Journal of Mathematics Education at Teachers College

Fall – Winter 2012

A Century of Leadership in Mathematics and its Teaching

TABLE OF CONTENTS

Preface

Articles

TABLE OF CONTENTS, continued

EQUITY NOTES FROM THE FIELD

Other

Journal of Mathematics Education at Teachers College
Fall–Winter 2012, Volume 3

How Curriculum and Classroom Achievement Predict Teacher Time on Lecture- and Inquiry-based Mathematics Activities

Julia H. Kaufman
University of Pittsburgh

Rita Karam
John F. Pane
RAND

Brian W. Junker
Carnegie Mellon University

This study drew on data from a large, randomized trial of Cognitive Tutor Algebra (CTA) in high-poverty settings to investigate how mathematics curricula and classroom achievement related to teacher reports of time spent on inquiry-based and lecture-based mathematics activities. We found that teachers using the CTA curriculum reported more time on inquiry-based activities and less time on teacher lecture activities overall compared to non-CTA teachers. However, both CTA and non-CTA teachers of the highest-achieving students spent more time on inquiry-based activities compared to teachers of lower-achieving students. Additionally, CTA teachers in classrooms with the most low-achieving and non-gifted students reported almost as much time on lecture-based activities as their non-CTA counterparts. Qualitative findings suggest that CTA teachers engaged in more traditional lecture-based activities and fewer inquiry-based activities when they thought their lower-achieving students could not tackle the reading open-ended activities in the curriculum without explicit demonstration and traditional practice problems. CTA thus appeared to increase inquiry-based activities in teachers' classrooms overall. However, lower-achieving students may have had needs unaddressed by the CTA curriculum. These findings thus imply that districts should think carefully about how to implement CTA and—potentially—other inquiry-based curricula in order to support teachers of students with the highest needs and least preparation.

Keywords: mathematics curriculum, inquiry-based activities, peer achievement

Introduction

Students in low-achieving classrooms and schools are consistently exposed to more teacher lecturing and less complex problem-solving activities compared to their high-achieving peers (Nystrand & Gamoran, 1988; Oakes, Gamoran & Page, 1992; Oakes, Ormseth, Bell & Camp, 1990). In part as a response to these differences in instruction for lower- versus higher-achieving students, the National Council of Teachers of Mathematics released standards documents (1989; 2000) motivating a new wave of mathematics curricula de-emphasizing teacher lecturing and traditional worksheets in favor of teachers engaging students in more open-ended, conceptually-based activities, with the teacher as facilitator. We refer to such curricula as "inquiry-based" in this paper.

In some settings where teachers implemented inquiry-based curricula with a high level of fidelity and quality, student achievement increased (Briars & Resnick, 2000; Knapp, Shields & Turnbull, 1995; Silver & Stein, 1996). However, studies in these settings were observational rather than experimental. Furthermore, the studies did not explore the extent to which teachers' implementation of inquiry-based curriculum tasks was dependent on characteristics of the students that those teachers served. Because inquiry-based tasks can be challenging for teachers to implement well (Ball, 2001; Cohen, 1990) and dependent on teachers' preconceptions about their students (Boaler, 2002), a relationship between student characteristics and teachers' use of inquiry-based tasks is likely present.

In this study, we drew on data from a large, randomized trial of Cognitive Tutor Algebra (CTA) in high-poverty settings to investigate how achievement among teachers' students and other student characteristics were related to teacher reports of time spent on inquiry-based and more traditional lecture-based activities. We further considered whether use of CTA, which emphasizes inquiry-based classroom activities, influenced the relationship between classroom achievement and the time teachers spend on these activities.

Brief Review of Relevant Literature

Effects of classroom achievement on individual student mathematics achievement range from moderate effect sizes of .2 to .4 (Hanushek, Kain, Markham & Rivkin, 2003; Hoxby, 2000; Kang, 2007) to effect sizes of less than .2 (Burke & Sass, 2008; Lefgren, 2004; Vigdor & Nechyba, 2004) or non-significant effects (Angrist & Lang, 2004) depending on the study. The considerable variation among these findings points to potential unobserved differences in the mechanisms by which classroom achievement influences the achievement of individual students (Cooley, 2009; Lefgren, 2004). Because most students are not randomly sorted into classrooms—or schools—multiple unmeasured characteristics of lower- versus higher-achieving students, their teachers, and the schools they attend could be influencing individual student achievement.

One particular mechanism that has gone unexplored in studies of classroom achievement is how instruction in high- and low-achieving classrooms differs and whether curricula can influence that difference. While inquiry-based curricula are intended to address tendencies for teachers to neglect complex problem-solving activities and spend more time talking than listening to students, the work of Boaler (2002), Lubienski (2000) and others suggests that such curricula may not be used by teachers or students as intended, especially if teachers and/or students feel that the curriculum tasks are unfamiliar or challenging. To make matters more complicated, lower-achieving students likely benefit from at least some explicit, teacher-led instruction embedded within or apart from inquiry-based curriculum when they do not grasp complicated mathematical concepts embodied in open-ended tasks (Baker, Gersten & Lee, 2002; Baxter, Woodward & Olson, 2001; Woodward & Brown, 2006).

By exploring how classroom achievement influences time on inquiry-based problem-solving activities and lecture-based activities among teachers using Cognitive Tutor Algebra versus another curriculum, we fill a gap in our knowledge about how inquiry-based curriculum support teachers of diverse student populations. Additionally, we address two key but unexplored factors likely mediating the influence of classroom achievement on individual students: curriculum and teachers' reports about their instruction.

Methodology

Setting

We drew on data from a evaluation of Cognitive Tutor Algebra (CTA) curriculum, which used a school-level randomized experimental design with 150 middle and high schools across seven states among two separate cohorts of students over a two-year period. Control schools not assigned to CTA continued with whatever algebra curriculum they were already using. Our analysis focused only on teachers' reports about their classroom instruction apart from the computer lab portion of CTA or any control school curriculum. CTA emphasizes inquiry-based classroom instruction through small group learning and discussion in a learner-centered environment where students work together toward the mathematical goals of the lesson and the teacher acts as a facilitator of that learning (Cognitive Tutor Algebra I Development Team, 2007).

Data Sources

To estimate what student and teacher characteristics were associated with teachers' reports of time spent on lecture-based and inquiry-based student thinking activities, we used data from a teacher survey distributed online to all teachers at the end of the school year[1] in combination with student pre-test scores and student demographic information provided by schools at the beginning of each year. For the survey, teachers were asked to keep in mind a "target class" for which to respond about their time spent on classroom activities, and that target class was randomly assigned from a list of each teacher's algebra classes.

Our two dependent variables—instructional time spent on teacher lecture and student thinking activities—are two composites derived from a larger set of 27 survey items asking teachers to estimate the percentage of time they and their students spent on a variety of instructional activities. To create these two composites, we grouped 10 of the 27 items into two conceptual categories: (a) time on lecture-based "teacher talk" activities like "demonstrating how to do a procedure" or "reviewing previously covered material" (5 items) and (b) time on inquiry-based "student thinking" activities like student work on "solving non-routine mathematical problems" or "explaining their reasoning" (5 items). The individual items in each composite are listed in Table 1.

We then performed confirmatory factor analysis for the items across these two categories using a maximum likelihood extraction method with two factors and varimax rotation with Kaiser normalization. The rotated factor matrix provided clear evidence of two separate factors (goodness-of-fit chi square=111.39, df=26, $p<.001$): the five items that were part of the "teacher talk" factor had loadings above .57 alongside "student thinking" items with factor loadings of .06 to .19, and the five items that were part of the "student thinking" factor had loadings above .69 alongside "teacher talk" factor loadings of .08 to .23. Reliability analysis indicated a Cronbach α of .85 for "teacher talk" activities and Cronbach α of .89 for "student thinking" activities.

[1] The survey response rate in Year 1 was 84% for non-CTA teachers and 90% for CTA teachers; in Year 2, the response rate was 88% for non-CTA teachers and 89% for CTA teachers.

Table 1. Survey Items in Teacher Talk and Student Thinking Composites

Composites based on teachers' response to survey question, "What percentage of time did you/students spend on each activity in the classroom (when the whole class was not working on math exercises on the computer)?"
1=No instructional time; 2=<10%; 3=10-25%; 4=26-50%; 5=51-75%; 6=>75%
TEACHER TALK ACTIVITIES ($\alpha = .85$)
a. Teacher presenting new information
b. Teacher demonstrating how to do a procedure or solve a problem
c. Teacher reviewing previously covered material
d. Students taking notes
e. Students listening to teachers explain math concepts to the whole class or student groups
STUDENT THINKING ACTIVITIES ($\alpha = .89$)
a. Students solving non-routine mathematical problems
b. Students explaining their reasoning or thinking process
c. Students making estimates, predictions or hypotheses and investigating them
d. Students conducting proofs or demonstrations of their mathematical reasoning
e. Students analyzing data to make inferences or draw conclusions

At the beginning of the year, students took the CTB/McGraw-Hill Algebra Readiness Exam. Students' raw scores on the exam were mapped to a scale score using a three-parameter IRT model. Scores were then standardized across all students. Schools provided student demographic information including ethnicity, gifted status, and the presence of any special education disability.[2] For all student-level variables, we calculated a mean at the school-level, across all of each teacher's students, and—for teachers with more than one class participating in the study—the mean for the "target class" about which the teacher was requested to respond in the survey. Other survey variables used in this analysis included teachers' background characteristics (e.g. teaching experience), algebra professional development hours, perceptions about CTA curriculum, and other variables that could be tied to time spent on particular instructional activities.

Analysis

In our preliminary hierarchical linear models (HLMs), we considered (a) how mean student characteristics at the target classroom, teacher and school level (on their own and in interaction with being a CTA teacher) were related to teachers' reports of time on teacher talk and student thinking activities; and (b) the extent to which other teacher-level factors drawn from surveys mattered for time spent on teacher talk and student thinking activities.

We systematically examined the effects of all these variables in a series of HLMs, using the lme4 package (Bates, Maechler & Bolker, 2011) in R (R Development Core Team, 2011). In our final models, we included mixed effects that provided us with the best model fit, based on deviance statistics, and we omitted variables that did not make a difference for the fit of the model. Among the variables we omitted because of non-significant effects were the percentages of teachers' students from each ethnicity, teachers' background characteristics—including teacher experience, certifications and education—and teachers' perceptions of curricula. Thus, the only variables measuring student characteristics included in our final models were those related to achievement: mean pre-test scores, percentage of gifted students, and percentage of students with a disability. We also omitted random slopes for mean student-level and teacher-level variables within schools from our final models, as those random effects did not improve model fit.

We decided to use the mean of pre-test achievement, percentage gifted and percentage with disability in our final models at the *teacher level* only (across all of each teacher's students, rather than at the school level or for each teacher' target class about which they were instructed to respond in the survey). We made this decision because mean student pre-test achievement, percentage gifted, and percentage with disability at the school-, teacher- and target classroom-level were very highly correlated within one another ($p<.001$). For example, teachers with high pre-test scores in their target class generally had high pre-test scores across all the algebra classes they taught that were included in our study, and mean pre-test scores were similarly high across those teachers' schools. Thus, looking at means for school, teacher or target class separately in different models yielded similar effects. That said, peer effects at the teacher level (rather than the target class or school level) generally yielded the most significant effects. We thus included pre-test achievement and percentage students gifted/learning disabled as teacher-level averages in the models with the assumption that teachers may have found it difficult to differentiate among their classes when responding to the survey and/or their teaching was similar in all of their classes. The only teacher-level survey variable related to time on teacher talk or student thinking activities was teachers' hours of algebra professional development.

Table 2 provides means and standard deviations in Year 1 and 2 for variables included in our final models,

[2] Students labeled as having a disability includes students with autism, an emotional or learning disability, mental retardation, a hearing or speech impediment, a traumatic brain injury or multiple disabilities.

comparing CTA and non-CTA teachers. Independent t-tests indicated significantly less time on teacher talk activities and more time on student thinking activities among CTA teachers compared to their non-CTA counterparts in both years. Algebra PD hours were also higher for CTA teachers compared to non-CTA teachers in Year 1.

As expected given school-level randomization in this study, CTA and non-CTA teachers were similar in regard to percentages of gifted and disabled students. However, CTA teachers had students with significantly lower average pre-test scores compared to non-CTA teachers in Y1. Authors (2011) suggest that this imbalance may have been due to treatment schools' deliberate assignment of higher-achieving students to classes that did not use CTA. To confirm this imbalance did not impact our results, we included percentage of gifted students and percentage of students with disabilities—other potential proxies for pre-test achievement—in separate models from pre-test score. Since the percentage of gifted and percentage of disabled students are both highly correlated with pre-test score [respectively, correlations of .54 ($p<.001$) and –.42 ($p<.001$)], any triangulation of findings from these data sources would suggest that CTA versus non-CTA differences in pre-test score do not matter for our findings.

In our final models listed below, we included a random effect for teacher—as 122 teachers took the survey both years—and for school. Y_{ijk} represents Teacher Talk or Student Thinking at occasion i for teacher j at school k. The level 1 fixed effect for average algebra readiness was considered in separate models from fixed effects for percentage gifted or disabled students (model for Level 1A versus 1B in Figure 1) because of high correlations of pre-test with percentage of gifted or disabled, which would make it difficult to ascertain their predictive power if they were all in the same model.

Following our quantitative analyses, we briefly consider potential explanations for the quantitative findings based on focus group interviews with one third of all CTA teachers. Using these focus group interviews, we specifically examined differences in descriptions of instruction and reasons for instructional choices among teachers of lower-versus higher-achieving students.

Findings

Being a CTA teacher had an appreciable negative effect on Teacher Talk and a positive effect on Student Thinking in HLMs (see Table 3). Additionally, teachers with a higher percentage of students with a disability reported more Teacher Talk activities, whereas teachers of students with a higher pre-test score and more gifted students reported more Student Thinking activities.

The interaction between being a CTA teacher and pre-test score/percentage of gifted students also mattered a good deal for Teacher Talk activities. Specifically, CTA teachers of students with higher pre-test scores and more gifted students reported less time on Teacher Talk than non-CTA teachers of higher-achieving, gifted students. For example, teachers with highest-achieving students (teachers in the top quantile in terms of average pre-test scores) reported Teacher Talk averages almost three-quarters of a standard deviation lower than non-CTA teachers in the top quantile for average pre-test scores. On the other hand, almost no difference in teacher talk activities was present for CTA versus non-CTA teachers in the lowest average pre-test score quantile. Thus, while CTA did make a difference for Teacher Talk among teachers of highest-achieving students,

Table 2. Means and Standard Deviations for Variables Used in Analysis

		Non-CTA Teachers			CTA Teachers		
		N	*M*	*SD*	*N*	*M*	*SD*
Time on Teacher Talk activities	Y1**	125	3.56	.88	105	3.20	.85
	Y2**	130	3.45	.78	97	3.12	.70
Time on Student Thinking activities	Y1**	124	2.60	.97	103	3.06	1.05
	Y2*	130	2.60	.94	97	2.88	.85
Pre-test score (teacher-level average)	Y1*	118	-.01	.71	105	-.20	.59
	Y2	131	-.06	.80	97	-.22	.62
% gifted students (teacher-level average)	Y1	108	.07	.19	94	.08	.19
	Y2	129	.17	.27	97	.12	.22
% disabled students (teacher-level average)	Y1	104	.08	.22	89	.07	.19
	Y2	120	.07	.19	91	.09	.23
Algebra PD hours (log)	Y1*	121	2.78	1.35	98	3.10	.88
	Y2	123	2.38	1.43	92	2.48	1.45

*$p<.05$ in independent t-tests comparing CTA and non-CTA teachers; **$p<.01$

Level 1A (percentage of gifted students and students with disability excluded from this model):

$$Y_{ijk} \text{ (Teacher Talk or Student Thinking)} = \beta_{0jk} + \beta_1(CTA_{jk}) + \beta_2 \text{ (average algebra readiness}_{ijk}) + \beta_3 \text{ (CTA x average algebra readiness}_{ijk}) + \beta_4 \text{(algebra PD hours}_{ijk}) + e_{jk,},$$
$$e_{ijk} \sim N(0, s^2)$$

Level 1B (average algebra readiness pre-test scores excluded from this model):

$$Y_{ijk} \text{ (Teacher Talk or Student Thinking)} = \beta_{0j} + \beta_1(CTA_{jk}) + \beta_2 \text{ (\% gifted students}_{ijk}) + \beta_3 \text{ (\% students w/ disability}_{ijk}) + \beta_4 \text{ (CTA x \% gifted students}_{ijk}) + \beta_5 \text{ (algebra PD hours}_{ijk}) + e_{j,},$$
$$e_{ijk} \sim N(0, s^2)$$

Level 2 (teacher):

$$\beta_{0jk} = \pi_{00k} + r_{0jk,}, \qquad r_{0jk} \sim N(0, t^2)$$

Level 3 (school):

$$\pi_{00k} = \gamma_{000} + \mu_{00k,}, \qquad \mu_{00k} \sim N(0, t^2)$$

Figure 1. Models for Level 1A, Level 1B, Level 2, and Level 3

it made little difference for teachers of the lowest-achieving students. We also found a significant, positive effect for algebra professional development hours, where teachers with more professional development hours reported more time on Student Thinking activities.

We also conducted focus group interviews with one third (n=43) of all the CTA teachers, and we used these data to investigate explanations for differences in instruction among teachers of lower- versus higher achieving students. Twenty-four of the 43 teachers in focus groups taught students in the bottom 50% in terms of teacher-level average pre-test scores. Teachers of these lower-achieving students at six different schools indicated that the extensive reading required within the CTA curriculum presented a major challenge for their students and limited inquiry-based activities. Teachers of lower-achieving students across eight schools also spoke about not integrating group and inquiry-based activities on a regular basis, for reasons ranging from students being unmotivated and not able to engage in open-ended group work to students not having the basic background knowledge. One such teacher said, "I don't [do groups] anymore because the kids, there's so little self-motivation with the kids that they weren't doing anything until I got to their group... we try [facilitation], but it is hard. They are used to the typical, the teacher's in front of the class. That's how they're used to [the] teaching, and that's how we're used to teaching." Because many teachers felt their students could not handle the reading and complex problem-solving work, they spoke of integrating more teacher explanations, repetition and traditional practice problems into their lessons.

Summary and Discussion

Our findings suggest that Cognitive Tutor Algebra curriculum helped teachers focus more on inquiry-based activities compared to curricula used in control schools. However, by itself, CTA curriculum did not ameliorate differences in time spent on such activities among teachers of lower- versus higher-achieving students. In our study, CTA teachers did report more time on student thinking activities and less time on teacher lecture activities overall compared to non-CTA teachers. However, CTA teachers of lowest-achieving and fewest gifted students reported almost as much time on lecture-based activities as their non-CTA counterparts. CTA teachers of lower-achieving and more non-gifted students also reported more time in inquiry-based activities—on average—than non-CTA teachers of lower-achieving and non-gifted students. However, average time on inquiry-based activities was predicted as highest overall for CTA teachers of higher-achieving and gifted students.

Qualitative findings from focus groups suggest that CTA teachers may have engaged in more lecture-based and fewer inquiry-based activities when teachers perceived that students could not tackle the large amount of reading in the curriculum and more open-ended activities without some explicit demonstration, practice on traditional problems, and teacher guidance. These data cannot provide insight into whether teachers' perceptions of their students were accurate or whether their decisions resulted in the best choices of instructional activities for these students.

One limitation of this research is our dependence on teacher survey data. Such data is susceptible to response biases, including that teachers may have overestimated time spent on more inquiry-based, "reform-oriented" practices and underestimated time on more traditional lecture activities (Banilower, Smith & Weiss, 2002; Mayer, 1999; Mullens & Kasprzyk, 1998). However, our finding that—on one hand—teachers of students with disabilities and CTA teachers of lower-achieving students reported more teacher talk activities and—on the other—that teachers of lower-scoring students and fewer gifted students reported less student thinking activities suggests that response bias cannot mask all actual differences in reported time on lecture-based and inquiry-based activities for teachers of lower- versus higher-achieving students. An additional limitation of our study is the presence of more low-achieving students in the

Table 3. Hierarchical Linear Models on Teacher Talk and Student Thinking

	Teacher Talk Activities		Student Thinking Activities	
	Model 1 Beta (SE)	Model 2 Beta (SE)	Model 1 Beta (SE)	Model 2 Beta (SE)
Intercept	3.42 (.10)***	3.30 (.10)***	2.34 (.12)***	2.23 (.13)***
CTA teacher	-.37 (.10)***	-.21 (.11)*	.44 (.13)**	.44 (.14)**
Pre-test score	.08 (.08)		.22 (.10)*	
% gifted students		.41 (.23)		.70 (.27)**
% students with disability		.78 (.21)***		.39 (.24)
CTA teacher x Pre-test score	-.37 (.14)**		-.11 (.16)	
CTA teacher x % gifted		-1.16 (.36)**		-.39 (.42)
Algebra PD hours (log)	.02 (.03)	.03 (.03)	.09 (.03)**	.09 (.03)**
Intercept for teacher (π_{00k})	.30	.28	.36	.37
Intercept for school (γ_{000})	.06	.06	.15	.20
Residual	.28	.27	.31	.29

***p<.001; **p<.01; *p<.05

CTA versus non-CTA condition. However, results in models with percentages of gifted and disabled students—rather than pre-test achievement—indicated similar patterns to models including pre-test achievement, which suggests that the imbalances in pre-test scores for the CTA versus non-CTA condition did not affect our findings.

Cognitive Tutor Algebra curriculum thus likely influences what teachers do. However, lower-achieving students may have needs that are not addressed by the CTA curriculum. Specifically, teachers of lowest-achieving students may integrate more Teacher Talk and less Student Thinking because students need more adaptations, scaffolding, and explanations than the curriculum provides or because teachers need more support to lead complex, facilitative discussions with lower-achieving students. These findings thus imply that districts should think carefully about how to implement Cognitive Tutor Algebra and—potentially—any inquiry-based curricula in order to support teachers of students with the highest needs and least preparation.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A070185to the RAND Corporation and through Grant R305B1000012 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Angrist, J. D. & Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, 94: 1613–1634.

Authors. (2011)

Baker, S., Gersten, R. & Lee, D.-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *Elementary School Journal*, 103(1): 51–73.

Ball, D. L. (2001). Teaching with respect to mathematics and students. *Beyond classical pedagogy: Teaching elementary school mathematics*. B. S. Nelson and J. Warfield. Mahwah, NJ: Lawrence Erlbaum.

Bates, D., Maechler, M. & Bolker, B. (2011). *Lme4: Linear mixed-effects models using s4 classes*: Available at http://cran.r-project.org/web/packages/lme4/index.html.

Baxter, J., Woodward, J. & Olson, D. (2001). Effects of reform-based mathematics instruction on low achievers in five third-grade classrooms. *The Elementary School Journal*, 101(5): 529–547.

Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education*, 33(4): 239–258.

Briars, D. J. & Resnick, L. B. (2000). *Standards, assessment—and what else? The essential elements of standards-based school improvement*. Los Angeles: Center for the Study of Evaluation.

Burke, M. & Sass, T. (2008). *Classroom peer effects and student achievement, Working Paper 18*. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Cognitive Tutor Algebra I Development Team (2007). *Algebra I teacher's implementation guide, Volume 1*. Pittsburgh, PA: Carnegie Learning, Inc.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12(3): 311–329.

Cooley, J. (2009). Can achievement peer effect estimates inform policy? A view from inside the black box. *WCER Working Paper no 2010-3*.

Hanushek, E. A., Kain, J. F., Markham, J. M. & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics: Special Issue on Empirical Analysis of Social Interactions*, 18: 527–544.

Hoxby, C. M. (2000). *Peer effects in the classroom: Learning from gender and race variation*. National Bureau of Economic Research Working Paper No. 7867.

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics*, 61: 458–495.

Knapp, M., Shields, P. M. & Turnbull, B. J. (1995). Academic challenge in high-poverty classrooms. *Phi Delta Kappan*, 76(10): 770–777.

Lefgren, L. (2004). Educational peer effects and the Chicago Public Schools. *Journal of Urban Economics*, 56: 169–191.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Nystrand, M. & Gamoran, A. (1988). *A study of instruction as discourse*. Madison, WI: Wisconsin Center for Education Research.

Oakes, J., Gamoran, A. & Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes and meanings. *Handbook of research on curriculum*. P. W. Jackson. New York: MacMillan**:** 570–608.

Oakes, J., Ormseth, T., Bell, R. & Camp, P. (1990). *Multiplying inequalities: The effects of race, social class and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: RAND.

R Development Core Team (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, available at http://www.R-project.org.

Silver, E. A. & Stein, M. K. (1996). The QUASAR project: The "revolution of the possible" in mathematics instructional reform in urban middle schools. *Urban Education*, 30(4): 476–521.

Vigdor, J. & Nechyba, T. (2004). *Peer effects in North Carolina Public Schools*. NBER Working Paper.

Woodward, J. & Brown, C. (2006). Meeting the curricular needs of academically low-achieving students in middle grade mathematics. *The Journal of Special Education*, 40(3): 151–159.