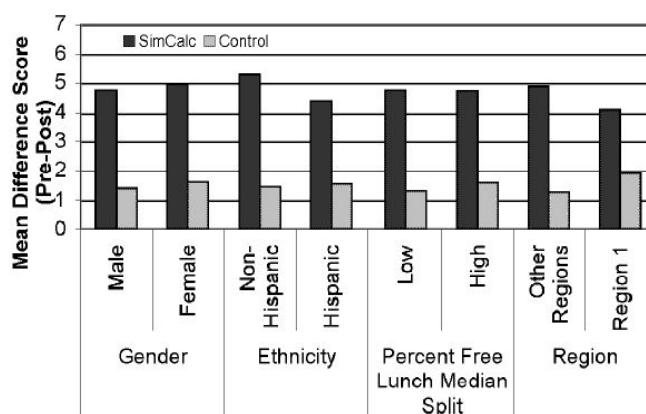


Figure 1. Mean student-learning gains by subpopulation group (Seventh-Grade Study).

### 2. Within-Classroom Spread of Learning Gains

While equity of learning occurred at this highly aggregated level, at the classroom level, do students learn equitably? To begin to examine this issue, we discuss the work of Empson, Greenstein, Maldonado, and Roschelle (2009) who report comparative case studies of three teachers who participated in the Seventh-Grade Study and taught with the SimCalc replacement unit. Two of the teachers, “Ms. Garfield” and “Ms. Driver” (pseudonyms), had mean classroom gains at or above the average mean gain found in the treatment group. The third teacher, “Mr. Simmons,” had a mean classroom gain that was lower than the average by a third of a

standard deviation. Note that mathematical knowledge for teaching (MKT) was not a predictor of student learning in these three cases. Mr. Simmons scored the highest of the three on our MKT assessment. Indeed, he was a former computer specialist. In a more detailed analysis of MKT we argue that MKT is just one of a variety of resources in the classroom instructional system (Shechtman, Roschelle, Haertel, & Knudsen, in press).

To investigate equity, we examine two aspects of learning in the classroom: overall gains and variation between students in gains. As Figure 2 shows, while the median gain in Mr. Simmons' class was relatively low, there was also a notable spread in achievement among his students. A few students performed very well in this classroom, but many students did much more poorly. In the two other cases, gains were both higher and less variable (as evidenced by less spread) among students. If one defines equity as equal learning gains for *all* students, then Mr. Simmons' classroom illustrates inequity.

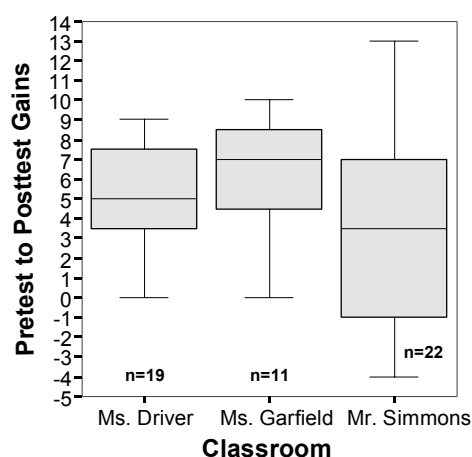


Figure 2. Boxplots of distributions of student gains on the 30-item assessment across the case study classrooms.

### 3. Within-Classroom Pretest-Posttest Slopes Tend to Be Steeper in Lower-Achieving Classrooms

Another approach to investigating equity is by examining the relationship within a classroom between students' prior knowledge (i.e., their pretest scores) and their learning gains. Pierson's (2008) dissertation investigated some of these patterns across classrooms. To illustrate, Figure 3 shows student achievement data in which each student is represented by an ordered pair where the x-coordinate represents his or her pretest score, and the y-coordinate represents the corresponding posttest score. The figure displays such data and best-fit regression lines for two classrooms. The variable of interest with respect to equity is the pretest-posttest achievement slope. Some classrooms have steeper slopes and some have flatter slopes. In other words, pretest scores are stronger determinants of achievement in some classrooms. Some claim that more equitable classrooms are those in which the pretest-posttest slope is closer to 0. In other words, *all* students achieve equally regardless of their prior knowledge.

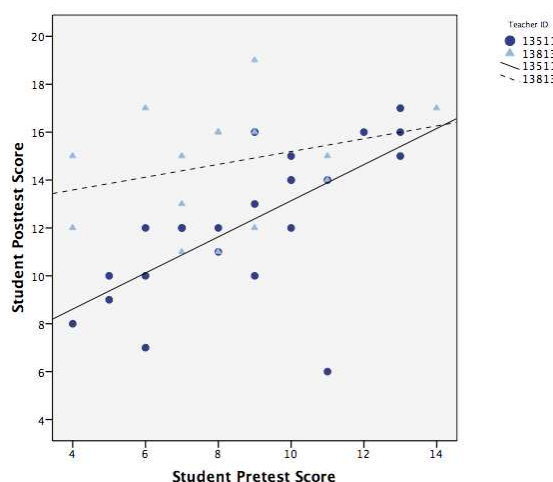


Figure 3. Scatter plot and regression lines for student pretest and student posttest scores in two SimCalc classrooms.

Pierson found in her HLM model a negative relationship between the mean achievement and pretest-posttest slopes at the class level ( $\tau_{01} = -.4309$ ). In other words, the higher the class posttest average, the smaller the pre-post achievement slope.<sup>1</sup> Classrooms that learned more overall (i.e., had higher average posttest scores) also learned more equitably.

#### 4. The “Low-Low” Effect

We also investigated equity of learning across students’ prior achievement levels (both as rated by their teachers and indicated by their pretest scores). We asked teachers to rate their students as low, medium, or high achieving prior to using the SimCalc replacement unit. Analyses found that the teachers’ rating accurately predicted outcomes: students designated by their teacher as low-achieving learned less. Of course, the opposite effect would be desirable: if teachers knew which students were low-performing, it would be preferable for those students to receive *more* resources and guidance so that they could learn more.

More broadly, we are investigating a “low-low effect,” designated as such because it appeared for low-ranked students in classrooms with low pretest means. We found this effect in both our Seventh- and Eighth-Grade studies, but for brevity we only present the Seventh-Grade data in Figure 4. Three graphs are presented side by side, representing classrooms with low-, medium-, and high-mean pretest scores (based on a tertile split across the treatment group). Each graph plots the teachers’ categorization of their students (as low, medium, or high achievers) against the mean of those students’ gain scores. The students in the low classroom have a particular pattern: low students in low classrooms had the lowest gains.

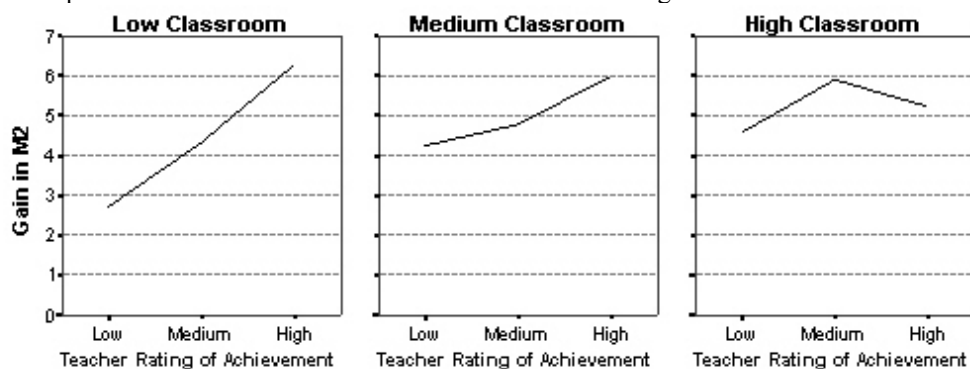


Figure 4. The “low-low effect” in the Seventh-Grade treatment group.

We have built an HLM of this phenomenon, and it is statistically significant in both our Seventh- and Eighth-grade studies. For each study, we ran an HLM of student gains, nesting students within classrooms. We then constructed the model using classroom mean pretest scores as a predictor at the classroom level and indicator variables of teacher-rated student achievement level (low and middle levels, with high level as the reference category) as predictors at the student level. To test for the low-low effect, we examined the coefficient for the interaction between classroom pretest and the indicator for low-achieving student. This coefficient was positive and significant in both the seventh-grade study ( $\beta = 0.29, p < .05$ ) and the eighth-grade study ( $\beta = 0.40, p < .001$ ), indicating the presence of the low-low effect. We also tested the same model on control classrooms where SimCalc materials were not used and found no presence of low-low effect. This suggests that the low-low effect is specific to the SimCalc materials.

We are in the process of investigating the low-low effect more deeply to examine possible alternative explanations for this phenomenon, including the possibility that it is an artifact of some kind. For example, it could be that only low classrooms have students with extremely low pretest scores. And it could be that students need a certain level of knowledge at baseline in order to engage meaningfully with the curriculum, or that teachers need certain additional skills to use these materials that engage lower-achieving students. For present purposes, the important point is that we see this as an indication of more and less equitable implementations of the SimCalc replacement unit.

### Conjectured Mechanisms of Within-Classroom Equity and Inequity

While our research was not designed to pinpoint particular mechanisms that produce more or less equitable classroom implementations, empirical analyses have produced a number of candidate explanations.

#### 1. Blocked Access to Learning Resources during Instruction

One conjecture that has emerged in Empson et al.’s (2009) case studies is that some teachers may block learning resources from students, thus causing inequitable learning opportunities. This grows out of the idea that

SimCalc can work in multiple configurations of learning resources. Full class discussion around a shared simulation display, guided by a knowledgeable teacher, can be one kind of resource. Individual student activities with the software, guided by a workbook, can be another kind of resource. Empson and her team found that Ms. Garfield and Ms. Driver each had classroom implementations which featured different configurations of resources. What was distinctive about Mr. Simmons was that he blocked access to both full class discussion and individual student activities for many students. For example, Mr. Simmons concentrated whole-group discussion on interaction with a small subset of vocal boys, who made over 10 times more contributions than rest of the class. He also curtailed students' interactions with SimCalc by specifying each step to take in running the simulation and by requiring students to frequently turn away from the computers. Blocking access to resources in these ways decreased students' opportunities to engage meaningfully with the content and to make substantive connections on their own.

The pattern of blocking access was also prominent in Dunn's case studies in the Eighth-Grade study (Dunn, 2009). One teacher, "Marilyn," had particularly low student gain scores. She was on the low end of the range in total time allocated to teaching with the SimCalc materials, and she skipped more material than the other teachers. In particular, Marilyn seemed to have a pattern of beginning each day's class with a new workbook lesson, regardless of whether there had been sufficient time to complete the previous workbook lesson in the previous class sessions. As the workbook was designed to build ideas sequentially, skipping material in this way could block coherent development of mathematical ideas. Marilyn also was at the high end of the spectrum for the time spent introducing students to ideas via lecture, with a tendency to "walk" students through significant portions of each workbook lesson, in most cases at the beginning of class. This preempted independent student interaction with these portions, and compromised the time available for subsequent autonomous student work and class discussion/teacher feedback on this later work. This uneven allocation of time and guidance blocked consistent access and, once again due to the sequential design of the materials, may also have blocked coherent development of the unit's mathematical ideas.

The complement to the idea of blocked access is the idea of multiple learning resources and multiple configurations of learning opportunities. It could be that a key design principle for equity on a large scale is to provide multiple means for students to engage with the big ideas within the same intervention.

## 2. Favorable Teacher Discourse Moves

A second major conjecture examines the role of particular types of teacher talk in the classroom – responsiveness to student ideas and how much intellectual work they demand of students. Pierson's (2008) dissertation used HLM to examine teacher discourse across 13 teachers who taught using SimCalc materials. In particular, Pierson analyzed videotapes and transcripts of the same SimCalc lesson for all teachers. She developed the ideas that teachers (a) can be more or less responsive<sup>2</sup> to their students and (b) can demand more or less intellectual work<sup>3</sup> from their students. Using a slightly different HLM model than that described above for the low-low effect, she ran her analysis on student posttest achievement scores, nesting students within classrooms. She used classroom discourse variables as predictors at the classroom level and pretest achievement scores as a predictor at the student level.

There were two findings relevant to this discussion of equity. First, Pierson found that both high responsiveness and high intellectual work correlate with classroom-level *gains*. Second, these factors also correlated with the pretest-posttest slope (see discussion above). Specifically, Pierson found that an *increase* of 10% in highly responsive moves corresponded to a *decrease* in the class's pretest-posttest slope of .29. Interestingly, responsiveness and intellectual work seem to be factors in more equitable classrooms, as shown by a decrease in the impact of pretest scores on posttest achievement.

Although Pierson's study showed good evidence *that* high levels of responsiveness and intellectual work have a positive relationship with more equitable patterns of achievement, we should consider *how* these discursive constructs might be related to equity. We hypothesize the following three more refined mechanisms as possible links between equitable achievement and patterns of discourse.

First, discussions high in responsiveness and intellectual work can act as formative assessments, which research indicates is positively related to student learning (Fennema et al., 1996; Wiliam, Lee, Harrison, & Black, 2004). These types of discourse moves afford teachers opportunities to gather evidence of student understanding in order to adjust and customize instruction, resulting in better learning opportunities for everyone.

Second, repeated patterns of responsiveness and intellectual work in classroom discourse socialize students into specific ways of being. These normative discourse patterns communicate expectations for how to engage with one another (listening to, learning from, and critiquing others' ideas); expectations about learning (mistakes are a natural process in learning; and learning is about thinking, not remembering); and beliefs about one's own abilities. Students, particularly low-achieving students, are likely to develop increased confidence, motivation, persistence, and more positive mathematical identities as their ideas are valued, taken up, and extended.

Third, high levels of intellectual work and responsiveness create classroom cultures that require students to routinely work on cognitively demanding tasks that require argumentation, justification, and cognitive struggle. The teachers' role here is critical in that they must uphold the expectation that all students will generate their own solutions and explain them, position students as capable problem solvers, and carefully orchestrate student involvement in classroom discussions. Creating this type of learning community might benefit students with lower levels of prior knowledge reduce that gap and achieve at levels comparable with their peers at the top of the class.

### 3. Emerging Conjectures in Other SimCalc Case Studies

Other case studies conducted in the context of the Scaling Up SimCalc studies are still underway:

- Work by Michelle McLeese and Deborah Tatar considers the social climate of the classroom. This study is exploring the idea that certain forms of laughter signal a more positive social climate and contribute to better learning outcomes.
- Another study, by Meg Kurdziolek and Deborah Tatar considers the configuration of technology in the classroom. For example, they observed SimCalc implementations in a computer lab, with a laptop cart in the teacher's regular classroom, and with only one computer per classroom. Surprisingly, the implementation in the classroom with only one computer was successful; the teacher found ways to actively involve all students despite the resource limitation. They also have found instances where a classroom has more computers and students work in pairs, but a more dominant partner blocks engagement of a less dominant partner. Likewise, both Dunn (2009) and Empson et al. (2009) observed classrooms with a sufficient number of computers where student activity with the computer was routinely interrupted or blocked by the teacher, leading to poor outcomes. At the very least, this suggests that analysis of equity in relationship to the numbers of computers in a classroom needs to take into account pedagogical adaptations and patterns of student interaction when they work in small groups.
- We are also analyzing workbook materials collected in diverse classrooms, containing student work. Preliminary analysis of these workbooks has found strong correlations from students' pretest scores to the completeness and correctness of their workbooks. This may suggest an equity concern that students who begin with higher math scores are better prepared to use the workbooks. Furthermore, because completing relevant portions of the student workbook is correlated with student gains, providing more support to students who are falling behind in completing the workbook might contribute to more equitable outcomes.

As these additional studies mature, they may contribute more nuance to our concepts of access to learning resources and favorable teacher discourse moves, or may contribute additional equity-relevant constructs.

## Discussion

Equity is an important policy objective and often an important goal in the design of interventions resulting from Learning Science theories and design experiments.

By presenting results from the Scaling Up SimCalc work, we have highlighted a few key ideas:

1. Equity and inequity occur simultaneously on different levels and scales; equity may be present at some levels and absent in others.
2. Classroom implementations of the same intervention can be more or less equitable.
3. A range of mechanisms may contribute to equity and inequities in particular classroom interventions.

With regard to the last point, we identified two relevant mechanisms. First, even the best designed learning resources are not useful if students' access to those resources is blocked. Second, when teachers are responsive to student ideas and engage students in doing public intellectual work, all students seem to benefit, not just those students that the teacher is interacting with. However, findings pertaining to students in the low-low effects cases suggest that additional teacher development may be needed in order for instruction to support lower-achieving students' productive engagement with the content. We need to know more about how the teaching knowledge and practices are involved in being responsive and posing questions with high intellectual demand for these students.

These observations may benefit *design*: if we become aware of different sources of inequity, we may be able to design features into the intervention to mitigate some. For example, we could give teachers more guidance about making SimCalc resources available to students in multiple configurations and giving them time to use the materials autonomously.

The observations may also benefit *method*. We suggest that a mix of large-scale experimental designs with embedded case studies is a particularly fruitful method to ferret out issues of inequity, for a few reasons. Large-scale studies can be designed with a structure that captures data at the multiple levels at which equity and

inequity can occur. For example, our designs considered the regional, school, classroom, and student levels. Hierarchical linear modeling can highlight cross-level patterns of inequity; some classrooms may have flatter slopes correlating student pretests to gains (i.e., more equitable outcomes), while other classrooms have steeper slopes (e.g., the rich get richer). Case studies, however, are also important for detecting mechanisms of inequity. For instance, it was far from obvious to us at the beginning of the experiments that we might see one-computer-per-classroom implementations that were more equitable than implementations with multiple computers.

We also suggest that this research could lead to fruitful paths in Learning Science *theory*. Although much Learning Science research looks at discourse, less looks at the implications of various classroom discourse strategies for equity. The finding that responsiveness and intellectual demand may help level the playing field in the classroom suggests that the subject of teacher discourse moves is worthy of further investigation.

Overall, we are encouraged about the range of settings in which we found equitable outcomes and energized by the potential to better pinpoint sources of inequity and to better address them in future design iterations.

## Endnotes

<sup>1</sup> A ceiling effect is unlikely because highest average pretest score for any class in Pierson's study was 9.00 points/problems (out of 18 possible) with the average pretest score across all classes being 5.877 points ( $\bar{x} = 2.28$ ).

<sup>2</sup> *Responsiveness* is an attempt to understand what another is thinking displayed in how one builds, questions, clarifies, takes up, or probes that which another says.

<sup>3</sup> *Intellectual work* reflects the cognitive work requested from students with a given turn of talk. High levels of intellectual work extend thinking and include discursive moves, such as providing justifications, examples, conjectures, explanations, and challenges; making connections across representations; generating problems and scenarios (contextualizing); or requesting these activities from students.

## References

- Baxter, J., Woodward, J. & Olson, D. (2001). Effects of reform-based mathematics instruction on low achievers in five third-grade classrooms. *The Elementary School Journal*, 101, 529-547.
- Dunn, M. (2009). *Investigating variation in teaching with technology-rich intervention: What matters in teaching and teacher training at scale?* Unpublished Ed.D., Rutgers University, New Brunswick, NJ.
- Empson, S. B., Greenstein, S., Maldonado, L., & Roschelle J. (2009). Scaling up innovative mathematics in the middle grades: Case studies of "good enough" enactments. Manuscript submitted for publication.
- Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, 27(4), 403-434.
- Kaput, J. (1994). Democratizing access to calculus: New routes using old roots. In A. Schoenfeld (Ed.), *Mathematical thinking and problem solving* (pp. 77-155). Hillsdale, NJ: Erlbaum.
- Kaput, J., Hegedus, S., & Lesh, R. (2007). Technology becoming infrastructural in mathematics education. In R. Lesh, E. Hamilton & J. Kaput (Eds.), *Foundations for the Future in Mathematics Education* (pp. 173-192). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lynch, S. J. (2000). *Equity and science education reform*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Neal, D. (2005). Why has Black-White skill convergence stopped? Retrieved June 2005 from <http://home.uchicago.edu/~n9na/>.
- Pierson, J. (2008). *The relationship between patterns of classroom discourse and mathematics learning*. Unpublished doctoral dissertation, The University of Texas at Austin.
- Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J. & Gallagher, L. (in press). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal*.
- Roschelle, J., Tatar, D., & Kaput, J. (2008). Getting to scale with innovations that deeply restructure how students come to know mathematics. In A. E. Kelly, R. Lesh & J. Y. Baek (Eds.), *Handbook of design research methods in education* (pp. 369-395). New York: Routledge.
- Roschelle, J., Tatar, D., Shechtman, N., & Knudsen, J. (2008). The role of scaling up research in designing for and evaluating robustness. *Educational Studies in Mathematics*, 68, 149-170.
- Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (in press). Investigating links from teacher knowledge, to classroom practice, to student learning in the instruction system of the middle school mathematics classroom. *Cognition and Instruction*.
- Wiliam, D., Lee, C., Harrison, C. & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11(1), 49-65.

## **Acknowledgments**

This material is based in part upon work supported by the National Science Foundation under Grant Number 04-37861. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.