Technical Report: Analysis of Effects of MathForward^{TM} on Student Achievement in Richardson Independent School District (RISD)

William R. Penuel
SRI International

Context for the Technical Report

The Texas Instruments (TI) MathForward^{TM} program is a systemic mathematics reform initiative aimed at improving achievement in middle and high school mathematics and algebra. The program is intended to provide students with increased instructional time in mathematics during which they experience challenging instruction supported by the TI-Navigator™ technology. The systemic supports include the creation of 100-minute blocks of mathematics, intensive professional development for teachers, and the establishment of common planning time to build school-based professional learning communities. In 2007–08, the program had nearly 140 teachers in 35 schools from 11 districts. During this past year, SRI International has been conducting an evaluation of the program, focusing both on program implementation and impacts on student achievement for each district in the program.

The current report presents analyses comparing student achievement for students participating in the MathForward program with students not in the program in Richardson Independent School District (RISD) in Texas. The report answers the question, "How do the achievement gains of students in the program compare with gains made by students not in the program?" Data from the Texas Assessment of Knowledge and Skill (TAKS) from spring 2007 and spring 2008 are the basis for the analysis. Background data on student grade, ethnicity, socioeconomic status, home language, gender, and special education status were also available for analysis. We fit separate models of effects of the program on achievement gains for each grade level, as well as an overall model of effects. Limitations of these data are that assignment of students to condition was not random and the comparison groups were not similar with respect to prior achievement. The concluding section of the report presents possible interpretations of the results, given these limitations, and suggests directions for future research.

The MathForward Program and Its Educational Significance

The Texas Instruments (TI) MathForward program is a systemic mathematics reform initiative aimed at improving achievement in middle school mathematics and algebra. The catalyst for reform has been the introduction of the TI-Navigator™ Classroom Learning System, TI-84, and TI-73 Explorer™ graphing calculators, which are intended to increase student participation and engagement in significant mathematics content and to inform both teachers and students about what they know and can do. In addition to the introduction of technology to enhance teaching and learning in the classroom, the MathForward program calls for strong administrator involvement and support, deep and ongoing teacher professional development, increased time for mathematics learning, and higher expectations of all students.

The core components of the program draw on scientific research that has addressed how best to meet the goals of MathForward. Classroom network technologies like the TI-Navigator™ Classroom Learning System have shown the potential for increasing students’ participation in class and their conceptual learning (Penuel, Roschelle, & Abrahamson, 2005). Professional development that is of an extended duration, focuses on content and classroom practice, and provides opportunities for teachers to learn from one another has been linked to increased teacher knowledge and improvements in classroom instruction (Desimone, Porter, Garet, Yoon, & Birman, 2002). When students spend more time working on complex assignments in mathematics, they learn more (Rowan, Correnti, & Miller, 2002). Likewise, heightened teacher expectations for students and teachers’ perceptions of greater administrative support have been linked to successful teaching and learning outcomes (Jussim & Eccles, 1992; O'Donnell & White, 2005; Rowan, Chiang, & Miller, 1997).

Three components of MathForward make it a systemic reform program with potential for broad impact. First, the program promotes alignment of professional development, curriculum, and assessment—a core feature of systemic reform (Knapp, 1997; Spillane & Jennings, 1997). Second, the program targets whole districts and schools for participation rather than individual teachers. Third, MathForward calls for changes not only inside the classroom, but also in how schools are organized to support mathematics learning. In these respects, MathForward is like certain states' efforts to promote systemic reform in mathematics and science; lessons from studies of implementation of these efforts (e.g., Penuel, Fishman, Gallagher, Korbak, & Lopez-Prado, 2008) regarding the challenges of transforming instruction in a systemic context are thus applicable to MathForward.

MathForward and Richardson Independent School District

RISD is an important district for Texas Instruments, since the district was where TI co-developed the MathForward program with district leaders. The program began there in the 2005-06 school year; by 2007-08, the year of the current study, the program was in its third year of implementation in the district. Early on, the components implemented came to be called “The Richardson Model” and now comprise the core elements of the overall program model.

Since its inception, RISD has sought to serve students who are underperforming as measured by state achievement test results. In the first year of the program, 100 percent of the participants in the program had not met the proficiency standards set by the state and measured by the Texas Assessment of Knowledge and Skills. In 2006-07, some of the students participating had met those proficiency standards but the majority did not. In the 2007-08 school year, 56.3 percent of participating students had not met proficiency standards in mathematics in the prior year.

Past research on RISD has shown the promise of the program. An analysis of achievement gains made in the first year of the program compared program students to two different comparison groups, matched with respect to student backgrounds, found that MathForward students made significantly greater gains than those students. These students differed significantly with respect to prior achievement; a modified regression discontinuity analysis provided some evidence that the greater gains of program students were not attributable solely to differences in initial achievement (Alexander & Stroup, 2007). Analysis of second year results indicated significant growth of students in the program relative to two different sets of comparison groups (Stroup, Pham, & Alexander, 2007). In their first analysis, gains of students in the program who had not met proficiency standards on the TAKS the previous year were compared to students who were not in the program and who also had not met proficiency standards. In their second analysis, a group of students in the program who had met proficiency standards were compared to similar students who were not in the program. In both analyses, program students outperformed those students most similar to them in prior achievement and student background.

There are important limitations in the prior research that stem from the method of assignment of students to the program. First, statistical tests of the similarity between program and comparison students with respect to prior achievement were not reported; in some comparison groups, the differences appeared quite large. These differences may be due to the method of assignment of students to the program, which systematically favored students with a particular achievement profile. Second, as the program grew, the method of assignment became more complex. The researchers attempted to handle the complexity by using a regression discontinuity design to analyze data for students falling below the proficiency cutoff who were assigned to the program. The attempt to apply this method to these data makes intuitive sense, but a standard assumption of evaluation practice regarding regression discontinuity is violated, that a single cutoff point is used in every context to assign students to the treatment condition (see Shadish, Cook, & Campbell, 2002).

The current research study approaches the analysis of achievement results using gains as outcomes, but it does not address the threats to internal validity posed by how RISD assigned students to the program. An analysis using gain scores is appropriate when assignment is based on a set of latent factors but not on a single cutoff score (Maris, 1998), as it appears to be in RISD. In addition, as previous researchers did, this study attempts to mitigate threats to selection bias by employing statistical controls for background variables that may have affected selection into the program. However, it is important to note that the region of overlap between program and comparison students with respect to their probability of being assigned to the program is so small as to make it impossible to say that the comparisons are unbiased estimators of impact. Instead, the analyses allow us only to analyze whether the gains made by program students were significantly different from those made by students in the same grade levels but who were not in the program, for whatever reason.

Methods

The study used a pre-post nonequivalent comparison group design to analyze effects on student achievement. The analysis focused on all 7^{th}, 8^{th}, and 9^{th} grade students in the district in the program and compared them to all students not in the program in those grades in participating schools. To account for nesting of students within schools, we fit a two-level hierarchical linear model to the data, using student gains in scale scores on the TAKS mathematics test from 2007 to 2008 as the outcome measure. Implementation data reported as evidence of fidelity to the model come from surveys SRI administered to teachers in fall 2007, winter 2008, and spring 2008.

Participants

Richardson Independent School District located in a suburban community of just under 100,000 residents in the Dallas (TX) metropolitan area. The district is ethnically diverse, with large numbers of White, African American, and Hispanic students. Just over a third of students in participating schools were eligible for free or reduced price lunches. Relative to other districts in the state, average mathematics scores in Richardson are above the state average. A total of 13 schools in the district participated in MathForward in 2007-08: 8 were middle schools, 4 were high schools, and 1 was a center serving freshmen. For the district, it was the third year of participation in the program.

A total of 52 teachers participated in the program in 2007-08. Of these teachers, 17 taught 7th grade, 16 taught 8th grade, and 4 taught both 7th and 8th grade. Fifteen taught in high schools, and of these, 10 taught Algebra I. On average, there were 4 teachers per school teaching in the program, with a range from 2 teachers to 6 teachers.

The participants for the study include all students in grades 7, 8, and 9 in the participating schools. The method of assignment to the program was on the basis of student need, which included consideration of the prior achievement of students in the program. As the program has scaled, it has expanded the pool of potential participants from students who failed to meet proficiency standards on the prior year's state mathematics test to include a wider range of factors. The participants were not intended to be representative of a larger group; however, the district's ethnic and socioeconomic diversity make it a particularly instructive context for analyzing impacts of the program for students from different backgrounds.

Descriptive statistics for the study sample appear below in Table 1. Robert Sorenson of RISD provided SRI with anonymous individual level data files that were the source for these statistics.

Table 1. Characteristics of Students in the Study Sample

	MathForward Students	Comparison Students
Grade		
7th	254 (31.1%)	1630 (33.5%)
8th	358 (43.8%)	1600 (32.9%)
9th	205 (25.1%)	1641 (33.7%)
Gender		
Female	373 (45.7%)	2460 (50.5%)
Male	444 (54.3%)	2411 (49.5%)

Table 1 (Cont'd)

	MathForward Students	Comparison Students
Ethnicity		
African American	318 (38.9%)	940 (19.3%)
Native American	1 (0.1%)	26 (0.5%)
Asian American	26 (3.2%)	444 (9.1%)
Hispanic	351 (43.0%)	1199 (24.6%)
White	121 (14.8%)	2262 (46.4%)
ELL Status		
Fluent English Speaker	670 (82.0%)	4639 (94.2%)
English Language Learner	147 (18.0%)	232 (4.8%)
Socioeconomic Status		
Eligible for Free or Reduced Price Lunch	555 (67.9%)	1730 (35.5%)

With respect to each of the above characteristics, with two exceptions, MathForward students were significantly different from comparison students. The two exceptions were that the proportion of the sample comprised of 7th graders was similar and the proportion of Native American students was similar for the two groups. At the same time, there was only 1 Native American student in the MathForward sample.

The Program as Implemented in RISD

Data from the teacher surveys conducted to document implementation across districts provide the basis for how the program was implemented in RISD. A full description of the measures and overview of implementation across districts can be found in Penuel, Ferguson et al. (2008). The program as implemented was largely congruent with the overall program model; however, although teachers made regular use of TI-Navigator tools with students, they reported only "sometimes" using interactive pedagogies in conjunction with these tools. Past research has found such pedagogies to be critical ingredients to effective use of network technologies (see, especially, Judson & Sawada, 2002).

With respect to the supporting conditions for implementation, RISD's implementation was congruent with the program model. As intended, all students in the program were assigned to double blocks of mathematics. In addition, all teachers participated in professional development related to TI-Navigator use in the classroom. Roughly twothirds had participated in the content-related professional development TI provided to selected districts in the program. School and district leaders were supportive of the program and of teachers' participation in it.

In the classroom, teachers in RISD implemented multiple TI-Navigator tools on a regular basis with students. As Table 2 indicates, they reported using Activity Center an average of 1 to 2 periods per day in spring and reported using Class Analysis and Screen Capture between 2 and 3 periods per day. Reported use of all Activity Center and Class Analysis tools was higher in Grade 7 than in the other grades.

Table 2. Teachers' Reported Use of TI Navigator Tools

	Activity Center	Class Analysis	Screen Capture
Grade 7	2.55	3.64	3.09
Grade 8	2.00	2.85	2.95
Grade 9	2.17	2.83	3.33

n = 32-33
Source: Spring 2008 Survey
1='Not at all' 2='1 period' 3='2 periods' 4='3 periods' 5='4 periods' 6='5 or more periods'

As Table 3 indicates, teachers on average "sometimes" asked students to discuss their answers with peers before or after answering a question using the TI-Navigator system. Teachers were somewhat less likely to ask students to re-answer after peer discussion or a teacher presentation of an explanation. Teachers reported doing these activities between "hardly ever/never" and "sometimes" on average. Teachers reported they sometimes facilitated whole-class discussion of student ideas when using TI-Navigator.

Table 3. Facilitated Peer and Group Discussion as Part of TI-Navigator Use

	Ask students to discuss with a peer before answering	Ask students to discuss with a peer after answering	Ask students to re-answer after discussing with a peer	Ask students to re-answer after a teacher explanation	Facilitate a whole-class discussion of students' ideas
Grade 7	2.00	2.00	1.91	2.09	2.40
Grade 8	1.85	1.77	1.62	1.54	2.08
Grade 9	2.00	2.00	1.40	1.60	2.00

n = 32-33
Source: Spring 2008 Survey
Scale: 1='Hardly ever/never' 2='Sometimes' 3='Most of the time' 4='Nearly every time'

As Table 4 indicates, across all three grades, teachers reported that they used feedback sometimes to adjust instruction, with the exception of Grade 9 teachers. Grade 9 teachers on average reported hardly ever/never making use of feedback to adjust instruction for the next class session.

Table 4. Feedback and Adjustment of Instruction

	Mean Use of Feedback to Adjust Instruction *During Class*	Mean Use of Feedback to Adjust Instruction *for the Next Class Session*
Grade 7	1.73	1.90
Grade 8	1.77	1.62
Grade 9	1.80	1.20

n = 32-33
Source: Spring 2008 Survey
Scale: 1='Hardly ever/never' 2='Sometimes' 3='Most of the time' 4='Nearly every time'

Composition of the Comparison Group

All students in the district for whom there were 2007 and 2008 scores but who were not in the program were ultimately included in the comparison group, despite the differences between the two groups both with respect to student background characteristics and prior achievement (see Table 5 below).

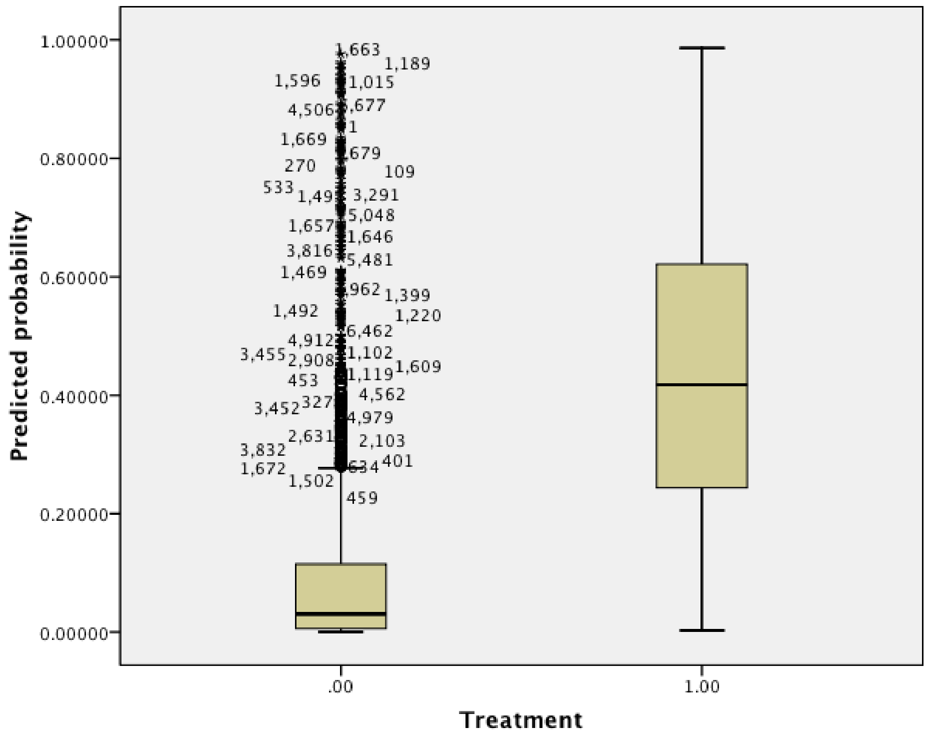
Table 5. Prior Achievement (TAKS Scale Score) of Students in the Study Sample

Condition	N	Mean Scale Score	Standard Deviation
MathForward™	817	2095.06***	113.89
Comparison	4871	2388.79	211.42

*** $p < .0005$

The approaches used in prior years to creating similar comparison groups were not possible for 2007-08, due to the change in the way students were assigned to the program. The program expanded in 2007 and 2008 to include many more students, since the district judged it to be a successful program. We attempted to use propensity score matching to develop comparison groups that were similar to the treatment group with respect to prior achievement and student background. However, although we were able to identify characteristics associated with the likelihood of being assigned to the program, we were not able to achieve groups that were similar even when using weighting techniques (Frank et al., 2008). Figure 1 below shows the limited overlap between the treatment and comparison groups with respect to their propensity scores.

Figure 1. Distribution of Probabilities of Being Assigned to MathForward



Key: MF = 1, Comparison = 0

Prior analysts sought to implement a regression discontinuity analysis of Richardson data, given the method of assignment to condition. However, such an approach was not appropriate given our modeling results, since there were multiple predictors of propensity to be assigned to the treatment, some of which could not have been used consciously by educators to assign students to the program. In such circumstances, it is more appropriate to use gain scores as the outcome measure and include statistical controls for background characteristics that could affect the results.

A concern with such an approach is the potential for regression of scores to the mean, given the distribution of scores in the data. MathForward students may gain more, simply because they started out with lower scores on average. Comparison students may not be able to gain as much, since their initial scores were, on average, much higher. It would be desirable in future years to either (1) assign students to MathForward or comparison classrooms on the basis of a specific cutoff TAKS score, permitting regression discontinuity analysis; or (2) randomly assigning some students to MathForward and comparison classrooms for purposes of experimental analysis. If this is not feasible, then attempts to estimate impact without potential for bias will continue to be impossible.

Comparing Students with One Versus Two Years' Data

In longitudinal studies, data loss is inevitable and poses a threat to the validity of inferences. For our dataset, we have 2008 scores for all students, but a total of 1,086 students' 2007 scores are missing. Students whose scores are missing had lower scores on average than those for whom we obtained matched data. Scores of students with missing data had mean mathematics scale scores of 2142 (*SD* = 214) on the outcome measure, which were significantly lower than the scores of students for whom we had matched data (*M* = 2317, *SD* = 221).

Outcome Measure

The outcome measure used for the analysis is gains from 2007 to 2008 on the mathematics portion of the Texas Assessment of Knowledge and Skills. This test is given in the spring of each school year to all eligible students. The scale scores from year to year on this test are intended to be comparable; however, no test equating has been performed to assess the validity of the approach, and the construction of the test is such that items do not repeat from year to year. In addition, the standards tested vary by grade level, so the content for the tests is not the same. Hence, we report here on overall results, which should be interpreted with caution, and grade-by-grade analyses, in part to account for possible differences between test difficulties across grades.

Approach to Analysis

Because we had school, but not classroom, assignment information for each student, we fit a 2-level HLM model to the data. Our outcome for the analysis was gain scores calculated by subtracting the 2007 mathematics TAKS scale score from the 2008 scale score. Preliminary analyses of an unconditional model using gains as outcomes indicated that there was a significant proportion of variance (Table 6) at the school level to warrant modeling the data in this way ($\chi^2 = 452.70$, df = 11, $p < .001$).

Table 6. Variance Components

	Variance Component	Percent
Level 1 (Student)	26160.51	91.7
Level 2 (School)	2353.67	8.3
Total	28514.18	100

Predictors included in initial models were a treatment variable, ethnicity, ELL status (binary), socioeconomic status (binary), gender, and special education designation (binary). We removed variables that were nonsignificant, and for any predictors that were significant, we created interaction terms to test for interactions with the treatment. The final model, for which overall results are reported here was:

LEVEL 1 MODEL

$$GAIN0708_{ij} = \beta_{0j} + \beta_{1j}(ETH\_AA_{ij} - \overline{ETH\_AA}_{..}) + \beta_{2j}(ETH\_HI_{ij} - \overline{ETH\_HI}_{..}) + \beta_{3j}(TIMF_{ij}) + r_{ij}$$

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

In which the coefficient for the treatment effect is $\beta_{3j}(TIMF_{ij})$, and the effect of being African American on gain scores is represented as $\beta_{1j}(ETH\_AA_{ij} – ETHN\_AA_{..})$ and the effect of being Hispanic on gain scores is $\beta_{2j}(ETH\_HI_{ij} – ETHN\_HI_{..})$. Subsequent models by grade include each of these background variables.

Results

Overall Model Results

The model results (Table 7) indicated a significant main effect for treatment and for being African American or Hispanic. The average change in score for a White or Asian student in the comparison group was -38.4 scale score points, indicating a significant drop in scores from 2007 to 2008 overall. By contrast, the overall treatment effect, controlling for ethnicity, was 67 scale score points improvement from 2007 to 2008. However, there were no differential treatment effects for either African American students or Hispanic students, and no other demographic variables were significant predictors of gains.

Table 7. Overall Model Results for Analysis of TAKS Gain Scores

Variable	Coefficient	Standard Error
Intercept	-38.36*	14.24
Treatment	67.21***	6.35
African American	-40.78***	5.18
Hispanic	-39.13***	5.75

* $p < .05$, *** $p < .001$

Grade Level Model Results

The model results (Table 8) indicated a significant positive main effect for treatment in 7th grade. The effect for 8th grade did not reach significance; however, using a 1-tailed *t*-test of significance, the coefficient could be interpreted to show a significant positive association between treatment assignment and achievement gains. At the same time, there was a significant negative main effect for treatment in 9th grade. In grades 8 and 9, achievement scores were lower for African American and Hispanic students, regardless of whether they were in the MathForward program or in the comparison group.

In addition to the main effects of the program on achievement gains, we examined interaction effects to determine whether achievement gaps were closing for MathForward students relative to other students in the district. In 9th grade, there was an interaction between treatment assignment and ethnicity: Hispanic students assigned to the program were more likely to make gains than Hispanics not assigned to the programs (p = .06). These results indicate some closing of the achievement gap for Hispanics in 9th grade. No other interactions were significant, indicating no evidence from this study that MathForward is closing achievement gaps for African American or Hispanic students in grades 7 or 8. Only longitudinal analysis can adequately assess whether gaps are in fact closing, given the instability in estimates of gain scores from year to year.

Table 8. Grade-by-Grade Model Results for Analysis of TAKS Gain Scores

	7th Grade (n = 1,884)		8th Grade (n = 1,958)		9th Grade (n = 1,846)	
Variable	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Intercept	163.99***	9.13	13.61**	3.60	38.79†	11.59
Treatment	182.41***	10.79	13.64†	7.71	-34.54**	11.47
African American	-6.34	9.65	-29.31***	7.59	-61.46***	9.06
Hispanic	-23.33	8.47	-23.80***	6.88	-60.66***	8.67

† $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

The overall model results indicated that MathForward participation was associated with significantly higher gains in mathematics achievement, when compared to students in the district not in the program. Grade-by-grade analyses suggest that gains associated with being assigned to the program were primarily in Grades 7 and Grade 8. Furthermore, except for Hispanics in 9th grade, there was no evidence from this study that achievement gaps were closing for students in the program, relative to students not in the program.

In grade 7, implementation rates were higher than in the other grades, suggesting one possible explanation for this pattern of results. Teachers made more frequent use of the most powerful TI-Navigator tools in this grade level, and 7th grade teachers also were more likely to adjust their instruction on the basis of formative assessment data collected using TI-Navigator than were teachers in other grades. The fact that the treatment effect was strongest in this grade suggests that the achievement gains are at least partly attributable to the program.

These results are suggestive of the *promise* of the intervention, but the models tested here do not permit us to conclude that we have unbiased estimates of program impact. There were significant differences between the two groups with respect to both student background and prior achievement, and propensity score matching did not yield groups with enough overlap to create a matched comparison group. The use of gain scores can mitigate potential effects of differences, but the fact that program students had more room to grow may have affected the results. Thus, we cannot conclude that the significant gains observed in this study were caused by the program.

Because of the threats to internal validity, there are limits to both generalizability and potential significance for policymakers beyond RISD. We know little about how the achievement gains of low-performing students compared to those of students with similar profiles as program students, since so many students in the district participated that a matched comparison group was impossible to construct. Comparison groups from outside the district may yield better estimates of impact in future years, but these students may not share enough of the same policy and district context to yield valid results.

The limitations of the study do not prohibit either TI or RISD from drawing lessons from the study. The relationship between implementation and gains suggests the promise of the program for high-implementing classrooms; it also suggests the need to understand how to support such implementation in the future. For RISD, the gains made by students are confirmation of its policy and approach: MathForward students *are* making significant gains, at least in 7^{th} and 8^{th} grade. The poor results in 9^{th} grade suggest, furthermore, that the district take a closer look to uncover why these results were not as strong as for the other two grades.

References

Alexander, C., & Stroup, W. (2007). *Richardson Math Project final report: Math TAKS results*. Austin, TX: University of Texas.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(2), 81-112.

Frank, K. A., Sykes, G., Anagnostopoulous, D., Cannata, M., Chard, L., Krause, A., et al. (2008). Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters? *Educational Evaluation and Policy Analysis, 30*(1), 3-30.

Judson, E., & Sawada, D. (2002). Learning from past and present: Electronic response systems in college lecture halls. *Journal of Computers in Mathematics and Science Teaching, 21*(2), 167-181.

Jussim, L., & Eccles, J. S. (1992). Teacher expectations: Construction and reflection of student achievement. *Journal of Personality and Social Psychology, 63*(6), 947-961.

Knapp, M. S. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research, 76*(2), 227-266.

Maris, E. (1998). Covariance adjustment versus gain score - revisited. *Psychological Methods, 3*, 309-327.

O'Donnell, R. J., & White, G. P. (2005). Within the accountability era: Principals' instructional leadership behaviors and student achievement. *NASSP Bulletin, 89*, 56-71.

Penuel, W. R., Ferguson, K., Singleton, C., Shea, S., Borelli, K., & Korbak, C. (2008). *MathForward implementation quality report*. Menlo Park, CA: SRI International.

Penuel, W. R., Fishman, B. J., Gallagher, L. P., Korbak, C., & Lopez-Prado, B. (2008). The mediating role of coherence in curriculum implementation. In *Proceedings of the 8th International Conference of the Learning Sciences*. Utrecht, the Netherlands: Erlbaum.

Penuel, W. R., Roschelle, J., & Abrahamson, A. L. (2005). Research on classroom networks for whole-class activities. In *Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education* (pp. 222-229). Los Alamitos, CA: IEEE.

Rowan, B., Chiang, F., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on student' achievement. *Sociology of Education, 70*, 256-284.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record, 104*(8), 1525-1567.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Spillane, J. P., & Jennings, N. E. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *Teachers College Record*, *98*, 449-481.

Stroup, W., Pham, V., & Alexander, C. (2007). *Richardson MathForward Project second year final report: Math TAKS results*. Austin, TX: University of Texas.