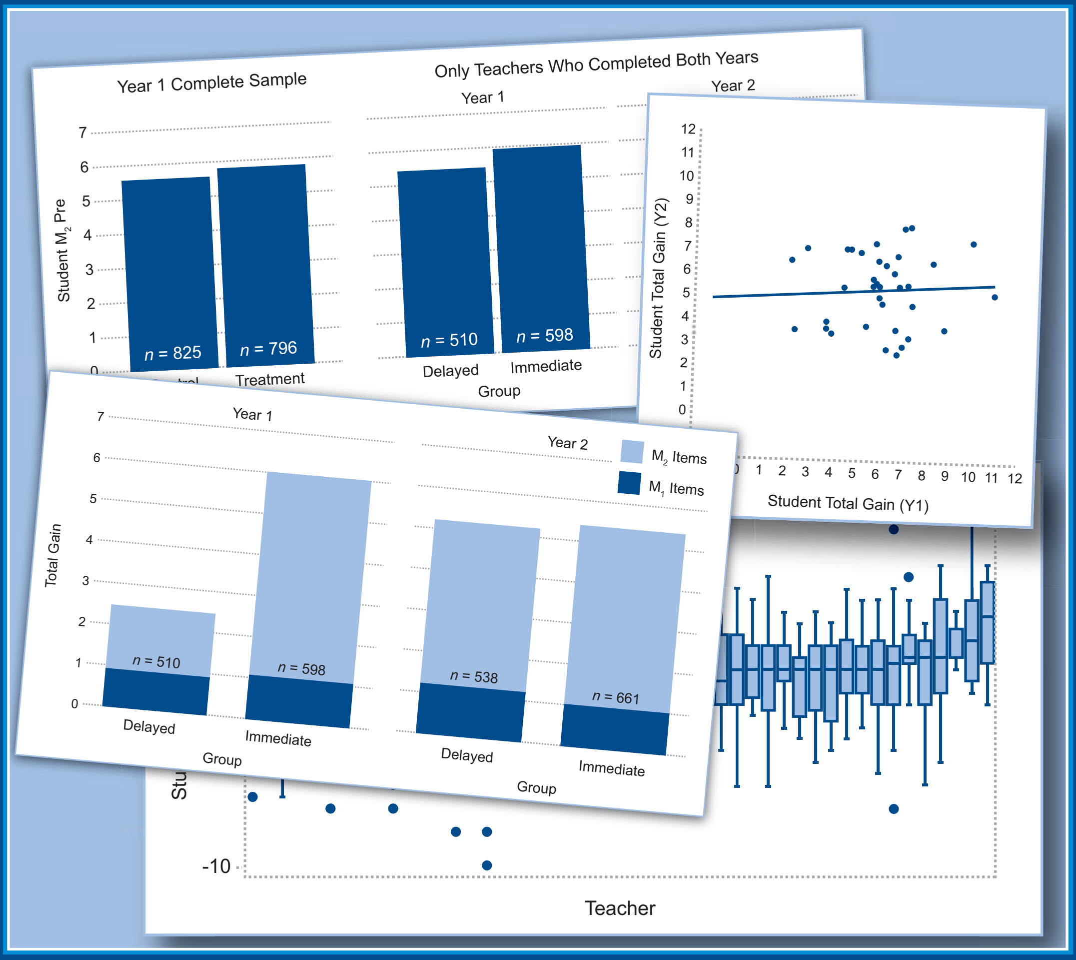
Scaling Up SimCalc Project

The Data Outtakes Reel: Archive of Unreported, Unreportable, and Irreproducible Findings



Technical Report 08 | April 2010

Nicole Shechtman, Sara Carriere, Jeremy Roschelle

Report Series Published by SRI International

Research collaborators:

University of Massachusetts, Dartmouth; Virginia Polytechnic Institute and State University;
The University of Texas at Austin; and the Charles A. Dana Center at the University of Texas at Austin

SRI International
Center for Technology in Learning
333 Ravenswood Avenue
Menlo Park, CA 94025-3493
650.859.2000

The Data Outtakes Reel: Archive of Unreported, Unreportable, and Irreproducible Findings

Prepared by:

Nicole Shechtman, SRI International

Sara Carriere, SRI International

Jeremy Roschelle, SRI International

Acknowledgments

This report is based on work supported by the National Science Foundation under Grant No. 0437861. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation.

We thank G. Haertel, J. Knudsen, K. Rafanan, P. Vahey, S. Carriere, L. Gallagher, H. Javitz, T. Lara-Meloy, M. Robidoux, S. Empson, S. Hull, L. Hedges, S. Goldman, H. Becker, J. Sowder, G. Harel, P. Callahan, F. Sloane, B. Fishman, K. Maier, J. Earle, R. Schorr, and B. McNemar for their contributions to this research. We thank the participating teachers and Texas Educational Service Center leaders from regions 1, 6, 9, 10, 11, 13, 17 and 18; this project could not have happened without them. We thank and remember Jim Kaput, who pioneered SimCalc as part of his commitment to democratizing mathematics education.

The Data Outtakes Reel: Archive of Unreported, Unreportable, and Irreproducible Findings

To inform future researchers as they design and implement large-scale, multifaceted research, this technical report documents the unreported, unreportable, and irreproducible findings on the cutting room floor of the Scaling Up SimCalc Project. It also provides a few good statistical puzzles.

There are three kinds of lies: lies, damned lies, and statistics.
— Mark Twain

Introduction

The Scaling Up SimCalc Project comprised a set of large-scale studies implemented throughout the state of Texas over 4 years. Since 2004, in peer-reviewed journal articles, book chapters, conference presentations, technical reports, doctoral dissertations, and master's theses (listed in Appendix A), we have reported a rich variety of findings and methodological issues. The core research design and findings are summarized in the abstract in our article in the *American Educational Research Journal:*

> We present three studies (two randomized controlled experiments and one embedded quasi-experiment) designed to evaluate the impact of replacement units targeting student learning of advanced middle school mathematics. The studies evaluated the SimCalc approach, which integrates an interactive representational technology, paper curriculum, and teacher professional development. Each study addressed both replicability of findings and robustness across Texas settings with varied teacher characteristics (backgrounds, knowledge, attitudes) and student characteristics (demographics, levels of prior mathematics knowledge). Analyses revealed statistically significant main effects, with student-level effect sizes of .63, .50, and .56. These consistent gains support the conclusion that SimCalc is effective in enabling a wide variety of teachers in a diversity of settings to extend student learning to more advanced mathematics. (Roschelle et al., 2010, see Appendix A)

That article and other publications present our main findings; the role of mathematical knowledge for teaching; theory about the use of representational infrastructure; technology and English language learners; qualitative and quantitative examinations of variations in classroom implementations; impacts of the professional development model, learner identity, and teacher and student ethnicity; and methodological considerations in large-scale research implementation and assessment development.

Here, we document the data outtakes that have been lying on the cutting room floor—findings that are not in any of these reports. Our vast dataset (including more than 100 unique sources of quantitative and qualitative data collected over the 4 years) yielded additional findings that were unexpected, too puzzling, or in some way interesting but not highlighted in a major report. We document some of them here to inform future researchers as they design and implement large-scale, multifaceted research.

In addition, Appendix B exhaustively documents our modeling findings across all the substantive variables we measured.

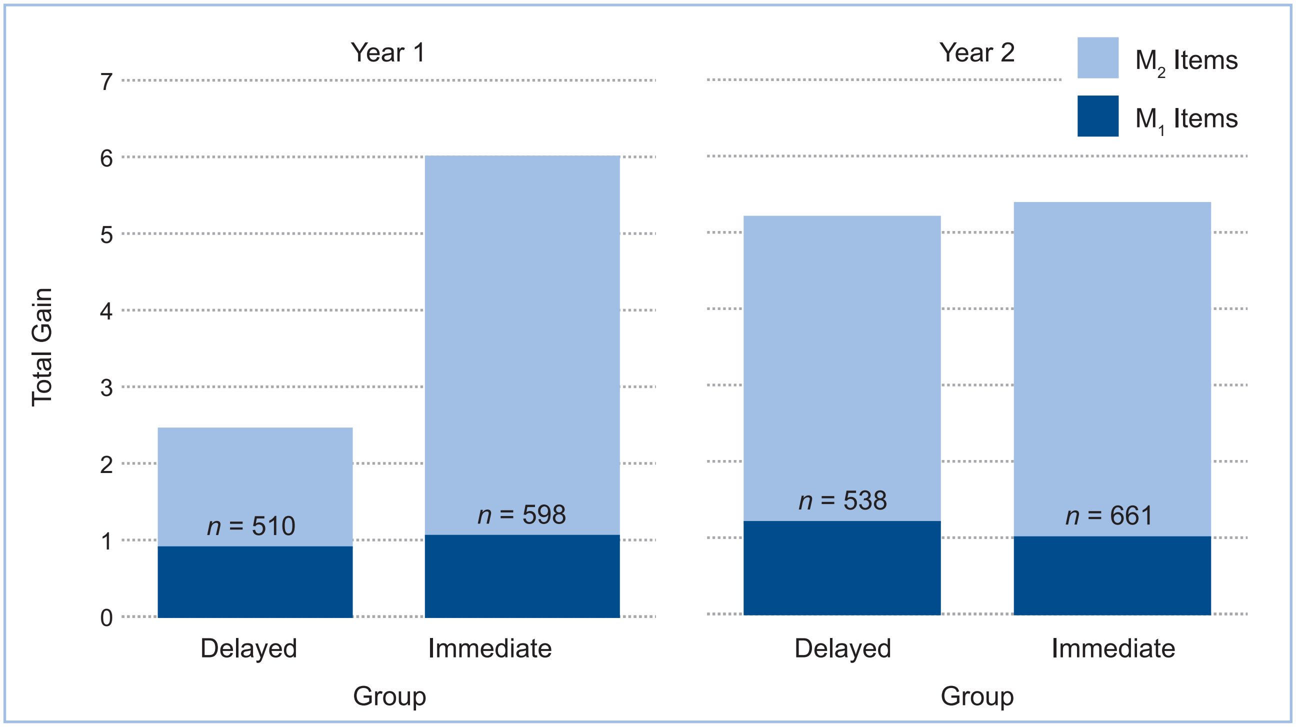
The Immediate Treatment Group in Year 2?

In the second year of the experiment, some teachers (the immediate treatment group) were using the SimCalc intervention for the second time. Before the experiment began, our team members and advisors had speculated about whether students' test scores would increase when teachers implemented the intervention again.

Yet readers familiar with our publications may note that we chose to report the Year 2 results for the other group of teachers, those who did not implement SimCalc in the first year but did in the second (the delayed treatment teachers). We viewed this delayed treatment group as representative of an embedded quasi-experiment comparing the implementations of delayed treatment teachers in Year 1 (doing business as usual) and Year 2 (using SimCalc). This focus was useful because it provided an additional replication of our Year 1 experimental comparison between nonimplementing and implementing teachers.

But let's see what happened to the immediate treatment group in Year 2. Figure 1 shows the results for all teachers who completed their data collection on their students' proportional reasoning in both years (37 immediate treatment and 30 delayed treatment teachers). As described in our previous technical reports, on this assessment of proportional reasoning the M_1 items covered the mathematics typically covered in seventh-grade mathematics and on the Texas state test, and the M_2 items—on which we saw the greatest gain difference between groups—covered the mathematics that goes beyond the typical and prepares students for algebra and high school science.

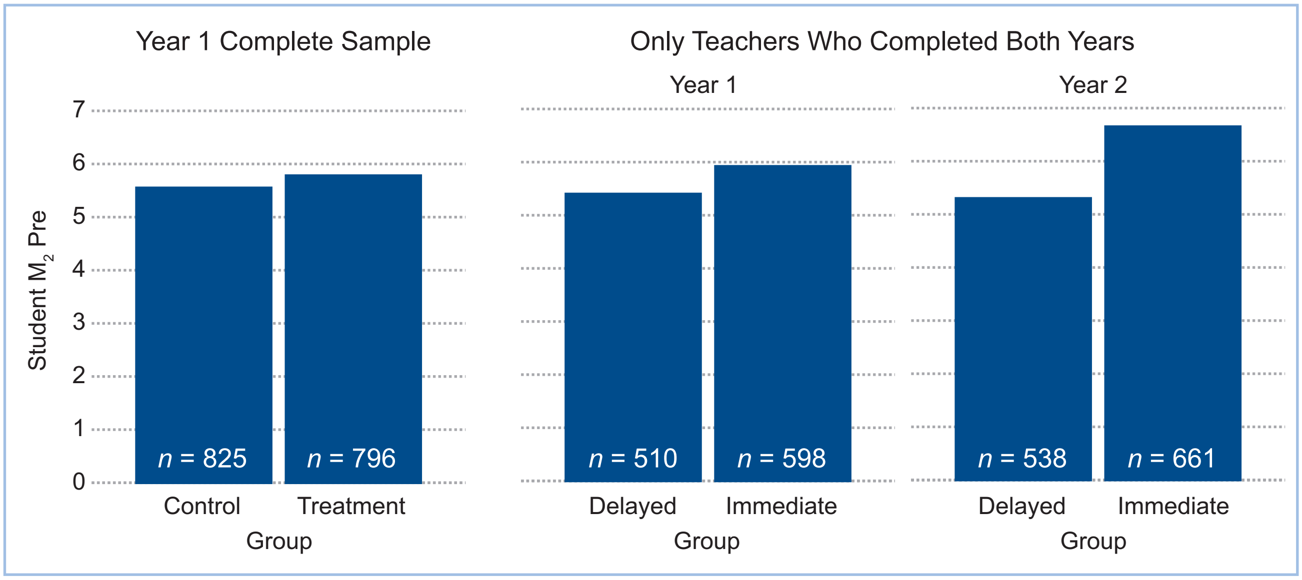
Figure 1. Pretest to posttest gains on the 30-item seventh-grade assessment across all classrooms of teachers who completed data collection both years.



Comparing the gains for the delayed treatment group's students in Years 1 and 2 reveals the quasi-experiment results. (When we reported on this quasi-experiment by itself, we only included the 30 delayed treatment teachers who completed both years; this chart, however, shows students of *all* teachers in Year 1.) The gain for the immediate treatment group's students in Year 2 was a bit lower than in Year 1; the gains are not statistically different. Overall, this was an encouraging finding—once again, the teachers were successful at teaching with SimCalc. But the speculators who believed the gains for these teachers would be greater in the second year were surprised this did not play out in the findings.

When we engaged in some detective work, we found evidence of possible selection bias. In Year 1 there were no baseline differences between groups, but in Year 2 the immediate treatment group had higher pretest scores (Figure 2). This difference was statistically significant. Although we have only weak evidence from interviews to support our conjecture about a self-selection bias, we suspect that teachers who implemented SimCalc in the first year purposely selected higher achieving classes in their second year to participate in the data collection (i.e., rather than follow our random selection of a target class period, as they appeared to in Year 1).

Figure 2. M_2 pretest gains across groups across years. The immediate treatment group in Year 2 had significantly higher pretest scores. This difference was not significant in Year 1.



In light of the possibility of a self-selectivity bias and the fact that there was apparently no particularly interesting shift in implementation success from the first to second year, we decided to focus on the more sound quasi-experimental comparison between years in the delayed treatment group.

The Past Does Not Predict the Future

A further question about the immediate treatment classrooms was whether teachers consistently achieved higher or lower gains among their students across years. The quick answer is no. An interesting and surprising finding was that although pretest scores were significantly correlated from one year to the next, there was no correlation in gains from year to year within the immediate treatment group (Figure 3).

It is not clear why this would be the case, but we speculate that it is related to two factors: (1) the majority of the variation in student gains was at the student level (71.2%) as opposed to the classroom level (28.8%), and (2) the range of classroom mean difference scores was small and highly overlapping. As Figure 4 shows, variation in gain within classrooms was considerable, but variation in the medians across classrooms was small. The wide variation in student-level growth and restriction in range of mean classroom-level growth might make it challenging to find meaningful correlative relationships at the classroom level. The restriction of range in classroom-level means was even greater in Year 2 (Figure 3).

Figure 3. Classroom-level gains by year for the 37 immediate treatment group teachers who completed both years.

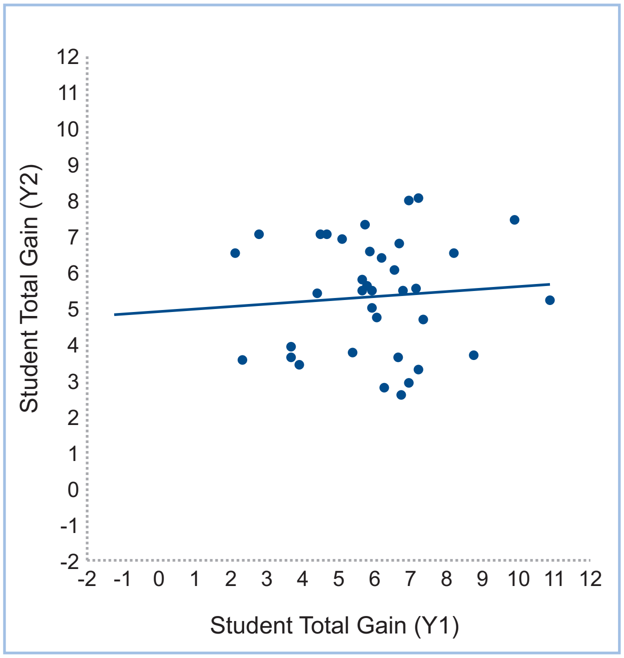
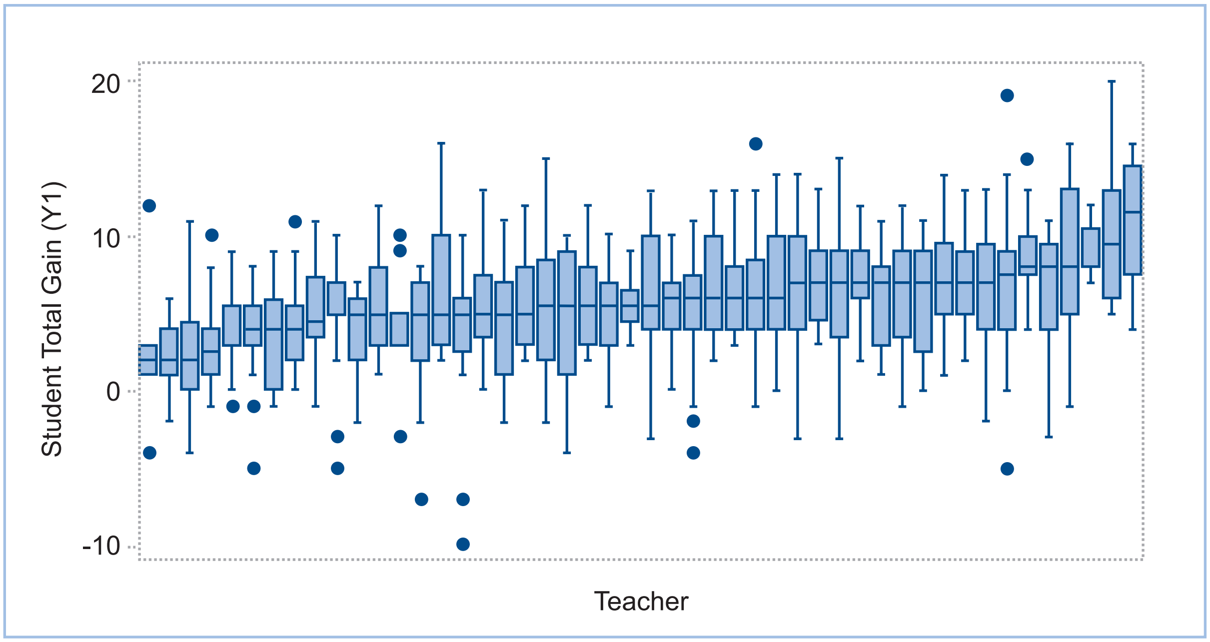


Figure 4. Student pretest score gains by classroom in seventh grade, Year 1.



Puzzling Socioeconomic Correlates

Readers may recall from our *American Educational Research Journal* paper (Roschelle et al., 2010, Appendix A) that the main effects were robust across all demographic groups. However, there was a somewhat puzzling finding that we still have not been able to decipher.

Table 1 shows the correlations between school demographic characteristics and student achievement. In both seventh grade years and the eighth-grade study, one robust (but unfortunate) finding was that socioeconomic factors are highly correlated with pretest scores. In Year 1 of the seventh-grade study and the eighth-grade study, student *gains* were not correlated with demographic factors. This suggests equitability in learning. However, in one group–the delayed treatment group in seventh grade Year 2–we see that gains were correlated with these demographic factors.

Why, and why only in one case? This is the puzzle.

Table 1. Correlations between school demographic characteristics and student achievement at the classroom-level

% of Students in School	Seventh Grade Year 1			Seventh Grade Year 2			Eighth Grade		
		Gain			Gain			Gain	
	Pretest *N*=95	Del. *N*=47	Immed. *N*=48	Pretest *N*=67	Del. *N*=30	Immed. *N*=37	Pretest *N*=56	Ctl *N*=23	Tx *N*=33
Qualify for free lunch	-.42***	.21	.05	-.52***	-.48**	-.11	-.45***	.08	-.14
Caucasian	.24*	-.18	-.04	.47***	.55**	-.13	.26	.16	.17
Hispanic	-.28**	.18	-.01	-.51***	-.56**	.05	-.26*	-.12	-.21

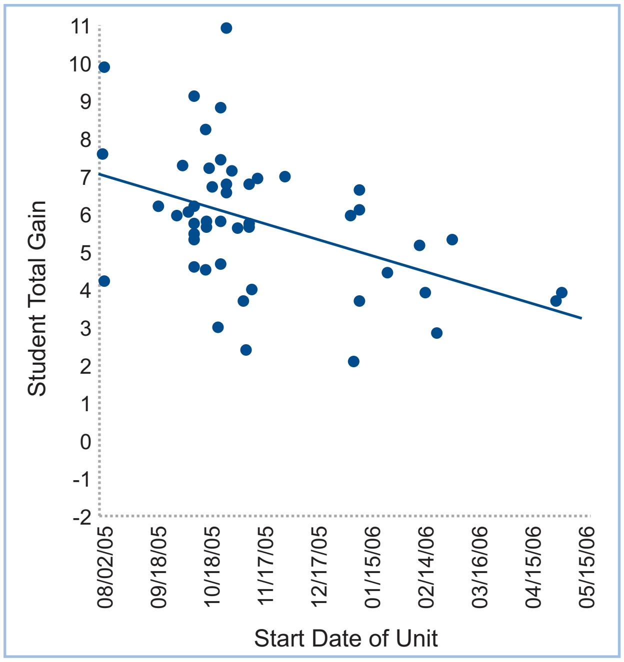
* $p < .05$; ** $p < .01$; *** $p < .0001$

All in the Timing

Another strange finding related to when a teacher started teaching with SimCalc. We found a significant result but only in one particular experimental group, the seventh-grade study Year 1 immediate treatment group. In this group, the earlier in the year the teacher began the unit, the more students tended to learn (Figure 5). The correlation was $r(48) = -.45$, $p < .01$. This finding was not replicated with significance anywhere else in the studies, although there was a similar nonsignificant trend in the eighth-grade study treatment group.

This could be due to a number of factors, one or many acting at the same time. One possibility is that students simply know less (and thus have higher potential gains) at the beginning of the year. In fact, a nonsignificant but positive correlation existed between start date and pretest score—a classroom's pretest score tended to be higher the later it was administered during the school

Figure 5. Start date of the unit and student learning gains, seventh-grade Year 1 study, immediate treatment group.



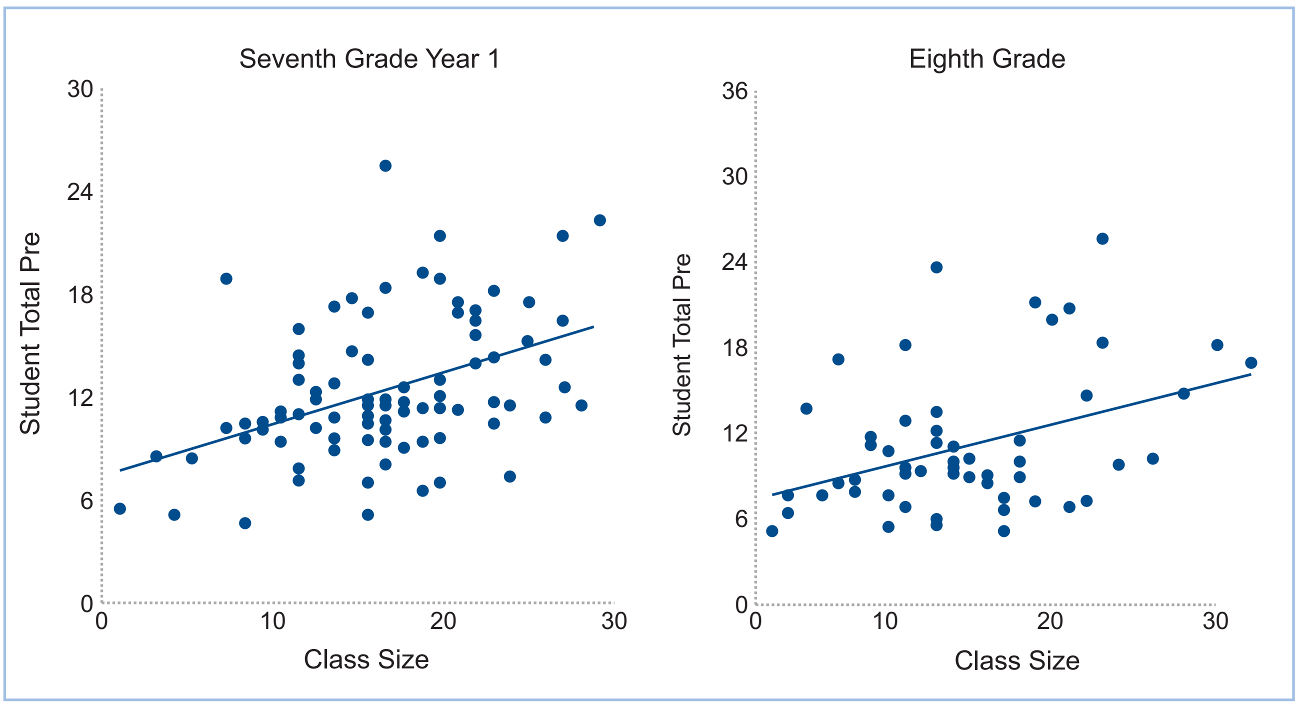
year. Another possibility is that teachers recall more of their summer training in the fall semester. Yet another possibility is that the unit occurs before the students get bored, disruptive, or otherwise inattentive. Finally, it may be that the more enthusiastic teachers that do the unit earlier in the year.

Class Size: Bigger Is Better?

Another finding on the cutting room floor is the correlation between class size and pretest scores. We found in both studies that class size was correlated with student pretest scores (Figure 6). This was significant in both cases: r(95) = .43, $p < .0001$ in seventh grade, and r(56) = .39, $p < .01$ in eighth grade. Furthermore, although gains were not correlated with class size in seventh grade, in the eighth-grade study treatment group, the larger the class the higher the gains: r(33) = .43, $p < .05$. This correlation was not significant in the eighth-grade control group.

Although this finding is counterintuitive and we do not know definitively why it is the case, the way it was discovered might offer some clues. In our analysis of the SimCalc workbooks, we had three master middle school math teachers review each classroom set of workbooks. They observed informally that the smaller classroom sets were more difficult to get through–the student work had more mistakes and misconceptions, more language issues, and less legible writing. Their observations led to this analysis, which revealed that the smaller classes had lower pretest scores. We conjecture from this that the smaller classes might be “resource” (special education) classes of some type.

Figure 6. Correlations between class size and pretest.



Rate of Change of Attitudes about Math of Change: k = 0

Curious at the outset about whether SimCalc could actually change how much kids like math, we asked students three questions at the end of the pretest and posttest (Exhibit 1).

We focus here on findings from the eighth-grade study (because we collected this data only in Year 2 of the seventh-grade study and analysis of that study is complicated). These questions cohere satisfactorily into a single factor with alpha .86 (with the 608 eighth-grade students who completed the survey). We call this factor LikeMath and use it in the following analyses.

Figure 7 shows that there were no changes. In fact, HLM analyses showed that note of these trends were significant. The treatment group did not start out liking math significantly more than the control group, and the decline in liking math is not significant.

Overall, it is not surprising that a 3-week replacement unit is not enough to substantially shift student attitudes about mathematics. But even if it did, this way of measuring the shift may not really get at what is important. It is interesting to look at the distribution of attitude shifts. Most students' attitudes stayed relatively consistent, but some shifted way up or way down (Figure 8).

Also, the variation in attitudes between classrooms is remarkable (Figure 9). In some classrooms, almost all students did not like math at all, whereas in others students had an overall positive attitude toward math. What might be going on in these different classrooms, and how might it affect students? Figure 10 gives us some clue. At the beginning of the year, we asked teachers to rate each student in their classroom as low, medium, or high achieving. We did not give them any criteria. Teachers' ratings were closely related to both student test scores and their gains and—now we see—also students' attitudes toward mathematics.

Not surprisingly, the higher students scored on the pretest, the more they liked math. The effect was not large, but it was significant: $\beta = .02$, $z = 3.02$, $p < .01$. Similarly, a small but significant relationship existed between how much students in the treatment group liked math and their M_2 gain scores: $\beta = .05$, $z = 2.06$, $p < .05$. The effect was similar in the control group.

Although these findings do not allow us to make any particular conclusions about student attitudes, they point to the need to explore and understand students' experiences in the classroom as they engage in these learning activities.

Exhibit 1. Questions probing whether students' attitudes about mathematics changes with exposure to SimCalc.

The following questions ask your opinion about math and your math class:

a. How much do you like math? Mark [X] one box.

Not at all- 1 2 3 4 5 6 7 - Love it!

☐ ☐ ☐ ☐ ☐ ☐ ☐

b. Think about the unit that you most recently completed in your math class. Think about the activities that you did and the math that you learned. How much did you enjoy your math class during this unit? Mark [X] one box.

Not at all- 1 2 3 4 5 6 7 - Love it!

☐ ☐ ☐ ☐ ☐ ☐ ☐

c. Again, think about the unit that you most recently completed in your math class. If math class were always like this, how excited would you be about taking math classes in the future? Mark [X] one box.

Not at all- 1 2 3 4 5 6 7 - Very excited!

☐ ☐ ☐ ☐ ☐ ☐ ☐

Figure 7. LikeMath pre- and post-unit across groups.

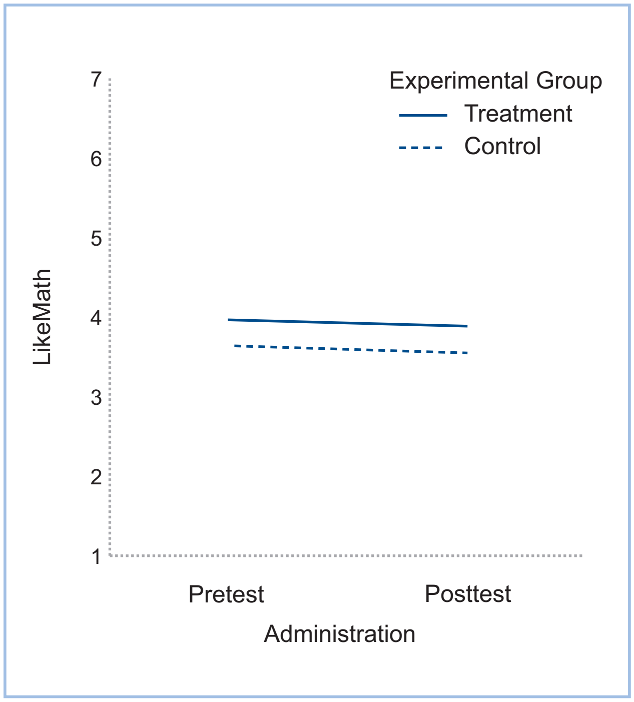


Figure 9. Mean pretest LikeMath by classroom.

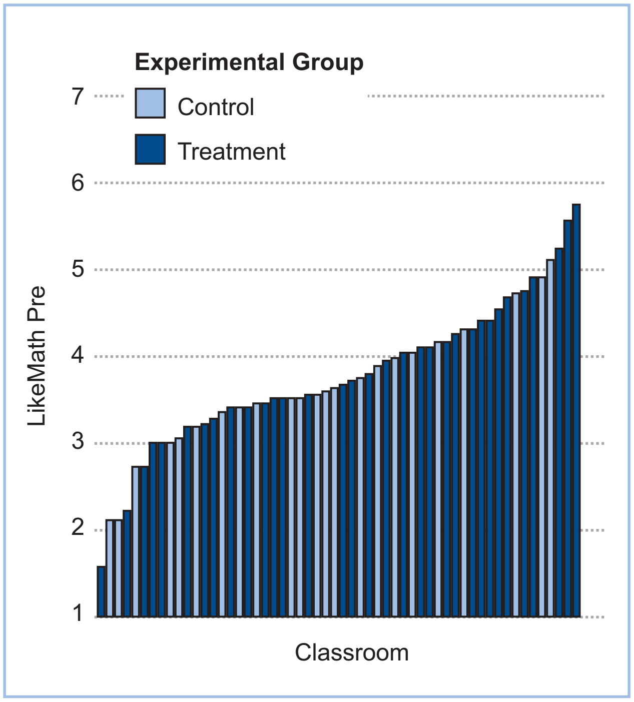


Figure 8. Distribution of the gain score for LikeMath from pretest to posttest.

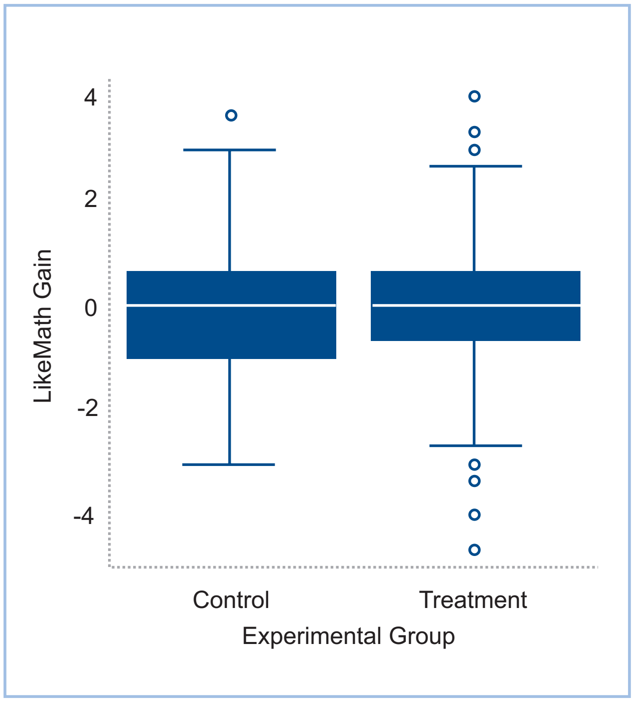
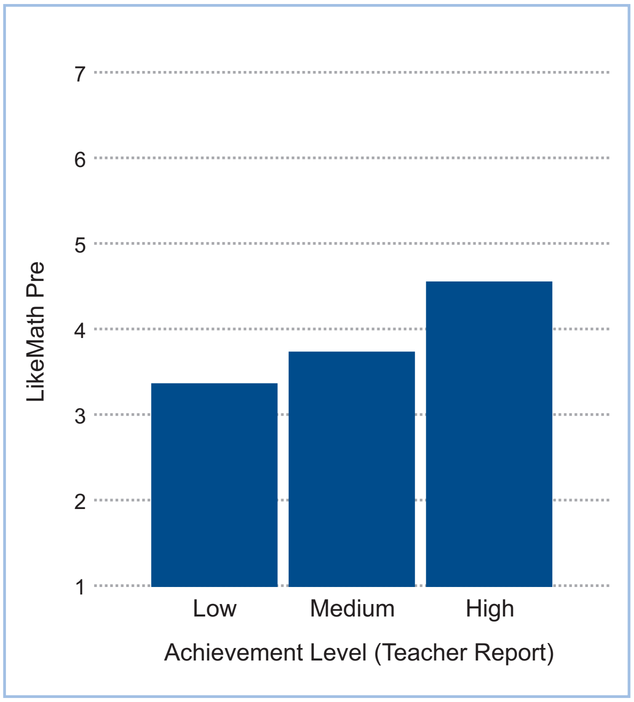


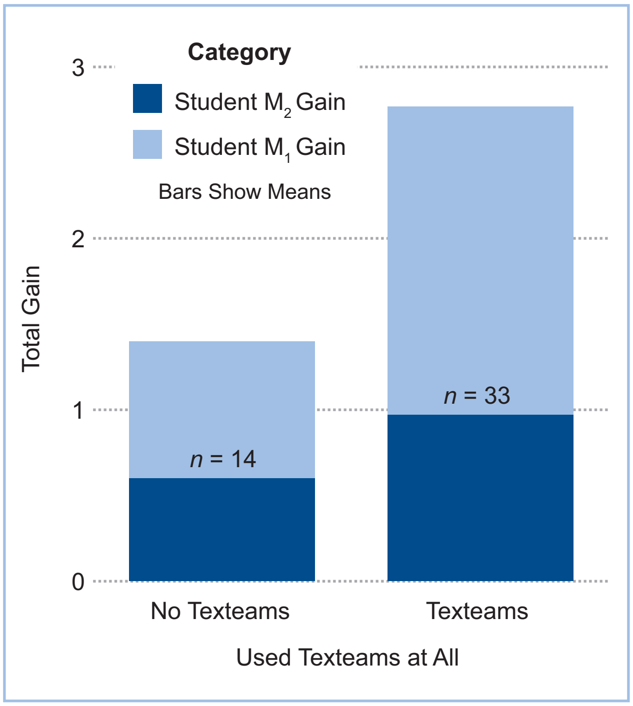
Figure 10. LikeMath by teacher rating of student achievement level.



Students Learn M_2 Concepts with Our Control Intervention, TEXTEAMS!

In the seventh-grade study, teachers in both the treatment and control groups received the TEXTEAMS training and materials to take back to their classrooms. TEXTEAMS was offered as a high-quality professional development experience to expose teachers to proportionality as a function relating an input and output by a multiplicative constant, i.e., y = kx. Within the control group, we conducted a quasi-experimental comparison to see whether the students of teachers who actually used the TEXTEAMS materials in their classrooms learned more of our target mathematics than teachers who did not use them. In fact, that was the case (Figure 11), and the difference was statistically significant. We do not know why teachers decided to use the TEXTEAMS materials or not. Clearly, there is a self-selection bias, so these findings should be interpreted with caution. Also note that the magnitude of gains from using TEXTEAMS is about half that from using SimCalc.

Figure 11. Quasi-experimental comparison of gains within the seventh-grade Year 1 control group of classrooms of teachers who did and did not use the TEXTEAMS materials in their classrooms. Note that even using the TEXTEAMS materials, the group did not approach the overall gain of 5.8 in the treatment group.



Girls Get an A for Effort

Gender inequity is a serious concern in mathematics, and some of our data did show inequity.

- *Baseline math knowledge.* On the seventh-grade Year 1 pretest, girls scored about 1 out of 30 points lower than boys (β = -.90 , z = -4.1, $p < .0001$), and on the eighth-grade pretest, girls scored about 2 out of 36 points lower than boys (β = -2.0, z = -4.8, $p < .0001$).
- *Learning gains.* Although girls in the treatment group scored higher than those in the control group in both studies, some gender difference was evident within the treatment group in the eighth-grade study: a marginally significant trend that girls gained about half a point less on M2 than boys (β = -.52, z = -1.9, $p = .06$). The overall mean gain in the treatment group was 4.8.

Girls did come out on par with or better than boys in a few areas.

- *Learning gains.* In the seventh-grade Year 1 study, treatment group girls gained much the same as boys, with a nonsignificant trend of edging toward closing the gap (β = .13, z = .55, $p = .58$).
- *Workbook completion.* Also in the seventh-grade Year 1 study, in which we looked at workbook completion, overall girls came out 3.6% ahead of boys in their workbook completion (β = 3.6, z = 4.25, $p < .0001$). Because workbook completion was related to learning gains, this extra effort may have something to do with their approaching closing the gap.
- *Liking math.* Perhaps most surprising was that girls liked math at baseline just a bit more than boys. In the eighth-grade study, girls were about a third of a point higher on the 7-point LikeMath scale (β = .30, z = 2.33, $p < .05$).

What Teachers Don't Know They Don't Know

In their 1999 research article, Kruger and Dunning[1] described how incompetence can mask the ability to recognize incompetence.

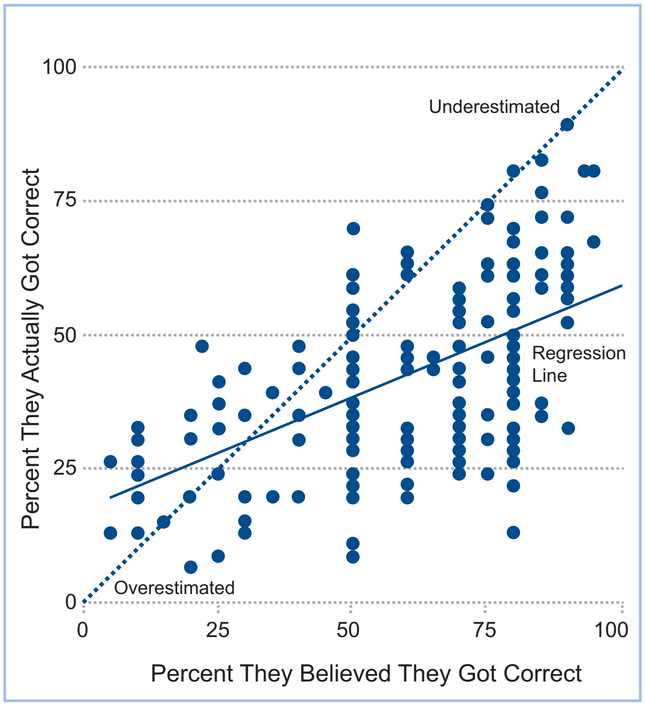
> People tend to hold overly favorable views of their abilities in many social and intellectual domains. The authors suggest that this overestimation occurs, in part, because people who are unskilled in these domains suffer a dual burden: Not only do these people reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realize it. Across 4 studies, the authors found that participants scoring in the bottom quartile on tests of humor, grammar, and logic grossly overestimated their test performance and ability. Although their test scores put them in the 12th percentile, they estimated themselves to be in the 62nd. Several analyses linked this miscalibration to deficits in metacognitive skill, or the capacity to distinguish accuracy from error. Paradoxically, improving the skills of participants, and thus increasing their metacognitive competence, helped them recognize the limitations of their abilities. (p. 1121)

This effect, called the Dunning-Kruger effect, can be a barrier to professional performance and growth because individuals do not understand the boundaries of their own knowledge.

We looked to see whether teachers in our study demonstrated the Dunning-Kruger effect with their mathematical knowledge. We conducted a secondary analysis on the field-test data for the seventh-grade teacher mathematical knowledge for teaching assessment. The original purpose of these data was to inform the development and refinement of the teacher assessment instrument. The data came from a mass mailing to 1,000 middle school math teachers randomly sampled throughout the United States; 179 teachers responded. At the end of the assessment, teachers were asked, "What percentage of the math items do you believe you got correct?"

Overall, there did appear to be a Dunning-Kruger effect: The mean percentage believed to be correct was 61.4 (SD = 23.8) and the mean percentage actually correct was 42.9 (SD = 17.6), a statistically significant difference ($t(165) = 33.1$, $p < .0001$). Although the correlation was high ($r(166) = .56$, $p < .0001$), Figure 12 shows an overall tendency for the teachers to overestimate their score. Furthermore, we found that the higher the actual score, the more accurate teachers were in estimating it. We calculated the absolute value of the discrepancy between the believed score and actual score and found that the higher the actual score, the lower the discrepancy ($r(166) = -.34$, $p < .0001$).

Figure 12. Comparison of believed and actual percentage correct in the teacher assessment field test.



1 J. Kruger & D. Dunning. (1999). "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessment," *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

These findings are consistent with those of Kruger and Dunning and suggest that there may be a metacognitive component of mathematical knowledge. Teachers in general overestimated their score on the assessment, but the higher the teachers' knowledge, the more accurate they were in their estimations. As in the original studies, teacher training may help teachers understand what they do not understand and seek support.

Conclusion

In reporting on the Scaling Up SimCalc project, we have often focused on "scale" and "robustness" as main messages. In general, we found considerable concurrence in the major findings across a series of replications of the main comparison between teachers in the treatment and control groups. However, it is important to remember that not every planned comparison works: Some project team members expected a second year of implementation to yield higher scores and it did not. It is also important to remember that p values represent actual uncertainty, and sometimes a comparison that is statistically significant in one study will not be so in the next study. We found a nonreplication with regard to the relationship between socioeconomic status and student learning gains, for example. Finally, not every statistically significant comparison has a meaningful interpretation. We do not know why classes with more students had higher pretest scores, for instance. We hope by reporting some of these bad hunches, nonreplicable findings, and head-scratchers we can support other researchers in their own efforts to design meaningful and important experiments.

Appendix A. Real Publications

Key Articles

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J., & Gallagher, L. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal, 47*(4), 833-878.

Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (2010). Investigating links from teacher knowledge, to classroom practice, to student learning in the instruction system of the middle school mathematics classroom. *Cognition and Instruction, 28*(3), 317-359.

Tatar, D., Roschelle, J., Knudsen, J., Shechtman, N., Kaput, J., & Hopkins, B. (2008). Scaling up innovative technology-based mathematics. *Journal of the Learning Sciences, 17*(2), 248–286.

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*(2), 149–170.

Hegedus, S., & Lesh, R. (Eds.) (2008). Democratizing access to mathematics through technology: Issues of design and implementation. *Educational Studies in Mathematics Special Issue, 68*(2), 81–93.

Book Chapters

Hegedus, S., Kaput, J., & Lesh, R. (2007). Technology becoming infrastructural in mathematics education. In R. Lesh, E. Hamilton, & J. Kaput (Eds.), *Foundations for the future in mathematics and science* (pp. 173–192). Mahwah, NJ: Lawrence Erlbaum.

Roschelle, J., Tatar, D., & Kaput, J. (2009). Getting to scale with innovations that deeply restructure how students come to know mathematics. In A. Kelly, R. Lesh, & J. Baek (Eds.), *Handbook of design research methods in mathematics, science and technology education.* Mahwah, NJ: Lawrence Erlbaum.

Doctoral Dissertations

Dunn, M. B. (2009). *Investigating variation in teaching with technology-rich interventions: What matters in training and teaching at scale?* Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.

Pierson, J. (2008). *The relationship between patterns of classroom discourse and mathematics learning.* Unpublished doctoral dissertation, University of Texas at Austin.

Stroter, A. (2008). *The effects of teacher-student racial and ethnic congruence on student math learning.* Unpublished doctoral dissertation, Virginia Tech.

Manuscripts Under Review and In Preparation

Empson, S. B., Greenstein, S., & Maldonado, L. (2010). *Scaling up innovative mathematics in the middle grades: case studies of "good enough" enactments.* Manuscript submitted for publication.

Hegedus, S. J., & Moreno-Armella, L. (2010). *Enhancing instrumental genesis with dynamic representations.* Manuscript submitted for publication.

Hegedus, S. J., Roschelle, J., & Moreno-Armella, L. (2010). *The mathematics of change and variation in middle school: Theoretical perspectives and findings from the SimCalc Research Program.* Submitted to the Journal of Research in Mathematics Education. Under revision.

Pierson, J. (2010). Gatekeepers, exiles, and citizens: *The effect of identity on mathematics learning.* Manuscript submitted for publication.

Stroter, A., & Tatar, D. (2010). *An evaluation of recruitment strategies, outcomes, and implications for a randomized-controlled experimental design with teachers.* Manuscript in preparation.

Technical Reports

Estrella, G., Shechtman, N., & Roschelle, J. (2010). *Designing the logistics for large-scale randomized controlled trials: Six strategies for implementation at scale.* Menlo Park, CA: SRI International.

Fishman, B., Penuel, W., Hegedus, S., Moniz, R., Dalton, S., Brookstein, A., Beaton, D., Tatar, D., Dickey, M., & Roschelle, J. (2009). *What happens when the research ends? Factors related to the sustainability of a research-based innovation.*

Knudsen, J. (2010). *Design and development of curriculum units and professional development for the scaling up simcalc project.* Manuscript in preparation. Menlo Park, CA: SRI International.

Roschelle, J., Tatar, D., Shechtman, N., Hegedus, S., Hopkins, B., Knudsen, J., & Stroter, A. (2007). *Scaling Up SimCalc Project: Can a technology enhanced curriculum improve student learning of important mathematics?* Menlo Park, CA: SRI International.

Roschelle, J., Tatar, D., Shechtman, N., Hegedus, S., Hopkins, B., Knudsen, J., & Dunn, M. (2008). *Extending the SimCalc Approach to grade 8 mathematic*s. Menlo Park, CA: SRI International.

Shechtman, N., Haertel, G., Roschelle, J., Knudsen, K., & Singleton, C. (2010). *Design and development of the student and teacher mathematical assessments.* Menlo Park, CA: SRI International.

Shechtman, N., Carriere, S., & Roschelle, J. (2010). *The data outtakes reel: archive of unreported, unreportable, and irreproducible findings.* You're reading it right now. Consider the paradoxical consequences of a publication citing itself…

Tatar, D. & Stroter, A. (2009). *Recruitment strategies, outcomes, and implications for a randomized-controlled experiment with teachers.* Menlo Park, CA: SRI International.

Vahey, P., Lara-Meloy, T., & Carriere, S. (2010). *Middle School Students' Mathematics Learning: An Analysis of a Student Population in Texas.* Manuscript in preparation.

Appendix B. Documentation of All Models

In this Appendix, we document each of the models run on the total gain scores and the M_2 gain scores across the three major studies.

Note that there may be minute inconsistencies with statistics reported in other publications. That is due to decisions about inclusion of missing data and does not have a substantial impact on the findings. Sample size discrepancies are annotated.

Table 2. Two-level MLM Models Run in Each Study for Each Factor Predicting Total Gains

Note: Full model is $Y_{ij} = \gamma_{00} + \gamma_{01} T_j + \gamma_{02} X_{ij} + \gamma_{03} T_j * X_{ij} + r_{ij} + u_j$, where X_{ij} may be a level 1 or level 2 covariate.

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
	(Main effect is experimental condition)		(Main effect is year 1 vs. year 2)		(Main effect is experimental condition)	
	n = 1,621		*n* = 1,048		*n* = 825	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
School is in Region 1						
Main Effect	3.89***	0.447	3.22***	0.258		
Region 1	0.38	0.739	1.14	0.885		
Region 1 Interaction	-1.04	1.060	-1.86***	0.543		
Intercept	2.00***	0.310	2.25***	0.368		
Level 2 Variance	2.26		2.08			
Residual Variance	13.20		13.35			
% School Hispanic						
Main Effect	3.70***	0.410	2.78***	0.226	4.00***	0.821
Percent School Hispanic	0.22	0.844	0.61	0.972	-1.50	2.980
Percent School Hispanic Interaction	-0.15	1.202	-3.05***	0.650	-2.32	3.736
Intercept	2.08***	0.286	2.48***	0.330	2.81***	0.627
Level 2 Variance	2.30		1.95		5.09	
Residual Variance	13.20		13.22		21.96	
% School Caucasian						
Main Effect	3.70***	0.409	2.78***	0.226	3.91***	0.819
Percent School Hispanic	-0.36	0.913	-0.79	1.037	1.56	2.367
Percent School Hispanic Interaction	0.30	1.297	3.37***	0.697	1.53	3.344
Intercept	2.08***	0.285	2.48***	0.327	2.84***	0.622
Level 2 Variance	2.29		1.95		5.19	
Residual Variance	13.20		13.20		21.97	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
School SES						
Main Effect	3.67***	0.411	2.75***	0.227	3.86***	0.840
School SES	0.70	1.044	0.76	1.215	0.83	2.761
School SES Interaction	0.18	1.440	-3.39***	0.813	-3.64	3.699
Intercept	2.10***	0.288	2.50***	0.328	2.84***	0.641
Level 2 Variance	2.30		1.97		5.47	
Residual Variance	13.19		13.28		21.94	
Student Hispanic[1]						
Main Effect	3.81***	0.469	3.89***	0.338	4.48***	0.895
Student Hispanic	-0.07	0.360	0.46	0.389	1.23	0.738
Student Hispanic Interaction	0.04	0.494	-1.98***	0.466	-2.08*	0.890
Intercept	2.14***	0.331	2.26***	0.375	2.53***	0.674
Level 2 Variance	2.12		1.88		5.34	
Residual Variance	13.45		13.39		21.88	
Student Female[2]						
Main Effect	3.64***	0.448	2.85***	0.319	4.21***	0.902
Student Female	-0.18	0.271	0.28	0.334	0.03	0.599
Student Female Interaction	0.31	0.386	-0.10	0.468	-0.80	0.736
Intercept	2.21***	0.314	2.35***	0.373	2.86***	0.697
Level 2 Variance	2.09		1.97		5.21	
Residual Variance	13.53		13.53		22.25	
Teacher Female						
Main Effect	1.70**	0.705	2.14***	0.506	3.69***	0.822
Teacher Female	-1.51***	0.501	-1.49*	0.753	-1.97	2.858
Teacher Female Interaction	2.57***	0.756	0.84	0.567	3.03	3.309
Intercept	3.27***	0.483	3.64***	0.657	2.97***	0.628
Level 2 Variance	2.06		1.86		5.02	
Residual Variance	13.14		13.47		22.03	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
Achievement Level (Teacher Report)[3]						
Main Effect	3.83***	0.478	2.88***	0.352	3.62***	0.924
Low Achieving (nominated)	-0.23	0.337	-0.32	0.410	-1.70*	0.754
High Achieving (nominated)	0.29	0.356	-0.16	0.422	0.82	0.689
Low Achievement Interaction	-0.63	0.495	-0.69	0.573	-0.08	0.930
High Achievement Interaction	0.47	0.518	0.67	0.585	0.44	0.885
Intercept	2.11***	0.325	2.65***	0.378	3.20***	0.705
Residual Variance	13.50		13.54		21.41	
Teacher has Masters[4]						
Main Effect	3.64***	0.441	2.52***	0.252	3.87***	0.850
Teacher has Masters	-1.14	0.545	-0.93	0.663	-0.45	0.746
Teacher has Masters Interaction	0.48	0.778	1.55**	0.586	-2.23	1.794
Intercept	2.32***	0.308	2.67***	0.352	3.00***	0.656
Level 2 Variance	2.34		2.00		5.24	
Residual Variance	13.16		13.42		21.93	
Teacher Hispanic						
Main Effect	4.12***	0.437	3.13***	0.261	3.95***	0.837
Teacher Hispanic	1.04	0.543	0.88	0.738	1.31	1.181
Teacher Hispanic Interaction	-1.88	0.764	-1.40**	0.537	-1.64	1.426
Intercept	1.85***	0.305	2.29***	0.367	2.81***	0.632
Level 2 Variance	2.22		2.06		5.41	
Residual Variance	13.16		13.41		21.95	
Years Teaching Math[5]						
Main Effect	3.73	0.383	2.90***	0.228	3.83***	0.815
Years Teaching Math	0.01	0.032	0.08*	0.038	-0.03	0.040
Years Teaching Math Interaction	0.08	0.039	-0.04	0.037	0.10	0.055
Intercept	2.04	0.269	2.29***	0.286	2.90***	0.619
Level 2 Variance	1.83		1.27		5.16	
Residual Variance	12.88		13.10		21.93	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
MKT Pretest Score						
Main Effect	3.67***	0.402	2.80***	0.228	3.83***	0.835
MKT Pretest	0.05	0.058	-0.04	0.067	0.05	0.059
MKT Pretest Interaction	0.03	0.071	0.10	0.058	-0.06	0.091
Intercept	2.06***	0.281	2.50***	0.334	2.87***	0.631
Level 2 Variance	2.20		2.06		5.45	
Residual Variance	13.18		13.46		21.97	
Class Size						
Main Effect	3.72***	0.411	2.67***	0.242	3.54***	0.810
Class Size	0.02	0.044	-0.05	0.031	0.06	0.066
Class Size Interaction	-0.02	0.063	0.04	0.025	0.08	0.085
Intercept	2.05***	0.288	2.59***	0.333	2.98***	0.617
Level 2 Variance	2.27		1.99		4.91	
Residual Variance	13.20		13.47		21.85	
Days in Class						
Main Effect	2.35***	0.487			1.90	1.067
Days in Unit	0.54***	0.123			0.05	0.145
Days in Class	-0.66***	0.129			-0.15	0.146
Days in Unit Interaction	-0.41***	0.140			0.30	0.190
Days in Class Interaction	0.56***	0.142			-0.19	0.184
Intercept	3.18***	0.369			3.68***	0.859
Level 2 Variance	2.24				3.81	
Residual Variance	12.97				22.01	
Days in Comp Lab						
Main Effect	2.23**	0.795			1.92	1.185
Days in Unit	-0.04	0.056			-0.09	0.072
Days in Comp Lab	0.54*	0.280			0.13	0.194
Days in Unit Interaction	0.08	0.064			0.15	0.118
Days in Computer Lab Interaction	-0.46	0.286			0.25	0.222
Intercept	3.30***	0.722			3.65***	1.040
Level 2 Variance	2.33				3.04	
Residual Variance	13.15				22.14	

Model	Seventh-Grade Year 1 Experiment Value	SE	Seventh-Grade Quasi-Experiment Value	SE	Eighth-Grade Experiment Value	SE
Days in Unit						
Main Effect	3.72***	0.416	2.76***	0.240	3.79	0.824
Days in Unit	-0.03	0.055	0.05	0.049	-0.07	0.080
Days in Unit Interaction	0.06	0.064	-0.09	0.098	0.18	0.133
Intercept	2.03***	0.295	2.51***	0.338	3.01	0.631
Level 2 Variance	2.27		2.14		4.98	
Residual Variance	13.20		13.48		22.01	
Days with Individual Student Work						
Main Effect	3.72***	0.409	2.86***	0.256	3.93***	0.687
Days with Individual Student Work	-0.15*	0.073	-0.15*	0.063	0.08	0.077
Days in Unit	0.07	0.071	0.14*	0.063	-0.13	0.089
Days in Unit Interaction	-0.04	0.080	-0.23	0.121	0.16	0.128
Days in Indiv. Stud. Work Interaction	0.17*	0.078	0.19*	0.089	0.36***	0.116
Intercept	2.02***	0.290	2.36***	0.337	3.06***	0.531
Level 2 Variance	2.16		2.03		2.99	
Residual Variance	13.19		13.44		21.79	
Days with Student Pair Work						
Main Effect	3.49***	0.431	2.55***	0.252	3.47***	0.911
Days with Student Pair Work	0.11	0.064	0.21**	0.068	0.12	0.116
Days in Unit	-0.09	0.064	-0.06	0.061	-0.08	0.086
Days in Unit Interaction	0.12	0.072	0.05	0.108	0.15	0.144
Days in Pair Work Interaction	-0.10	0.068	-0.21**	0.084	0.07	0.141
Intercept	2.24***	0.316	2.72***	0.337	3.12***	0.696
Level 2 Variance	2.20		2.01		6.25	
Residual Variance	13.20		13.39		21.71	
Days with Student Small Group Work						
Main Effect	3.60***	0.430	2.55***	0.252	4.11***	0.873
Days with Student Small Group Work	0.07	0.075	0.21**	0.068	0.01	0.096
Days in Unit	-0.04	0.056	-0.06	0.061	-0.07	0.086
Days in Unit Interaction	0.07	0.065	0.05	0.108	0.19	0.138
Days in Small Group Work Interaction	-0.06	0.077	-0.21**	0.084	0.27	0.228
Intercept	2.13***	0.313	2.72***	0.337	2.00***	0.649
Level 2 Variance	2.29		2.01		5.27	
Residual Variance	13.19		13.39		21.97	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
Days with Teacher Demonstration						
Main Effect	3.74***	0.423	2.73***	0.245	3.79***	0.850
Days with Teacher Demonstration	0.02	0.077	0.05	0.076	0.04	0.159
Days in Unit	-0.04	0.070	0.03	0.057	-0.08	0.086
Days in Unit Interaction	0.04	0.084	-0.07	0.111	0.21	0.151
Days in Tchr. Demonstration Interaction	0.04	0.093	-0.05	0.094	-0.07	0.185
Intercept	2.03***	0.301	2.54***	0.349	3.04***	0.652
Level 2 Variance	2.37		2.26		5.15	
Residual Variance	13.18		13.48		22.03	
Days with Whole Class Discussion[6]						
Main Effect	3.51***	0.489	2.81***	0.270	3.80***	0.819
Days with Whole Class Discussion	-0.09	0.080	-0.12	0.093	-0.13	0.148
Days in Unit	0.01	0.064	0.08	0.056	-0.06	0.080
Days in Unit Interaction	-0.16	0.100	-0.13	0.106	0.15	0.135
Days in Whole Class Discussion Interaction	0.33**	0.120	0.19	0.117	0.23	0.171
Intercept	1.83***	0.352	2.41***	0.353	2.95***	0.628
Level 2 Variance	2.44		2.23		4.84	
Residual Variance	13.16		13.46		22.03	
Days with Whole Class Lecture						
Main Effect	3.93	0.421	3.06***	0.262	3.86***	0.807
Days with Whole Class Lecture	-0.21***	0.069	-0.27***	0.068	0.06	0.136
Days in Unit	0.05	0.062	0.12	0.052	-0.08	0.081
Days in Unit Interaction	-0.01	0.071	-0.12	0.102	0.29*	0.136
Days in Whole Class Lecture Interaction	0.17*	0.081	0.25***	0.080	-0.31	0.159
Intercept	1.85***	0.300	2.22***	0.324	3.03***	0.619
Level 2 Variance	2.24		1.78		4.70	
Residual Variance	13.14		13.34		21.86	
Emphasis on Complex Goals						
Main Effect	3.41***	0.389	2.43***	0.257	3.71***	0.856
Complex Goals	0.81*	0.337	1.11***	0.326	-0.04	0.694
Complex Goals Interaction	0.34	0.529	-1.32**	0.497	0.72	0.985
Intercept	2.22***	0.272	2.92***	0.335	2.86***	0.658
Level 2 Variance	1.90		1.71		5.08	
Residual Variance	13.17		13.39		22.02	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
Emphasis on Simple Goals						
Main Effect	3.62***	0.419	2.51***	0.254	3.83***	0.871
Simple Goals	-0.06	0.332	-1.35**	0.482	0.33	0.928
Simple Goals Interaction	-0.49	0.447	1.25*	0.556	-1.24	1.051
Intercept	2.08***	0.294	2.81***	0.342	2.80***	0.672
Level 2 Variance	2.36		1.89		5.72	
Residual Variance	13.16		13.42		21.85	
Emphasis on M1 Topics						
Main Effect	3.69***	0.403	2.66***	0.247		
M1 Topics	0.36	0.336	0.55	0.324		
M1 Topics Interaction	-0.94*	0.471	-0.52	0.386		
Intercept	2.10***	0.282	2.66***	0.339		
Level 2 Variance	2.20		1.92			
Residual Variance	13.18		13.48			
Emphasis on M2 Topics						
Main Effect	3.37***	0.431	2.32***	0.338		
M2 Topics	1.15***	0.295	0.56*	0.286		
M2 Topics Interaction	-1.53***	0.444	-0.45	0.436		
Intercept	2.54***	0.303	2.99***	0.406		
Level 2 Variance	2.16		1.83			
Residual Variance	13.09		13.48			

Table 3. Two-level MLM Models Run in Each Study for Each Factor Predicting M_2 Gains

Note: Full model is $Y_{ij} = \gamma_{00} + \gamma_{01} T_j + \gamma_{02} X_{ij} + \gamma_{03} T_j * X_{ij} + r_{ij} + u_j$, where X_{ij} may be a level 1 or level 2 covariate.

	Seventh-Grade Year 1 Experiment		**Seventh-Grade Quasi-Experiment**		**Eighth-Grade Experiment**	
	(Main effect is experimental condition)		(Main effect is year 1 vs. year 2)		(Main effect is experimental condition)	
	n = 1,621		*n* = 1,048		*n* = 825	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
School is in Region 1						
Main Effect	3.48***	0.365	2.75***	0.194		
Region 1	0.27	0.606	0.59	0.753		
Region 1 Interaction	-0.78	0.870	-1.36***	0.407		
Intercept	1.28***	0.255	1.49***	0.310		
Level 2 Variance	1.55		1.59			
Residual Variance	7.94		7.51			
% School Hispanic						
Main Effect	3.34***	0.334	2.42***	0.169	3.38***	0.505
Percent School Hispanic	0.14	0.690	0.27	0.797	-1.17	1.845
Percent School Hispanic Interaction	-0.27	0.979	-2.86***	0.481	-0.56	2.302
Intercept	1.34***	0.234	1.60***	0.268	1.29***	0.388
Level 2 Variance	1.57		1.38		1.80	
Residual Variance	7.94		7.34		10.12	
% School Caucasian						
Main Effect	3.33***	0.334	2.42***	0.168	3.34***	0.502
Percent School Caucasian	-0.34	0.746	-0.58	0.852	1.36	1.459
Percent School Caucasian Interaction	0.52	1.057	3.345***	0.518	0.04	2.049
Intercept	1.34***	0.233	1.61***	0.268	1.31***	0.384
Level 2 Variance	1.57		1.40		1.82	
Residual Variance	7.94		7.30		10.11	
School SES						
Main Effect	3.32***	0.336	2.40***	0.169	3.30***	0.510
School SES	0.48	0.856	0.52	1.010	0.58	1.678
School SES Interaction	-0.20	1.176	-3.15***	0.607	-1.67	2.256
Intercept	1.35***	0.236	1.63***	0.272	1.32***	0.391
Level 2 Variance	1.58		1.44			
Residual Variance	7.93		7.40			

Model	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
	Value	SE	Value	SE	Value	SE
Student Hispanic [1]						
Main Effect	3.65***	0.376	3.40***	0.252	3.63***	0.556
Student Hispanic	-0.25	0.281	0.09	0.292	0.34	0.491
Student Hispanic Interaction	-0.37	0.385	-1.76	0.348	-1.01	0.592
Intercept	1.47***	0.266	1.56***	0.294	1.23**	0.420
Level 2 Variance	1.44		1.25		1.93	
Residual Variance	8.01		7.45		9.78	
Student Female [2]						
Main Effect	1.56***	0.547	2.57***	0.240	3.52***	0.562
Student Female	-1.14***	0.389	0.22	0.251	-0.02	0.401
Student Female Interaction	2.30***	0.586	-0.31	0.352	-0.55	0.493
Intercept	2.24***	0.375	1.53***	0.306	1.35***	0.436
Level 2 Variance	1.24		1.49		1.85	
Residual Variance	7.93		7.62		10.05	
Teacher Female						
Main Effect	1.56***	0.547	2.11***	0.380	3.18***	0.480
Teacher Female	-1.14***	0.389	-1.66**	0.592	-0.24	1.713
Teacher Female Interaction	2.30***	0.586	0.43	0.425	2.13	2.017
Intercept	2.24***	0.375	2.90***	0.516	1.36***	0.369
Level 2 Variance	1.24		1.19		1.53	
Residual Variance	7.93		7.58		10.16	
Achievement Level (Teacher Report) [3]						
Experimental Treatment	0.45	0.271	2.52	0.260	0.43	0.467
Low Achieving (nominated)	-0.36	0.257	-0.32	0.302	-0.90	0.508
High Achieving (nominated)	3.42***	0.383	0.30	0.311	3.21	0.584
Low Achievement Interaction	-0.91*	0.378	-0.88	0.422	-0.09	0.628
High Achievement Interaction	0.76*	0.396	0.76	0.431	0.21	0.599
Intercept	1.35***	0.261	1.64	0.302	1.49	0.448
Level 2 Variance	1.51		1.40		1.80	
Residual Variance	7.80		7.34		9.88	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
Teacher has Masters [4]						
Main Effect	3.34***	0.366	2.28***	0.189	-0.54	1.164
Teacher has Masters	-0.56	0.435	-0.32	0.528	-0.20	0.494
Teacher has Masters Interaction	0.34	0.621	0.90*	0.440	-0.20	0.494
Intercept	1.45***	0.256	1.69***	0.298	1.34***	0.407
Level 2 Variance	1.70		1.57			
Residual Variance	7.82		7.56			
Teacher Hispanic						
Main Effect	3.73***	0.353	2.64***	0.196	3.27***	0.524
Teacher Hispanic	1.02	0.433	0.62	0.602	-0.20	0.494
Teacher Hispanic Interaction	-1.79**	0.608	-0.81	0.403	-0.54	1.164
Intercept	1.12***	0.246	1.49***	0.307	1.40***	0.407
Level 2 Variance	1.50		1.56		1.82	
Residual Variance	7.91		7.56		10.13	
Years Teaching Math [5]						
Main Effect	3.39***	0.321	2.45***	0.173	0.01	0.026
Years Experience Teaching Math	-0.03	0.025	0.02	0.030	3.30***	0.497
Years Teaching Math Interaction	0.10***	0.032	0.03	0.028	0.03	0.037
Intercept	1.28***	0.226	1.50***	0.242	1.34***	0.379
Level 2 Variance	1.35		1.00		1.78	
Residual Variance	7.90		7.56		10.11	
MKT Pretest Score						
Main Effect	3.29***	0.325	2.45***	0.171	0.01	0.039
MKT Pretest	0.04	0.046	0.00	0.053	3.24***	0.516
MKT Pretest Interaction	0.07	0.056	0.07	0.043	-0.05	0.060
Intercept	1.32***	0.227	1.62***	0.282	1.34***	0.392
Level 2 Variance	1.48		1.57		1.96	
Residual Variance	7.89		7.57		10.10	
Class Size						
Main Effect	3.39***	0.333	2.40***	0.182	0.03	0.043
Class Size	0.03	0.035	-0.02	0.025	3.16***	0.491
Centered Class Size Interaction	-0.01	0.050	0.01	0.018	0.03	0.055
Intercept	1.23***	0.234	1.66***	0.282	1.39***	0.375
Level 2 Variance	1.54		1.55		1.66	
Residual Variance	7.94		7.59		10.13	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
Days in Class						
Main Effect	2.39***	0.396			2.27***	0.679
Days in Unit	0.42***	0.099			0.02	0.094
Days in Class	-0.50***	0.103			-0.07	0.094
Days in Unit Interaction	-0.38***	0.113			0.10	0.123
Days in Class Interaction	0.47***	0.114			-0.10	0.119
Intercept	2.19***	0.300			1.75***	0.549
Level 2 Variance	1.54				1.40	
Residual Variance	7.83				10.17	
Days in Comp Lab						
Main Effect	1.87***	0.644			1.92**	0.763
Days in Unit	-0.02	0.045			-0.05	0.047
Days in Comp Lab	0.56**	0.226			0.11	0.125
Days in Unit Interaction	0.04	0.052			0.12	0.143
Days in Computer Lab Interaction	-0.50*	0.230			0.03	0.076
Intercept	2.63***	0.584			1.97***	0.670
Level 2 Variance	1.61				1.18	
Residual Variance	7.90				10.17	
Days in Unit						
Main Effect	3.35***	0.341	2.50***	0.180	3.22***	0.515
Days in Unit	-0.01	0.044	0.03	0.038	-0.03	0.051
Days in Unit Interaction	0.03	0.051	0.04	0.074	0.04	0.085
Intercept	1.31***	0.242	1.63***	0.278	1.41***	0.396
Level 2 Variance	1.58		1.53		1.83	
Residual Variance	7.93		7.58		10.13	
Days with Individual Student Work						
Main Effect	3.34***	0.339	2.50***	0.193	3.29	0.433
Days with Individual Student Work	-0.07	0.059	-0.03	0.048	0.04	0.050
Days in Unit	0.03	0.057	0.05	0.048	-0.07	0.058
Days in Unit Interaction	-0.03	0.064	0.00	0.092	0.02	0.082
Days in Indiv. Stud. Work Interaction	0.09	0.064	0.05	0.067	0.23***	0.075
Intercept	1.31***	0.240	1.60***	0.282	1.45***	0.336
Level 2 Variance	1.55		1.53		1.09	
Residual Variance	7.94		7.59		10.03	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
Days with Student Pair Work						
Main Effect	3.19***	0.355	2.434***	0.190	3.09***	0.555
Days with Student Pair Work	0.08	0.052	0.12*	0.051	0.05	0.075
Days in Unit	-0.05	0.052	-0.04	0.046	-0.04	0.054
Days in Unit Interaction	0.06	0.058	0.14	0.082	0.02	0.090
Days in Pair Work Interaction	-0.07	0.055	-0.19	0.064	0.02	0.092
Intercept	1.46***	0.260	1.75***	0.276	1.45***	0.427
Level 2 Variance	1.56		1.44		2.16	
Residual Variance	7.93		7.54		10.06	
Days with Student Small Group Work						
Main Effect	3.26***	0.352	2.41***	0.188	3.37***	0.545
Days with Student Small Group Work	0.05	0.060	-0.07	0.055	0.05	0.063
Days in Unit	-0.02	0.045	0.05	0.039	-0.05	0.054
Days in Unit Interaction	0.03	0.052	0.07	0.082	0.05	0.087
Days in Small Group Work Interaction	-0.05	0.062	-0.01	0.079	0.04	0.147
Intercept	1.39***	0.255	1.69***	0.295	1.36***	0.407
Level 2 Variance	1.59		1.73		1.92	
Residual Variance	7.94		7.55		10.12	
Days with Teacher Demonstration						
Main Effect	3.37***	0.354	2.49***	0.183	3.21***	0.537
Days with Teacher Demonstration	0.07	0.062	0.10	0.058	-0.04	0.102
Days in Unit	0.04	0.068	-0.01	0.043	-0.03	0.055
Days in Unit Interaction	-0.05	0.057	0.16*	0.084	0.00	0.097
Days in Tchr. Demonstration Interaction	-0.02	0.075	-0.22***	0.070	0.08	0.120
Intercept	1.30***	0.251	1.67***	0.291	1.37***	0.413
Level 2 Variance	1.73		1.67		1.94	
Residual Variance	7.91		7.51		10.13	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
Days with Whole Class Discussion [6]						
Main Effect	3.12***	0.398	2.60***	0.204	3.30***	0.517
Days with Whole Class Discussion	-0.04	0.064	-0.08	0.071	-0.14	0.096
Days in Unit	0.01	0.051	0.05	0.042	-0.02	0.052
Days in Unit Interaction	-0.15	0.080	0.05	0.080	0.03	0.086
Days in Whole Class Discussion Interaction	0.25**	0.095	0.07	0.088	0.14	0.112
Intercept	1.21***	0.287	1.56***	0.286	1.34***	0.398
Level 2 Variance	1.70		1.55		1.82	
Residual Variance	7.90		7.58		10.13	
Days with Whole Class Lecture						
Main Effect	3.46***	0.346	2.76***	0.197	3.25***	0.496
Days with Whole Class Lecture	-0.14**	0.056	(-)0.17***	0.052	0.02	0.085
Days in Unit	0.04	0.050	0.07	0.040	-0.04	0.051
Days in Unit Interaction	-0.03	0.057	0.05	0.078	0.09	0.085
Days in Whole Class Lecture Interaction	0.14*	0.066	0.10	0.060	-0.17	0.101
Intercept	1.12***	0.246	1.44***	0.267	1.43***	0.382
Level 2 Variance	1.57		1.31		1.64	
Residual Variance	7.91		7.53		10.10	
Emphasis on Complex Goals						
Main Effect	3.04***	0.305	2.23***	0.193	3.24***	0.529
Complex Goals	0.80***	0.263	0.69**	0.248	-0.01	0.451
Complex Goals Interaction	0.44	0.412	(-)1.03**	0.377	0.21	0.644
Intercept	1.48***	0.213	1.89***	0.277	1.34***	0.408
Level 2 Variance	1.18		1.29		1.80	
Residual Variance	7.90		7.55		10.14	
Emphasis on Simple Goals						
Main Effect	3.34	0.341	2.32	0.191	3.25	0.525
Simple Goals	0.31	0.263	-0.58	0.366	-0.06	0.582
Simple Goals Interaction	-0.68	0.352	0.47	0.422	-0.17	0.668
Intercept	1.28	0.239	1.77	0.289	1.35	0.408
Level 2 Variance	1.61		1.50		1.92	
Residual Variance	7.91		7.58		10.11	

	Seventh-Grade Year 1 Experiment		Seventh-Grade Quasi-Experiment		Eighth-Grade Experiment	
Model	**Value**	**SE**	**Value**	**SE**	**Value**	**SE**
Emphasis on M1 Topics						
Main Effect	3.31***	0.323	2.35***	0.185		
M1 Topics	0.48	0.266	0.42	0.246		
M1 Topics Interaction	(-)0.89*	0.370	(-)0.58*	0.291		
Intercept	1.37***	0.226	1.75***	0.279		
Level 2 Variance	1.45		1.41			
Residual Variance	7.93		7.58			
Emphasis on M2 Topics						
Main Effect	3.01***	0.345	2.18***	0.255		
M2 Topics	1.01***	0.234	0.33	0.218		
M2 Topics Interaction	(-)1.26***	0.350	-0.56	0.330		
Intercept	1.75***	0.243	1.91***	0.327		
Level 2 Variance	1.42		1.35			
Residual Variance	7.87		7.59			

1 N varies for model: 7th Grade Y1 = 1,497, 7th Grade Quasi-Experiment =1,019, 8th Grade = 784.

2 N varies for model: 7th Grade Y1 = 1,520, 7th Grade Quasi-Experiment = 1,035, 8th Grade = 784.

3 N varies for model: 7th Grade Y1 = 1,463, 7th Grade Quasi-Experiment = 1,006, 8th Grade = 743.

4 N varies for model: 7th Grade Y1 = 1613.

5 N varies for model: 7th Grade Y1 = 1,546, 7th Grade Quasi-Experiment = 1, 019.

6 N varies for model: 7th Grade Y1 = 1,613.

Sponsor: The National Science Foundation, Grant REC - 0437861

Prime Grantee: SRI International. Center for Technology in Learning

Subgrantees: Virginia Polytechnic Institute and State University; University of Massachusetts, Dartmouth; The University of Texas at Austin; and The Charles A. Dana Center at The University of Texas at Austin

SRI International
Center for Technology in Learning
333 Ravenswood Avenue
Menlo Park, CA 94025-3493
650.859.2000

www.ctl.sri.com
www.math.sri.com