

Cognitive Complexity in Mathematics Teaching and Learning: Emerging Findings in a Large-Scale Experiment

Jeremy Roschelle, SRI International, Menlo Park CA USA, jeremy.roschelle@sri.com

Nicole Shechtman, SRI International, Menlo Park CA USA

Stephen Hegedus, University of Massachusetts, Dartmouth USA

Jessica Pierson, University of Texas at Austin USA

Michelle McLeese, Virginia Polytechnic Institute, Blacksburg VA USA

Deborah Tatar, Virginia Polytechnic Institute, Blacksburg VA USA

Abstract: Many Learning Sciences projects use technology to enhance the cognitive richness of teaching and learning, although few projects investigate the robustness of their approaches across a wide variety of teachers and classrooms. Our Scaling Up SimCalc project has completed experiments in which over a hundred mathematics teachers in 7th or 8th grade used either SimCalc or their existing materials. We found that students of teachers who used SimCalc learned more. In this paper, we look specifically at the issue of cognitive complexity through a case study of two teachers, contextualizing the case study within aggregate data. Aspects of the theme of cognitive complexity can be seen in (a) student performance on different types of test items (b) teachers' report of their own daily teaching goals and (c) discourse patterns in different classrooms. Although this analysis is preliminary, it reveals the potential of understanding Learning Science-based interventions more fully by combining large-scale and case study data.

Introduction

Many Learning Sciences projects use interactive technology to enhance teaching and learning. A common goal of this work is to increase students' opportunity to learn important, difficult concepts in science and mathematics. Further, a common approach is to draw upon the new representational affordances of technology, including visualization, simulation, and modeling. Although such work has been underway for many years, most projects work with just a few schools. Some important exceptions include Linn et al (2006), who found that "interactive visualizations combined with online inquiry and embedded assessments can deepen student understanding of complex ideas in science" across 16 schools in 5 states and Songer (2007), who carried out a deliberate program of research aimed at sustaining the "Kids as Global Scientists" program across a large, poor, urban area. These projects address science learning (see also Dede, Honan, & Peters, 2005; Schneider & McDonald, 2007). To our knowledge, our Scaling Up SimCalc project is the first random assignment experiment to examine a Learning Sciences-based approach in mathematics learning in more than 20 schools.

Ann Brown, a founder of the Learning Sciences, brought to our attention one important reason to study our approaches at scale: in the field, approaches can undergo "lethal mutations" which distort the designers' intent and undermine the mechanisms which lead to enhanced teaching and learning (Brown, 1991). In mathematics education, the TIMSS international comparison video studies revealed an important class of potentially lethal mutations (Hiebert et al., 2003). In particular, the TIMSS research found that teachers can modulate the cognitive complexity of the tasks offered to students in mathematics classes. In the United States, which has only average performance on TIMSS, *teachers* tend to do the more complex and conceptual parts of mathematical problems themselves, and tend to give *students* the simpler, more factual and procedural aspects. Teachers in high performing Asian countries, on the other hand, give students more responsibility for the complex aspects of mathematical tasks. This is not merely a cultural difference; research in Chicago found that students learn more when they receive more demanding intellectual assignments from their teachers (Bryk, Nagaoka, & Newmann, 2001). Similarly in Pittsburgh, Stein and colleagues found that students developed increased conceptual understanding of the mathematics in schools where teachers faithfully implemented more challenging problems (Stein, Grover, & Henningsen, 1996). Over-simplifying tasks thus could be an important (although not necessarily lethal) mutation for many Learning Sciences-based approaches; it could reduce the students' opportunities for rich cognitive engagement. Within our Scaling Up SimCalc project, therefore, we collected multiple measures relating to the cognitive complexity of both teaching and learning. Consequently, we examine variability in implementation and analyze the degree to which they increase or decrease learning.

In this paper, we report some emerging findings on the issue of cognitive complexity in mathematics teaching and learning, looking across different scales of our research and using multiple measures. In particular, we look both at aggregate data for 95 teachers and at a case study of two teachers. In the aggregate data, we find that students in SimCalc classrooms learned more and that, in particular, the gains occurred on the "complex" portion of our assessment, not on the "simple" portion. Further, the aggregate data shows that students learned

more when their teachers reported cognitively complex teaching goals. In the case study data, we look at a teacher whose students had highest learning gains, compared to a teacher with average gains. These teachers were selected from a pool of four teachers in case studies and we focus on these two teachers only because their data was analyzed first. In ongoing work, not available at the time of this writing, we will also look at teachers whose classrooms had lower performance. Given that we do not yet have a full spectrum of case study data available, our findings are necessarily tentative. Further, a limitation of our instruments is that they were not designed to capture a singular construct of “cognitive complexity” – as the famous Indian parable goes each “touches different parts of the elephant” (see http://en.wikipedia.org/wiki/Blind_Men_and_an_Elephant). Nonetheless, we see an important contribution taking shape in the relationships we can begin to see between teachers’ goals, teachers’ discourse moves, and student performance on more complex and simpler test items. We can also show that some less-than-ideal teaching moves do not eliminate the benefit of the SimCalc intervention we tested; teaching discourse moves have a bounded impact on student learning gains.

Overview of the Scaling Up SimCalc Project

The long-term mission of the SimCalc research group, based at the James J. Kaput Center at the University of Massachusetts, Dartmouth and extending to colleagues in many locations, is to “democratize access to the mathematics of change and variation” (Kaput, 1994). Key concepts in this strand of mathematics, which stretches from late elementary school through university courses, concern the mathematical analysis of rate and accumulation. These and additional related concepts in the strand are centrally important in much of science, technology, engineering and mathematics in secondary school and beyond. Early SimCalc work employed design research techniques to develop SimCalc MathWorlds software. This software presents the concepts of rate and accumulation in the context of animated motions, which may be controlled by linear and piecewise linear functions. Students can edit these functions in graphs with a mouse (or edit traditional algebraic symbols) and can execute the functions to see resulting animated motions. The software was always intended to be used in conjunction with curriculum materials that develop the concept of rate across multiple representations, including verbal, graphical, symbolic and tabular representations. As the project matured, an increasing focus was on the teacher professional development needed to support teachers in appropriate use of the integrated curriculum and software.

The Scaling Up SimCalc project, led by SRI International, focused on the research question: “Can a wide variety of teachers use an integration of curriculum, software, and professional development to increase student learning of complex and conceptually difficult mathematics?” The Scaling Up SimCalc project designed and conducted a series of large-scale experiments. This paper uses data from the first year of the 7th grade experiment. (The 7th grade experiment continued for an additional year and a separate 8th grade experiment was also conducted.) The 7th grade experiment compared students whose teachers used a 2-3 week replacement curriculum unit called “Managing the Soccer Team” with students whose teachers used their ordinary curriculum unit for the same topics. Teachers who used Managing the Soccer Team were trained in the use of the unit in the summer before implementation. The unit itself is deeply integrated with representational affordances SimCalc MathWorlds software; the text asks students to open particular software documents and most of the activities in the paper curriculum require students to edit and animate representations.

The experiment was conducted in Texas. We selected Texas because it is a large and varied state and has a stable and well-aligned system of standards in mathematics instruction. Further, our partner in Texas, the Charles A. Dana Center had the ability to reach teachers across the state and had a long-term commitment to increasing the number of students who enroll in AP Calculus, a course that occurs later in students’ progression on the mathematics of change and variation strand. We worked with the Dana Center to recruit volunteer teachers in 8 regions of Texas representing varied demographic conditions..

Teachers were randomly assigned to either use SimCalc in the first year or to use their ordinary curriculum materials in the first year and then use SimCalc materials in the second year. All teachers first went to a 2-day TEXTEAMS workshop, which introduces the concepts of rate and proportionality across multiple representations, giving all teachers good quality preparation for the topics we intended to measure. The SimCalc teachers stayed an additional three days to learn how to use the SimCalc materials. We collected several sources of data related to the theme of cognitive complexity. The student pre- and post-tests were designed to have both a “simple” and “complex” scale, corresponding to mathematics requiring simpler numeric computations and tasks requiring more elaborate analysis of functions across representations. Teachers were asked to complete a daily teaching log, which in part asked them to rate the cognitive complexity of their teaching goals. Further, we conducted case studies with a subset of teachers; for some of these cases, we can look at the same lesson as taught by different teachers.

Aggregate Data

We first consider the issue of cognitive complexity using aggregate data from the 95 teachers who completed the study reported more substantially in (Roschelle et al., 2007). The subsequent section presents case study data for two teachers.

Main Effect Findings: Gains on “Complex” Test Items

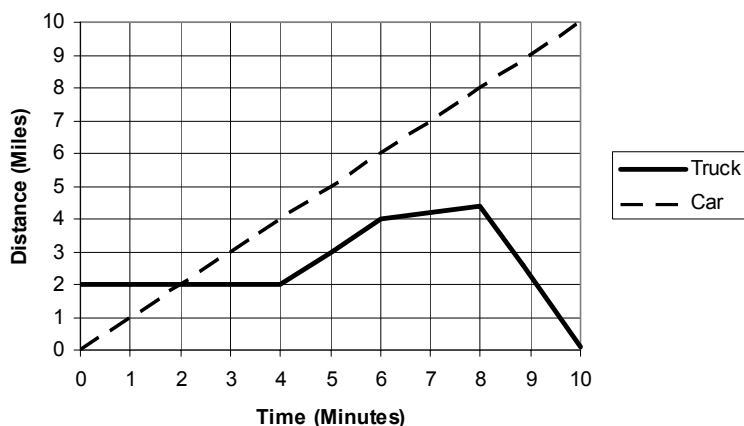
With a panel of mathematicians and mathematics education experts, we developed an assessment blueprint encompassing both simple and more complex aspects of rate and proportionality. The simpler items were based on items used on the Texas state test for seventh grade; these typically ask students to calculate using a proportional relationship stated either as a word problem or more mathematically. For example, a problem we will later refer to as the “car of the future” problem asked:

A car of the future will be able to travel 8 miles in 2 minutes.
How far will it be able to travel in 5 minutes?

Our expert panel also viewed proportionality as the basis of the first nontrivial function, $f(x) = kx$, that students learn in their multiyear journey through the mathematics of change and variation – the mathematical strand that continues through high school calculus. In preparation for future mathematics and science, they argued that students should learn to analyze this function across representations, including graphs, tables, and symbolic expressions. An item addressing this more complex approach asks students to complete several missing x and y values in a table describing a proportional function, then to write an algebraic expression for the function, and finally to sketch a graph of it.

Our expert panel also asked us to be sure to address common mathematical misconceptions. Hence, one item (see Figure 1) targeted a common misconception about position graphs: that the intersection of two graphed lines on a position vs. time graph indicates when two objects are moving at the same speed. The correct concept is that slope indicates speed, so that two objects have the same speed when their graphed lines have parallel slopes. We included this complex item because interpreting slope as a rate and connecting this representation to a narrative description of change over time is an important skill in all sciences: analysis of motion in physics, and interpreting rates of reaction in chemistry and rates of growth in biology are some of the many topics for which this concept is used.

A car and a truck are driving along a 10-mile road. The graph below shows their distance from the starting point over a 10 minute period.



When are the car and truck traveling at the same speed?

- A. Between 1 and 4 minutes
- B. At 2 minutes
- C. Between 4 and 6 minutes
- D. Between 6 and 8 minutes

Figure 1. An item which tests for the concept of slope as rate.

The overall 30-item test included 11 simple and 19 complex items. We carried out rigorous validation processes on the test, including cognitive interviews with students, item response theory analyses on field test data collected from a large sample of students, and expert panel reviews.

Our main effect was statistically significant and showed that students in the Treatment group learned more (see Figure 2 and Roschelle et al. 2007 for the primary discussion of these results). The overall effect size was 0.84, considered large in education studies ($t(93) = 9.1$, $P < 0.0001$, using a two-level hierarchical linear model with students nested within teacher). The difference between the groups occurred mostly on the complex portion of the test. The effect size of treatment on this portion was 1.22 ($t(93) = 10.0$, $P < 0.0001$). The effect size of the treatment on the simple portion was 0.17 ($t(93) = 1.8$, $P < 0.072$, n.s.); many students had mastered these concepts before the unit began, suggesting a ceiling effect. Because this data was analyzed in a hierarchical linear model, the statistical procedures yield an estimate of the variance at the teacher or classroom level vs. the variance at the student level. Only 21% of the “complex” portion and 18% of the total test score is accounted for by variance at the teacher or classroom level, suggesting that teachers had a limited ability to influence student learning with SimCalc, whether for better or worse, at least in the first year of implementation.

Because the SimCalc-related gains were predominantly realized on the “complex” portion of the test, it makes sense to further explore the issue of cognitive complexity within this data set.

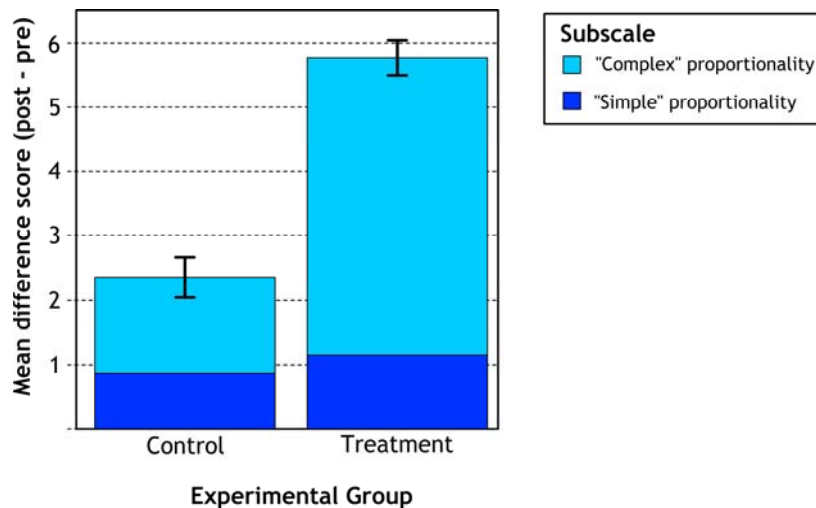


Figure 2. Student gain scores aggregated by teacher. Mean (\pm SE of total) difference scores on the 30-item student assessment.

Analysis of Simple and Complex Test Items

One way in which we further examined the nature of this finding was to look within the test at the particular items on which students in SimCalc classrooms exhibited strong gains. On the “car of the future” problem described above, approximately 75% of the students got the question right at pretest, suggesting that the item is pretty easy for many students, even before instruction on proportionality. After instruction (using either SimCalc or their existing curriculum) student performance was over 80%, but this gain is not statistically significant. This item is not trivial. However, it can be solved in a commonsensical way if students observe that in 4 minutes the car will go twice as far, covering 16 miles. Then the car goes for 1 more minute and would only go half as far, 4 miles. $16 + 4$ is 20 miles, the correct answer. Many students are competent with reasoning about “half as much” and “twice as much” years before 7th grade, making this a fairly simple item. Alternatively, the item can be solved procedurally, by setting up a proportional relationship and using the cross-multiplication procedure; this procedure is commonly taught in 7th grade instruction. This item was fairly typical of the “simple” items in our assessment in that students are given 3 numbers and asked to compute a missing number.

In contrast, on the item represented in Figure 1, student performance at pretest was around 23%, which is slightly worse than random guessing would produce. We argue that students find this item more difficult because it requires knowledge that goes beyond the arithmetic skills that predominate the early years of school. The target concept comparison of “speed” requires identifying a quantity which is not directly given by the representation. Instead it must be inferred from the slope of the line, which requires conceptual understanding of the meaning of the slope. Further, there is a significant perceptual distracter – the point of intersection. Indeed, most students did not guess randomly on the pre-test: most chose answer (b), which represents the target misconception. On the posttest, 55% of students who had been in classrooms using SimCalc got the item right, a statistically significant gain. In comparison, only 38.5% of students in non-SimCalc classrooms got the item right at posttest and more students (55%) chose the misconception-based distracter, answer (b).

The contrast between these two items reflects the distinction in our assessment rubric between “simple” items in which students are given 3 or 4 numbers and must compute an unknown value and more “complex” items in which students must use conceptual understanding of how functions describe rates across

multiple representations, often including graphs. In the case of the item in Figure 1, the function is represented in a graph and must be related to the verbal concept of “speed.”

Teacher Self-Report of Teaching Goals

We asked teachers in the study to complete a teaching log after each day of instruction. The log included a scale that asked teachers to rate their degree of focus for the day’s class on goals of low cognitive complexity (memorization and use of routine procedures) and high cognitive complexity (communicating conceptual understanding; making mathematical connections and solving nonroutine problems; and conjecturing, generalizing, or proving). This scale was based upon a previously validated scale (Porter, 2002).

For high-order goals, in both studies, Treatment teachers (versus Control) reported a stronger daily focus ($t(93) = 2.3$, $P < 0.05$) and there was an overall positive statistical association between teachers’ report of more complex goals and student gain on the complex math subscales ($P < 0.0001$). For lower-order goals, non-SimCalc teachers reported a stronger daily focus ($t(93) = 3.1$, $P < 0.01$), and there was an overall negative statistical association with classroom mean student gain on the complex math subscale ($P < 0.05$). Figure 2 shows a scatter plot of the association between a teacher’s self-report of cognitive complexity of their teaching goals (horizontal axis) and the mean student learning gain in a teacher’s classroom. When teachers report cognitively complex teaching goals their students learn more *regardless of condition* (although students in the SimCalc condition, represented by the symbol “x” in Figure 3 learned more overall).

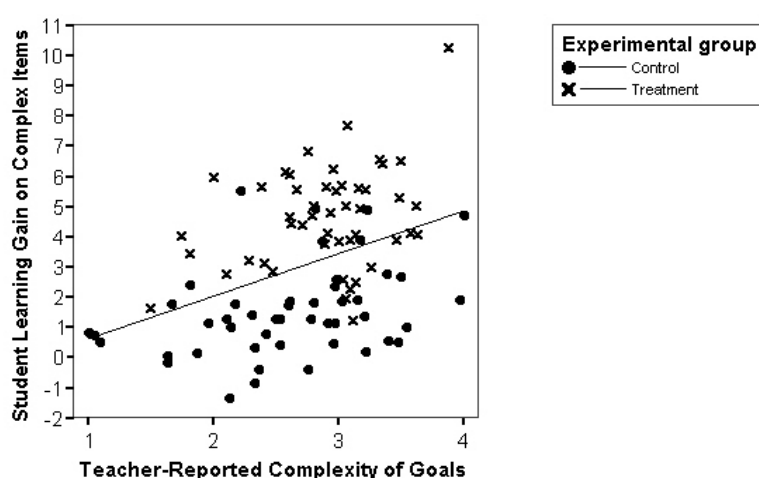


Figure 3. Teachers who reported more complex teaching goals had stronger classroom learning gains.

Case Study Data

In the context of her dissertation research, Jessica Pierson (a co-author of this paper) further analyzed discourse in SimCalc classrooms (Pierson, in progress). Discourse analysis is an important Learning Sciences methodology and has the potential to yield a thicker description of how cognitive complexity is enacted in SimCalc classrooms. Pierson selected teachers and performed her case studies before the aggregate data reported above was available. Two case study teachers, code-named “Mary” and “Nancy” are discussed below. These teachers were selected, in particular, because we had a video tape of each teaching the same lesson (both used SimCalc). Both teachers taught in small cities in Texas and had over 20 years of experience. They both had classrooms in which most students were white and about 1/3 were Hispanic. Mary’s class had pre-test scores that were approximately at the mean of all classrooms in the study; Nancy had students who scored about 1.5 test items above the mean (out of 30 total items). Pierson (2007) described her overall impressions of the two classrooms as follows:

Mary framed discussions using core mathematical concepts to organize the lesson’s trajectory. Discussions had a sense of cohesion, and procedures and formulas were introduced as a natural part of answering central questions related to the big ideas of speed and rate. She probed their thinking and allowed their observations and conjectures to be the impetus for further class discussion. She also encouraged students to test their thinking, reasoning, and problem solving with one another. Nancy was more task-oriented and preferred to present mathematics as a series of small, manageable tasks and calculations.

As it turned out, Mary’s students displayed the greatest learning gains of any classroom in the study; her classroom’s gain from pre-test to post-test was more than two standard deviations above the norm. Nancy’s students had above average gains, although within a standard deviation of the mean.

Discourse Analysis: Teacher Follow-Up of Student Responses

Because classroom discourse is broad and complex, Pierson focused on one aspect of discourse and its relationship to students' opportunities to learn: teacher follow-up of student responses. In particular, Pierson focused on the "prospectiveness" of teacher moves in the final slot of a traditional initiation-response-evaluation pattern. Instead of an evaluation (e.g. "good" or "that's right") that ends the sequence, a teacher can use the slot to encourage students to elaborate or extend their thinking. A move that uses the student response to drive the discussion forward is more "prospective" – it anticipates that students are capable of more extended argumentation and deeper thinking.

Pierson (2007) found that Mary had more prospective moves than Nancy. Whereas Mary used students' responses in giving or demanding more information, Nancy tended to merely "rebroadcast" the student response. Not surprisingly, turn-taking sequences on a topic were longer in Mary's class (average of 13.5 moves per sequence) than Nancy's class (average of 7.5 moves per sequence).

It is particularly interesting to observe the following dialog (from Nancy's class), which reflects the lesson in the unit that most directly relates to the misconception captured in the item in Figure 1. In this lesson, students are comparing the trips of a bus and car as represented in graphs that look similar to Figure 1; in addition, during the lesson students can animate the functions and observe the corresponding motions.

Turn	Speaker	Text
<i>Sequence 34</i>		
150	Teacher	Can you see when you replay the simulation when the bus started to slow down?
151	Student	Yes.
152	Teacher	Right before it got to the end, correct? Ok at what?
153	Student	140.
<i>Sequence 35</i>		
154	Teacher	Ok. So now let me ask you some questions. What, how long did it take them for this trip? Raise your hand. How long did it take em for this trip? Chris.
155	Chris	Three hours.
156	Teacher	How do you know it went three hours?
157	Chris	Cuz it ends, the line ends at three.
158	Teacher	Very good, cuz the line for the bus or the van?
159	Mult S's	Both.
160	Teacher	Both end at what?
161	Mult S's	Three hours.
162	Teacher	Three hours. So you know it took them three hours.
<i>Sequence 36</i>		
163	Teacher	Was the speed of the two vehicles the same?
164	Mult S's	No.
165	Teacher	No. You know that by how?
166	Student	They're not the same.
167	Teacher	Because of the (pause) graph right? And because of also running the simulation. They didn't stay right beside each other, did they?
<i>Sequence 37</i>		
168	Teacher	Ok, umm, how far did they go in 3 hours? Raise your hand. How far did they go in three hours? Charles?
169	Charles	** (unintelligible)
170	Teacher	How'd you know that?
171	Charles	**
172	Teacher	180 miles where? (pause - no student response). On the graph, right?
<i>Sequence 38</i>		
173		Ok. And let's see if you can calculate the speed. Go through, run your simulation and see if you can calculate the speed for the van and then calculate the speed for the bus and then also calculate the speed for the bus when it started to slow down since you guys have told me it started to slow down.

Table 1. Discourse Excerpt from Nancy's Class

Sequences 34-38 were orchestrated to accomplish the teacher's goal of calculating the speeds for the van and bus. Notice that this larger goal was not mentioned to the students until the last sequence (see Turn 173). Instead, it remained hidden in a series of loosely-related questions (from the students' perspective) that identified the necessary components for calculating speed. Additionally, as seen in this excerpt, little control or responsibility is given to the students as the teacher generated and controlled the topic of discussion.

Conversation was tightly scripted with little room for student ideas to shape the discussion. It is as if Nancy had a predetermined path that she was leading her students down: she may have heard their responses but did not incorporate them substantially into the flow of the lesson. In fact, there was very little of substance for her to incorporate since students' roles were limited to paying attention, performing basic calculations, recalling formulas, and answering yes/no or short-answer questions.

In contrast, in the discussion of the same graph in Mary's classroom, one student thought two piecewise linear functions depicted different routes for two vehicles instead of different speeds. Instead of correcting her, Mary let the class struggle with this interpretation (the sequence was 54 moves long) asking the student first to explain her reasoning and then providing her classmates the opportunity to share their thinking.

Supporting Data from Additional Measures and Analyses

Because Mary and Nancy were two of the 95 teachers in the larger study and because several kinds of analyses are taking place, we have additional data about these two teachers. As mentioned both teachers had many years of experience and taught in schools that are similar in many respects. Nancy's students were "smarter" although Mary had a smaller class.

Although Pierson was not aware of this when choosing the cases or analyzing the discourse data, the teachers differed in their goals and level of mathematical knowledge. Mary reported more complex teaching goals than Nancy. Indeed, in the graph in Figure 3, she is represented by the top-right "x." Nancy chose higher-than-average values on the Likert scales for cognitively complex teaching goals, but she also chose higher-than-average values on the scales for simple teaching goals, suggesting that she tried to balance between simple and complex teaching goals. On the test of "Mathematical Knowledge for Teaching" given to all teachers at the beginning of the study, Mary got a higher than average score and Nancy got a lower than average score. This suggests that Mary knew the mathematics covered in the SimCalc replacement unit better than Nancy did. We would conjecture that knowing more mathematics and adopting more complex teaching goals are related to and supportive of the discourse style that Mary adopted.

It is interesting to analyze the performance of students in Mary and Nancy's classroom on the "slope as rate" problem portrayed in Figure 1, as the discourse analyzed by Pierson related directly to this problem. In the overall data set, 55% of students who used SimCalc and 38.5% of students who used their ordinary curriculum got this item right on the post-test. In Mary's class, 67% of the students got this problem right – a better-than-average performance. In Nancy's class, only 35% of the students got the problem right and almost twice as many chose the misconception-based distracter.

Nancy's students did gain more than Mary's students (and more than students in the average classroom) on the "simple" portion of our test; thus her focus on cognitively simple tasks apparently did have a pay-off. Yet, we wonder if it was worth it. Recall that Nancy's started out with a higher-than-average classroom pretest. This suggests her students were likely ready for more complex mathematics and could have learned more if presented with an appropriate opportunity to learn. This raises the classic teaching problem of "diminishing returns" for a continued focus on skills that many students already know well.

Any conclusions drawn from this data can only be preliminary, as there is much more data under analysis. Another crosscutting analysis undertaken by sociology graduate student Michelle McLeese at Virginia Tech (another co-author of this paper) also picked out Mary as an unusual teacher based on her teaching behavior, starting from a different perspective. Mary was one of four teachers out of 24 teachers observed to explicitly mention Willie Nelson's song "On the Road Again" in teaching the unit called "On the Road." Willie Nelson is an iconic Texas figure and this song can be expected to be well known by the students. Of the four teachers, she was the only one to engage in substantial elaboration of this relationship; she wrote a song for the class. One way of conceiving Mary's success is that it reflects a kind of *mathematical opportunism*, which can lead to cognitively complex goals and to prospective interactions as well as elaboration of meaningful contexts.

Nancy's and Mary's classrooms represent only two of large set of teachers being studied in case studies from several perspectives. We have additional interview data for all teachers. Further, we will later consider teachers with classrooms with below-average pretest scores or below-average gains. In addition, we have data from both Nancy and Mary in a second year of implementation and Mary's second year class does not gain have the top learning gains. However, in the second year, Mary's class started with a considerably higher pre-test score (making it hard to replicate the gains) and reached a higher post-test score as well. This illustrates some of the nuance that must be considered before reaching conclusions.

Discussion

We have examined the issue of "cognitive complexity" across two different scales of data and many several different measures within a data set that presents a "main effect" of increasing student learning of complex mathematics. Although SimCalc has features in common with many other Learning-Sciences-based approaches for increasing opportunities to learn rich content, few projects have an opportunity to triangulate on the issue of cognitive complex from different scales of analysis and different data sources. Admittedly, we do

not have a single definition of cognitive complexity across all measures in our study. We have, however, explained the definition in use in each measurement instrument in the study.

Although our findings are preliminary, a convergent story is emerging. Teachers may moderate the impact of an intervention like SimCalc by choosing cognitively complex teaching goals. These goals may then be enacted in many ways in the classroom. One difference in enactment may be in the prospectiveness of teacher discourse moves in the “E” slot of the Initiate-Respond-Evaluate/Extend sequence. Other differences in enactment may relate to teacher’s knowledge of mathematics and comfort in working through “misconceptions” with students. The effects of these differences may be most profound on test items that address deep-seated conceptual difficulties.

Overall, however, it is worth bearing in mind that only 21% of the variance in student learning gains on the complex portion of our test is accounted for at the teacher level. The rest of the variance is at the student level. In this paper, this is reflected in the fact that Nancy’s students performed well despite some features of her teaching that may be less than ideal. Indeed, strength of the SimCalc intervention we studied is that students on a whole learn more with it, despite having a wide variety of teachers and classroom situations. Our conclusion therefore is not simple. First, the SimCalc intervention (on average) enabled students learn complex mathematics better than their existing curriculum did. Second, teachers who reported more complex goals had a measurable but limited impact in modulating the degree to which students learned more complex mathematics, perhaps because of the way discourse was enacted in their classrooms.

References

- Brown, A. (1991). Design experiments. Theoretical and methodological challenges in evaluating complex interventions in classroom settings. *Journal of the Learning Sciences*.
- Bryk, A. S., Nagaoka, J. K., & Newmann, F. M. (2001). *Authentic Intellectual Work and Standardized Tests: Conflict of Coexistence?* : Consortium on Chicago School Research.
- Dede, C., Honan, J. P., & Peters, L. C. (Eds.). (2005). *Scaling up success: Lessons learned from technology-based educational improvement*. San Francisco: Jossey-Bass.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching Mathematics in Seven Countries: Results From the TIMSS 1999 Video Study*. Washington DC: National Center for Educational Statistics.
- 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.
- Linn, M., Lee, H.-S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science*, 313(August 25), 1049-1050.
- Pierson, J. (2007). *The impact of moment-to-moment discourse moves on opportunities to learn mathematics*. Paper presented at the 29th Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education, Lake Tahoe, Nevada.
- Pierson, J. (in progress). *The Relationship Between Teacher Follow-up Moves and Mathematics Learning*. Unpublished doctoral dissertation, University of Texas at Austin.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.
- Roschelle, J., Tatar, D., Shechtman, N., Hegedus, S., Hopkins, B., Knudsen, J., et al. (2007). *Can a technology-enhanced curriculum improve student learning of important mathematics? Results from 7th Grade, Year 1* (No. 1). Menlo Park, CA: SRI International.
- Schneider, B., & McDonald, S.-K. (Eds.). (2007). *Scale-up in Education: Issues in Practice* (Vol. 2). New York: Rowman & Littlefield Publishers, Inc.
- Songer, N. B. (2007). Rethinking sustainability of curricular innovations: Notes from urban Detroit. In B. Schneider & S.-K. McDonald (Eds.), *Scale-up in education: Issues in practice* (Vol. 2, pp. 165-182). New York: Rowman & Littlefield Publishers, Inc.
- Stein, M. K., Grover, B. W., & Henningsen, M. A. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, 33(2), 455-488.

Acknowledgments

This material is based on work supported by the National Science Foundation under Grant No. 0437861. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The Scaling Up SimCalc Project required the contributions of a large and dedicated team and the participation of many teachers, school leaders and teacher educators. Although we cannot individually thank each person in this space, we are nonetheless grateful for his or her contributions.