Teacher Leaders for Mathematics Success (TL=MS)

Final Evaluation Report

Teacher Leaders for Mathematics Success (TL=MS)

Final Evaluation Report

Cheri Fancsali, Ph.D.

Submitted to

Institute for Literacy Studies

Lehman College

City University of New York

by

Academy for Educational Development

New York City

2004

Academy for Educational Development

Acknowledgements

Many individuals contributed to this evaluation and report. Thanks are due to Nancy Nevárez and Kari Nelsestuen who conducted research for the evaluation. Ann Cola also conducted interviews and observations. Dr. Sandra Langley and Jodi Paroff conducted data analyses and helped write the report. Dr. Ann O'Connell conducted the multilevel analyses and wrote the summary of results from those analyses. Drs. Langley and O'Connell also reviewed the report and provided helpful feedback. Eliana Orellana provided research support, Wanda Dallas entered the teacher survey data, and Noemi Corujo provided administrative assistance. Elayne Archer edited the report. Dr. Alexandra Weinbaum served as the senior technical advisor to the evaluation project.

Thanks are also due to all the staff and teacher consultants of the Teacher Leaders for Mathematics Success (TL=MS) project for their assistance and support with data collection and feedback on the findings. I am especially grateful to all the TL=MS participants and students who participated in this evaluation; they provided valuable information on TL=MS, which will advance the field's understanding of effective teacher professional models and their impact on student achievement.

Cheri Fancsali, Ph.D., Senior Program Officer
Academy for Educational Development
Center for School and Community Services
New York City
Fall 2004

What Teachers Say about TL=MS

Before TL=MS I was much more structured. Now, I let students explore more, and I use multiple approaches to help them learn.

I've learned how to expand lessons, and I pay a lot more attention to students. I spend a lot more time listening to what they are saying.

I try to create an atmosphere where it is okay to be wrong. I want to enrich their experiences in math and make sure they've participated inside the process and learned from it.

If you understand the content, it's easier to teach and give multiple ways of presenting a concept. Otherwise, you're stuck teaching only the way it is presented in the textbook.

[Through TL=MS], *I've seen so many different people solve things in different ways. That helps me as a teacher because I am able to explain it in more ways to students and to understand what their perspective might be.*

The most helpful aspect of TL=MS project was having a consultant visit and model engaging lessons using manipulatives, games, etc. I also learned a lot from the consultant modeling how to question the students on their mathematical thinking.

The most helpful aspect of the TL=MS project was the visits by the consultants in the classroom. It allowed time to reflect on practices and build a shift in my mathematical philosophy. It gave me a chance to conference and assess my own technique.

[In the study groups] *ideas were shared. It was a positive thing to find that others were experiencing similar problems/success.*

TL=MS taught me how to keep an open mind while teaching math and how to turn-key information to other teachers.

Not only have I become a more effective teacher of mathematics, but also a more effective staff developer.

Participation changed my professional life—I went from class teacher to staff developer. Now I have applied for math coaching jobs . . . I return with a whole new approach.

Table of Contents

Table of Tables and Figures

Executive Summary

TL=MS has definitely changed my thinking. First, I used to do what I thought was hands-on. But, actually, I was very much regimented and worksheet-based. I try to let the students lead the instruction now. I used to have to stick to my lesson plan, but now I am able to see what the students need and what they are interested in and remain flexible. (TL=MS teacher participant)

I used to hear kids say, "I hate math!" but I don't anymore. The kids ask me where I was when I miss a day. In one class, they cheer when I walk in the door! (TL=MS teacher participant)

Introduction

Teacher Leaders for Mathematics Success (TL=MS*)* is a five-year project designed to build the capacity of Bronx teachers and schools in supporting continued improvement in mathematics education for all students in a standards-based environment. The project, implemented by the Institute for Literacy Studies at Lehman College and funded by the National Science Foundation, seeks fundamental educational change by enhancing the understanding of mathematics content, standards-based curriculum, and performance standards, as well as student-learning strategies among teachers, principals, and other administrators.

The project facilitates discourse about and reflection on the relationships between content knowledge, pedagogy, student learning, and school change. Its goal is to create conditions for institutionalizing teacher leaders as agents for instructional reform in mathematics within schools and districts. It is founded on the notion that the "effectiveness of mathematics teaching and learning is a function of teachers' knowledge and use of mathematical content, of teachers' attention to and work with students, and of students' engagement in and use of mathematical tasks" (National Research Council, 2001).

Project Description

Working with three cohorts of approximately 20 schools and 80 teacher and administrator participants each, the project is organized around three levels of activities across three years for each participating cohort.

Level one immerses participants in an intensive study of mathematics topics, aligned with standards-based curricula implemented in the schools, and their relationship to performance standards and student learning. During this first year, all participants are asked to attend a 60-hour summer institute. Once the school year begins, participants attend monthly Saturday seminars (eight Saturdays for six hours each) and work with a teacher consultant on a biweekly basis. The teacher consultant provides a range of services, including meeting with, and conducting observations of, teachers, facilitating team meetings, coteaching classes, and assisting in lesson and project planning. The teacher consultants also provide support to school administrators and the district by participating in meetings, facilitating discussions, and conducting workshops. Through this immersion, participants enhance their understanding of mathematical concepts as well as develop effective strategies to teach these concepts in the classroom.

In level two, during the second year, TL=MS focuses on curriculum and leadership development as well as the development of a mathematics "leadership action plan" for the school. Participants continue to attend professional development sessions on Saturdays and after school, and teacher consultants continue to visit the school, although less frequently. During this second year, TL=MS participants also involve other teachers and administrators in their school in mathematics reform. In level three, the third and final year of each cohort's involvement, participants focus on implementing their school leadership plan and sustaining school-based leadership.

TL=MS Evaluation

The Academy for Educational Development (AED) conducted a five-year formative and summative evaluation of TL=MS. The evaluation addressed research questions related to program participation; outcomes for teachers, students, and schools; and district and administrative support for the project. These questions were investigated through multiple data collection methods including surveys of participants, in-depth interviews with administrators and project teachers at selected schools, classroom observations at selected schools, and analysis of pre-post student assessment data from 15 classrooms.

This report presents findings from all five years of the evaluation but focuses on results from the final teacher survey and pre-post student assessment data. (See AED's previous reports on TL=MS for a discussion of findings from earlier years.) Findings were triangulated with results from earlier data collection efforts.

Key Findings on Impact on Teachers and Other Participants

Major findings are summarized below.

- 80% or more of survey respondents reported that TL=MS had a great deal or good amount of impact on their 1) understanding of math content; 2) comfort level teaching math; 3) effectiveness as a teacher; and 4) teaching practices.
- High levels of impact in these areas were reported by teachers regardless of different grade levels taught, years of experience, and project cohort.

> *It is impossible to help children come to any understanding if we as teachers do not have deep understanding of the content.* (TL=MS participant)

- At the end of the project, TL=MS teachers reported very high levels of use of reform-based practices (such as solving math problems in small groups or with a partner). They also reported that their use of these practices increased as a result of their participation. For example, 90 percent or more said they asked students to explain how they arrived at their answers, had students discuss solutions to mathematics problems with other students, and used manipulatives at least once or twice a week. At the same time they reported decreasing their use of traditional classroom practices (such as lecturing and using worksheets) as a result of the project.

In summary, many TL=MS participants entered the project as self-described math "phobics" and with limited knowledge of mathematics content and limited understanding of constructivist approaches. They emerged from the project more confident and competent in their understanding of mathematics and their skills as mathematics teachers. Further, longitudinal survey data show that teachers—up to five years after first participating in TL=MS—continued frequent use of the reform-based approaches espoused by the project, indicating that TL=MS was sustained—at least at the classroom level.

Comparisons with a National Sample

In comparison to a nationally representative sample[1], TL=MS teachers were much more likely to use reform based practices than their peers and less likely to use traditional practices ($p \leq .01$). For example:

- Almost double the percentage of TL=MS teachers reported asking students to discuss solutions to math problems almost everyday compared with a national sample of teachers at the same grade level (86% vs. 44%).
- Over three times as many TL=MS teachers reported asking students to solve math problems in small groups or with a partner (80% vs. 26%) almost everyday.
- Over four times as many TL=MS teachers reported asking students to write a few sentences about how to solve a math problem almost everyday compared with teachers nationwide (59% vs. 14%).
- TL=MS teachers were less likely to report using multiple choice exams to assess student learning compared with teachers nationwide (12% vs. 20% reported using the practice 1-2 times a week).
- The majority of TL=MS teachers (88% or more) reported placing a heavy emphasis on developing an appreciation for the importance of math, learning how to communicate ideas in mathematics effectively, and developing reasoning and analytic ability to solve unique problems, areas that are consistent with a reform-based approach. Compared with teachers nationwide, TL=MS teachers were much more likely to report placing heavy emphasis on these areas ($p \leq .01$).
- The majority of TL=MS teachers (73% to 86%) also placed heavy emphasis on students' learning mathematics facts and concepts, as well as skills and procedures for solving routine problems, but to a much less extent than teachers nationally ($p \leq .01$). These findings indicate that TL=MS teachers practice reform-based practices to a much greater extent that most teachers nationally, while maintaining an emphasis on facts and concepts and the skills and procedures necessary to solve routine problems.

In summary these findings show that the impact of TL=MS on teachers was deep—affecting their comfort level with teaching mathematics as well as their content and pedagogical knowledge around teaching mathematics. Further, more than just increasing or decreasing their knowledge and their use of specific practices, teachers described changes in their whole approach to mathematics instruction.

Teachers attributed much of this shift to the powerful impact of the summer institute and the value of the work conducted by teacher consultants (TCs). TCs supported teachers and schools in a variety of ways, including conducting observations and providing individual feedback to teachers, modeling and demonstrating lessons, and participating in team meetings and planning

---

[1] The sample was provided by the National Assessment of Educational Progress (NAEP) survey of teachers. The NAEP is a nationally representative sample of students whose teachers took the survey.

sessions. In addition, many teachers spoke of the benefit of working and learning with a collegial group of TL=MS participants. TL=MS gave teachers a community and network through which they could share ideas, learn from each other, problem-solve, and support each other. These positive findings held for teachers at different levels of their career (novice, experienced and veteran), different grade levels, and for each of the projects' three cohorts.

Key Findings on Student Impact

The impact of TL=MS on students was equally impressive. The majority of teachers and administrators reported that they saw a great deal or good amount of impact on students' engagement in mathematics instruction and achievement. These findings were substantiated by an open-response, pre-post performance assessment completed by students in a sample of TL=MS and non-TL=MS classrooms.

TL=MS students saw large gains between pre- and post-assessments of their mathematical knowledge, strategic knowledge and mathematical communication.

> *Kids get extremely excited about math—it is the highlight of their day! Even those who are struggling with some of the math concepts still get to participate and "play." There are a lot of student-helping-student scenarios.* (TL=MS participant)

- Nearly nine out of 10 TL=MS students showed increases in their total scores on the assessment, and between 72% and 79% showed an increase in a specific sub-area (mathematical knowledge, strategic knowledge and mathematical communication).
- Multilevel analyses showed that, controlling for teacher characteristics and pretest scores, TL=MS classes gained an average of 3.29 points more (out of 12) than non-TL=MS students on the post-test. This is a very strong, statistically significant[2] finding, especially given the small sample size (12 TL=MS and 3 non-TL=MS classrooms).
- The positive effect of TL=MS held for a diverse group of low-income students, showing no differences by gender, race/ethnicity or English language learner status.

In conclusion, findings from multiple sources, including multilevel statistical analyses of pre-post student assessment data, converge to provide strong evidence that TL=MS provided teachers with a deep and lasting professional development experience that had an impact on teachers' practices and attitudes about teaching mathematics as well as on student achievement. This study is one of a small number of studies that have looked beyond impact of professional development on teachers to investigate the impact on student outcomes (Killion, 1998). Results from this study also support findings from other studies that have shown a connection between high-quality, sustained professional development opportunities, reform-based practices, and student achievement (Garet et al., 2001; Hamilton et al., 2004; Wenglinsky, 2000).

[2] ($p \leq .10$). A level of significance of .10 was used for identification of variable effects. The sample size of 15 teachers is small enough to warrant such a consideration; a strict .05 level of significance might mask some important tendencies or trends.

Introduction

> *I have definitely learned a lot* [through TL=MS] *and that learning has made me feel more and more comfortable teaching math. I've learned about the meaning, the "why" behind a lot of math. I've learned how there are patterns behind almost everything in math. Since I understand the why so much better, I can actually help students more because I can explain things in more than one way and understand why they can't understand*. (TL=MS teacher participant)

Teacher Leaders for Mathematics Success (TL=MS*)* is a five-year project designed to build the capacity of Bronx teachers and schools in supporting continued improvement in mathematics education for all students in a standards-based environment. The project, implemented by the Institute for Literacy Studies at Lehman College and funded by the National Science Foundation, seeks fundamental educational change by enhancing the understanding of mathematics content, standards-based curriculum, and performance standards, as well as student-learning strategies among teachers, principals, and other administrators.

The project facilitates discourse about and reflection on the relationships between content knowledge, pedagogy, student learning, and school change. Its goal is to create conditions for institutionalizing teacher leaders as agents for instructional reform in mathematics within schools and districts. It is founded on the notion that the "effectiveness of mathematics teaching and learning is a function of teachers' knowledge and use of mathematical content, of teachers' attention to and work with students, and of students' engagement in and use of mathematical tasks" (National Research Council, 2001). Working with three cohorts of approximately 20 schools and 80 teacher and administrator participants each, the project is organized around three levels of activities across three years for each participating cohort.

Level one immerses participants in an intensive study of mathematics topics, aligned with standards-based curricula implemented in the schools, and their relationship to performance standards and student learning. During this first year, all participants are asked to attend a 60-hour summer institute. Once the school year begins, participants attend monthly Saturday seminars (eight Saturdays for six hours each) and work with a teacher consultant on a biweekly basis. The teacher consultant provides a range of services, including meeting with, and conducting observations of, teachers, facilitating team meetings, coteaching classes, and assisting in lesson and project planning. The teacher consultants also provide support to school administrators and the district by participating in meetings, facilitating discussions, and conducting workshops. Through this immersion, participants enhance their understanding of mathematical concepts as well as develop effective strategies to teach these concepts in the classroom.

In level two, during the second year, TL=MS focuses on curriculum and leadership development as well as the development of a mathematics "leadership action plan" for the school. Participants continue to attend professional development sessions on Saturdays and after school, and teacher consultants continue to visit the school, although less frequently. During this second year, TL=MS participants also involve other teachers and administrators in their school in mathematics reform. In level three, the third and final year of each cohort's involvement, participants focus on implementing their school's leadership plan and sustaining school-based leadership.

To recruit schools and participants, TL=MS staff made presentations at principals' conferences in each Bronx district to outline program objectives and clarify criteria for nominating schools and team members. Schools were encouraged to nominate teams that included three teachers representing a mix of experienced and new teachers and a staff developer or administrator. Schools submitted applications to participate in the program to a steering committee comprised of district mathematics coordinators, district principals, the principal and co-investigators of the project, and Lehman College faculty. The committee selected schools for participation based on the following criteria: 1) school readiness, commitment to reform, and capacity; and 2) teacher, staff developer/administrator preparation and experience, in-service professional development related to nationally validated curriculum, and degree of exposure to standards-based curricula (Source: TL=MS project summary, Lehman College, undated). Participants received tuition-waved graduate credit or stipends for their involvement.

The Academy for Educational Development (AED) conducted a five-year formative and summative evaluation of TL=MS. The evaluation addressed research questions related to program participation; outcomes for teachers, students, and schools; and district and administrative support for the project. These questions were investigated through surveys of teachers, in-depth interviews with administrators and project teachers at selected schools, classroom observations at selected schools, and analysis of pre-post student assessment data from 15 classrooms.

This report presents findings from all five years of the evaluation but focuses on results from the final teacher survey and pre-post student assessment data. (See AED's previous reports on TL=MS for a discussion of findings from earlier years.) Findings were triangulated with results from earlier data collection efforts.

Evaluation Methodology

AED explored program participation and outcomes through multiple data collection methods. These included surveys of participating teachers from all three cohorts, site visits to a sample of participating TL=MS schools, and pre-post assessment of student performance on a mathematics task. Evaluation methods are summarized in the box below and described in the next section.

Evaluation Methods, Respondent size and Administration Periods

Method	Respondent size	Administration period
TL=MS participant pre-summer institute survey	All cohort-one (n=85) and two TL=MS participants (n=98)	Prior to attending the summer institute (cohort 1= 1999, cohort 2=2000)
TL=MS post-summer institute survey	All cohort-one TL=MS participants (n=62)	May 2000
Final teacher survey	All TL=MS participants (n=129)	Summer 2003 (mailed)

Method	Respondent size	Administration period
Site visits (included classroom observations and interviews with TL=MS teachers and administrators)	Five cohort-one schools (three elementary and 2 middle.) 2001: n=28 teacher/administrator interviews, 16 classroom observations; 2002: 19 teacher/administrator interviews, 14 classroom observations	2001 and 2002
Pre-post student assessment	12 TL=MS and 3 non-TL=MS 4th and 5th grade classrooms (n= 326 TL=MS and 74 non-TL=MS; students with both pre- and post-assessments =220 TL=MS and 58 non-TL=MS)	2002-03 (piloted in 2001-02)

Teacher Surveys

Teachers participating in TL=MS were asked to complete a total of three surveys. Teachers in cohorts one and two completed a survey before the summer institute—their first exposure to the project—and a follow-up survey at the end of their first year in the project. All three cohorts of teachers were asked to complete a final teacher survey in spring 2003. Each is described below.

Pre-Post Summer Institute Surveys

To establish a baseline measure of practices and to measure the early impact of TL=MS after one year, AED developed a pre-post survey instrument with the input of the project implementers. The pre-institute survey was administered to all cohort-one and -two participants before the summer institute—participants' first exposure to the project (summer 1999 for cohort one and 2000 for cohort two); the post-institute survey was administered at the end of the first year of implementation (May 2000) for cohort one. TL=MS staff administered each survey in person during a TL=MS event. The final teacher survey (described next) served as the post-survey for cohort two. A total of 85 cohort-one participants completed the pre-institute survey, and 62 completed the post-institute survey; 54 participants took both. A total of 98 cohort-two teachers completed the pre-institute survey, and 44 completed the post-survey (administered in spring 2003); 26 completed both.

Final Teacher Survey

The final teacher survey was mailed to every TL=MS participant in spring 2003 with several follow-up reminders. A total of 129 TL=MS participants completed the final teacher survey. This represents 58 percent of the total number of participants (223) and 63 percent of the participants for whom we had correct addresses, a relatively high response rate considering that the survey was administered and collected through the mail. Non-respondents were also telephoned and asked to complete the survey; 17 teachers for whom TL=MS did not have current addresses or telephone numbers could not be reached. The final teacher survey asked participants to report the frequency with which they used selected practices and to reflect on the project's impact on their use of these practices, as well as its impact on students. Survey items also asked teachers to evaluate the helpfulness of specific TL=MS activities, as well as to provide demographic information such as years of teaching experience, gender, and grade-level taught. Respondents were provided with a small monetary incentive for completing and returning the survey.

Site Visits

Evaluators visited five cohort-one schools (three elementary, one middle, and one K-8 school) representing four districts in the Bronx during years two and three of the evaluation. During each site visit, a team comprised of a trained evaluator from AED and a math teacher with extensive experience in TL=MS practices conducted interviews and observations with all participating TL=MS teachers. AED developed interview protocols for cohort-one teachers and principals in TL=MS schools. Project implementers contributed to the final instrument design. Questions for teachers focused on their teaching and assessment strategies; the impact of TL=MS on students; school and district support for mathematics reform; leadership opportunities; and the impact of TL=MS on teachers' knowledge of and comfort with mathematics. Questions for the principals were designed to provide information about mathematics reform in each school and the perceived impact of TL=MS on teachers and students.

Each teacher interview was completed in approximately 45 minutes during the school day. In total, 25 teachers and three principals from five schools were interviewed in year two. Sixteen of the teachers were also observed for one class period (approximately 45 minutes) in year two to gather examples of how TL=MS was implemented in the classroom and of student responses to TL=MS strategies and practices. In year three, 14 teachers and five principals were interviewed from the same five schools visited in year two. Site visitors also observed at least one class period for each interviewed teacher in year two. AED staff met with project implementers and teacher consultants to develop the observation protocol and discuss the indicators of success in TL=MS classrooms, including specific evidence of TL=MS teaching strategies.

Pre-Post Student Assessment

TL=MS teacher consultants created the student performance tasks to measure students' growth in mathematical knowledge, strategic knowledge, and written communication skills. The performance tasks were pilot-tested in a small number of classrooms in year three (2001-02). Based on the results of the pilot test, the tasks and scoring rubric were revised and administered in 15 classrooms in year four. The pretest was administered in fall 2002 and the post-test was administered at the end of the school year in spring 2003. Teachers who administered the tasks also completed a short survey investigating their classroom practices and background characteristics. They were provided a small monetary incentive for completing and returning the survey.

The tasks, written at a fourth- and fifth-grade level, asked students to solve a unique problem, explain their thinking, and show their work. Both tasks involved probability concepts—a common topic in fourth- and fifth-grade mathematics classes. The two tasks are presented in the box below.

Performance Task Version A: Probably Buttons
Carmella has a small bag containing 4 green buttons, twice as many blue buttons as green buttons, 2 red buttons, and 6 times as many yellow buttons as red buttons. What is the probability of Carmella choosing a yellow button? Explain your thinking and show your work.
Performance Task Version B: Can of Worms
A can of candy worms had 3 red worms, twice as many blue as red worms, 4 yellow worms, and 5 green worms. Then a student opened the can and ate two of the green worms. After that, what were the chances of pulling out a blue worm without looking? Explain your thinking and show your work.

Two versions of the assessment were created by TL=MS staff and teacher consultants. Approximately half of the classrooms in the sample took version A as the pretest and B as the post-test. The other half of the sample took version B as the pretest and A as the pre-test. This method of administration addressed potential issues related to a "prompt effect" (i.e., one prompt being more difficult than the other.) Preliminary analyses of the results indicated that students tended to score higher on one version. Therefore, as a precaution, "version" was taken into account in all subsequent analyses.

The student assessments were scored using a five-point rubric (0-4). The rubric was adapted from the Exemplars rubrics (www.exemplars.com) and measured mathematical knowledge, strategic knowledge, and written communication. Mathematical knowledge measures students' understanding of the ideas in the problem and the mathematics needed to solve the problem. To receive the highest score in this area, the student must use the correct mathematical terms and labels. Strategic knowledge measures students' understanding of all important parts of the problem. To receive the highest score in this area, it must be clear that the student had a plan for working out the problem and that he/she was able to work through the plan intelligently. Written communication measures the student's ability to give a complete, well-written explanation of the process used to solve the problem. To receive the highest score, the student must answer all the questions completely and clearly. High-scoring work may also include diagrams, graphs or charts. The complete scoring guide is located in the appendix.

Scoring Process

Scorers were mathematics teachers who also served as teacher consultants for Lehman College's New York City Mathematics Project. Most of the scorers were high school teachers who were not directly involved in the TL=MS project in any way; a few scorers were teacher consultants who did work with the project. To address potential bias arising from knowledge of the teachers and schools involved in the study, these scorers only scored student assessments from schools with which they were not involved. Although scorers were not told which assessments were from the "pre" administration and which were from the "post" administration, some of the student work was dated, from which scorers could have determined the order of administration. Each assessment was scored independently by two scorers. Differences in scores were reconciled by a third scorer.

TL=MS Teacher Demographics

TL=MS reached a variety of teachers in terms of background characteristics. Tables 1 to 3 show the demographics, grade-levels taught, and current job status of teachers who returned the final teacher survey administered in spring 2003.

Table 1: TL=MS Teacher Demographics
(n=129)

	Percent
Gender	
Female	86.8%
Position	
Classroom teacher	62.8%
Cluster teacher	9.3%
Staff developer	14.0%
Principal or administrator	4.7%
Other	9.3%
Number of Years Teaching	
1-3 years	11.6%
4-10 years	48.8%
11-15 years	13.2%
16-20 years	11.6%
More than 20 years	14.7%

Source: Final post-program teacher survey, spring 2003.

As shown in the table above, most TL=MS participants (87%) were female. The majority were classroom teachers (63%), while 14% were staff developers and 9% were cluster teachers. Five percent of respondents were principals or administrators. Participants varied in years of teaching experience: 12% had three years or fewer, almost half (49%) had between four and 10 years of experience; 13% had 11 to 15 years, and over one-fourth (26%) had 16 or more years of experience.

Most TL=MS participants taught at the elementary school level, with just over one-fourth (26%) teaching grade 7, 8 or 9 (Table 2).

Table 2: What Grade Are You Currently Teaching? (n=122)	
Kindergarten	27.9%
1st Grade	27.9%
2nd Grade	31.0%
3rd Grade	31.8%
4th Grade	39.5%
5th Grade	24.0%
6th Grade	17.1%
7th Grade	11.6%
8th Grade	10.9%
9th Grade	3.1%
Note: Multiple responses allowed.	

A substantial number of TL=MS participants (31%) had changed schools (including transferring to other schools involved with TL=MS) after they began participating in TL=MS, illustrating the high mobility of teachers served by the project. However, only a small percentage of respondents (4%) reported that they left the profession, as shown in the Table 3 below. Many of the teachers who did not return a survey may also have left their school or profession since first participating in TL=MS.

Table 3: Have You Changed Schools or Left the Teaching Profession Since You Began Participating in TL=MS? (n=120)	
NO	65.0%
YES, changed schools	30.8%
YES, left profession	4.2%

The rest of this report presents findings from multiple data sources over the five years of the evaluation. Findings are organized into two major sections: impact on teachers and impact on students.

Impact of TL=MS on Teachers[3]

> *TL=MS enabled me to look at student work in a different way and to allow students to spend more time looking at their work.* (TL=MS teacher participant)
>
> *I am much more comfortable facilitating rather than controlling. This* [TL=MS] *has affected my teaching in all subject areas!* (TL=MS teacher participant)

The TL=MS model of professional development is grounded in the theory and research showing that, to be effective, mathematics teachers need more than a set of teaching strategies. They also need a deep understanding of mathematics concepts and content, as well as a comfort level in teaching mathematics (Feiman-Nemser, 2001; Fennema & Frank, 1992; Heibert & Carpenter, 1992; Wenglinsky, 2000). For example, Wenglinsky (2000) found that "teachers with greater mastery of their subject and armed with richer and more sustained professional development are better able to teach higher-order thinking skills and engage in related practices, such as hands-on learning"—two practices associated with higher student achievement in mathematics.

Data from TL=MS teachers substantiated this finding. As one TL=MS teacher put it succinctly, "It is impossible to help children come to any understanding if we as teachers do not have deep understanding of the content." TL=MS addressed these needs through a multipronged approach that aimed to improve teachers' understanding and knowledge of mathematical content as well as of how students learn mathematics; increase their use of effective teaching strategies and pedagogy; improve teachers' comfort with and confidence in teaching mathematics; and support the development of teacher leaders. The following section summarizes evidence from teacher surveys, teacher and administrator interviews, and classroom observations related to the impact of TL=MS in these areas.

Overall Impact on Teachers

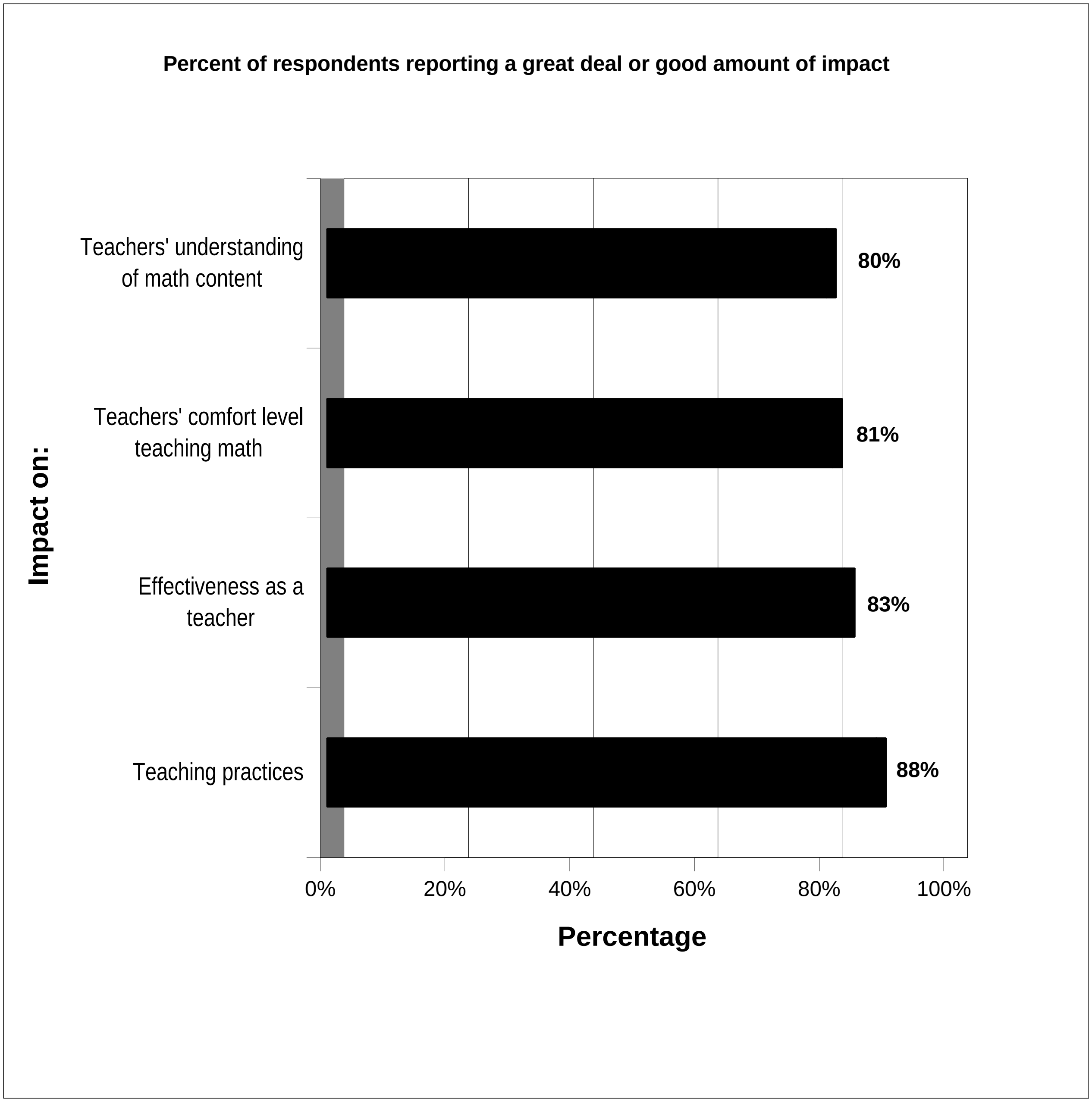
AED collected qualitative and quantitative data to determine the impact of TL=MS on teacher practices. Teacher surveys asked participants to assess the impact of TL=MS on their understanding of mathematics, comfort level with teaching mathematics, and effectiveness as a teacher using a five-point scale (a great deal, a good amount, some, a little, none.) As shown in the figure below, a large majority of participants reported that TL=MS had an impact on all these dimensions. Specifically, 80% or more said it had a good amount or great deal of impact on their understanding of math content and comfort level with teaching. A total of 83% said it had a good amount or great deal of impact on their effectiveness as a teacher, and nearly 90% said it had as large an impact on their teaching practices.

Survey respondents' open-ended responses verified these findings. For example, one teacher commented that before TL=MS she was "intimidated with the use of overheads, manipulatives, even conversations from students," and noted that TL=MS has made her feel more comfortable teaching mathematics. Another wrote that, even though she considered herself "strong in math"

[3] Because the majority of TL=MS participants were teachers, we report the impact on teachers; however participants also included staff developers and a few administrators.

and knowledgeable about mathematics content, she attributed learning “everything I know about the teaching of mathematics” to TL=MS. Teachers’ reports of impact on comfort level were highly correlated to their reports of the impact on teaching practices. Specifically teachers who reported that the project had a great deal of impact on their comfort level with mathematics were more likely to report a great deal of impact on their teaching practices and effectiveness.

Figure 1: Teacher-Reported Impact of TL=MS



Note: Respondents to each item used the following five-point scale: a great deal, a good amount, some, a little, none. The number responding to each item ranged from 126 to 128.

Differences by Groups of Teachers: Years of Experience, Grade Level, Teaching Position, and Cohort Group

Reported impact of TL=MS on practices was high, regardless of years of teaching experience, and was not substantially different for teachers at different stages of their career. Using the five-point scale noted above (a great deal, a good amount, some, a little, none), Veteran teachers (16+ years of experience) just as frequently reported a great deal or good amount of impact on their teaching practices as experienced (4-15 years) and novice teachers (fewer than 4 years). However, a slightly smaller percentage (but not a statistically significant difference) of new teachers reported a great deal or good amount of impact on their comfort level and understanding of math content than their more experienced counterparts. This may be related to the multiple challenges novice teachers face in their first years of teaching when struggling to develop content-area teaching skills, classroom management skills, and relationships with families and the community, as well as negotiating the school and district political environment.

Table 4: Reported Impact of TL=MS by Years of Teaching

Area of impact	All teachers n=126 to 128	As a result of participation in TL=MS, percent of teachers reporting a great deal or good deal of impact New teachers <=3 yrs n=15	Experienced teachers 4-15yrs n=80	Veteran teachers 16+yrs n=34
Teaching practices	88.3%	86.7%	88.6%	88.2%
Effectiveness	83.3%	86.7%	83.3%	81.8%
Comfort level	81.3%	73.3%	82.3%	82.4%
Understanding of math content	79.7%	73.3%	79.7%	82.4%

Note: Differences were not statistically significant using chi-square test. Respondents used the following five-point scale: a great deal, a good amount, some, a little, none.

Not all TL=MS teachers had been teaching mathematics their entire career. For example, five respondents were experienced teachers who were new to math. When reported impact is compared by teachers' years of teaching mathematics, rather than overall teaching experience, differences between the groups diminish, as shown below. In sum, disaggregating the data by overall years of teaching experience and years of teaching mathematics showed that teachers in the early stages of their careers benefited from TL=MS as much as teachers with substantial experience. This finding supports the literature and research that advocates for the need for serious and sustained learning opportunities at every career stage (Feiman-Nemser, 2001) and indicates that TL=MS is an effective professional development model for all teachers, not just those who are inexperienced or underskilled in teaching mathematics.

Table 5: Reported Impact of TL=MS by Years of Teaching Math				
		As a result of participation in TL=MS, percent of teachers reporting a great deal or good deal of impact		
Area of impact	**All teachers** n=126 to 128	**New math teachers** <=3 yrs n=20	**Experienced math Teachers** 4-15yrs n=83	**Veteran math teachers** 16+yrs n=26
Teaching practices	88.3%	85.0%	90.2%	84.6%
Effectiveness	83.3%	85.0%	84.0%	80.0%
Comfort level	81.3%	80.0%	82.9%	76.9%
Understanding of math content	79.7%	80.0%	80.5%	76.9%
Note: Differences were not statistically significant using chi-square test. Respondents used the following five-point scale: a great deal, a good amount, some, a little, none.				

Compared with school administrators and staff developers, classroom and cluster teachers have closer contact with students and what goes on in the classroom. For that reason, Table 6 is restricted to classroom and cluster teachers to focus more clearly on the reported impact of TL=MS on their teaching practice, effectiveness, comfort level, and understanding of math content.

Table 6: Reported Impact of TL=MS by Grade Level Classroom and cluster teachers only				
		As a result of participation in TL=MS, percent of teachers reporting a great deal or good deal of impact		
Area of impact	All teachers n=91 to 93	K-3 teachers n=49*	4-6 teachers n=52*	7-9 teachers n=18*
Teaching practices	89.9%	89.8%	94.2%	83.3%
Effectiveness	83.5%	83.7%	84.3%	76.5%
Comfort level	81.8%	85.7%	86.5%	66.7%
Understanding of math content	79.8%	85.7%	80.8%	66.7%
*Teachers could have indicated that they were teaching more than one grade level. Note: Differences were not statistically significant using chi-square test. Respondents used the following five-point scale: a great deal, a good amount, some, a little, none.				

Teachers working with students in grades 4 to 6 more frequently reported that TL=MS had a great deal or good deal of impact on their teaching practices (94.2% of fourth- to sixth-grade teachers) than teachers at the early elementary level (89.8% of K-3 teachers) and in the middle grades (83.3% of grades 7-9 teachers). Slightly fewer teachers working with seventh to ninth graders reported an impact on any of the four areas of change (practice, effectiveness, comfort or understanding) compared with elementary teachers. This was especially true in the areas of comfort level teaching mathematics and understanding of math content. For example, about 80% of all TL=MS teachers reported a great deal or good deal of impact on their comfort level or understanding of math content, but only 66.7% of teachers working with the middle grades reported this level of impact. While there appears to be a pattern of less satisfaction on the part of middle-grades teaches compared with that of elementary teachers, these differences are not statistically significant and may be related to the small number of middle-grades teachers in the sample.

Nonetheless, interview data and responses to open-ended comments indicated that some teachers thought the TL=MS professional development would have been improved if focused on specific grade levels, rather than across grade levels. For example, one middle-grades teacher suggested in an interview that professional development sessions separate teachers into different levels (e.g. elementary and middle) rather than mixing across levels. This teacher reported that she thought the mixed grouping "slowed things down." She also said, "There's not as much to learn from each other if you're at different levels, I'd learn a lot more talking with all middle school teachers." Regardless of the reasons that a smaller proportion of middle-level teachers reported a good or great deal of impact, it is important to note that the majority of this group did report a strong impact.

Classroom and cluster teachers reported very similar amounts of impact to staff developers but analyses by cohort group did reveal small (but not statistically significant) differences. Specifically, a greater proportion of respondents from cohort two of the initiative reported a good amount or great deal of impact from TL=MS in three of the four areas discussed above, especially in teaching practices and effectiveness as a teacher (see table 7 below).

Table 7: Reported Impact of TL=MS by Cohort Group

		As a result of participation in TL=MS, percent of teachers reporting a great deal or good deal of impact		
Area of impact	**All teachers** n=126 to 128	**Cohort 1** n=50	**Cohort 2** n=46	**Cohort 3** n=32
Teaching practices	88.3%	84.0%	93.5%	87.5%
Effectiveness	83.3%	79.6%	89.1%	80.6%
Comfort level	81.3%	80.0%	82.6%	81.3%
Understanding of math content	79.7%	82.0%	76.1%	81.3%

Note: Differences were not statistically significant using chi-square test. Respondents used the following five-point scale: a great deal, a good amount, some, a little, none.

Impact on Specific Practices

> *Being an active participant in learning math concepts and skills caused me to struggle as the students do in my classroom. When you know and feel what the struggle is for our children, then you're better able to deal with their needs. I'm more sensitive and more patient with my students now. My students are more relaxed—they feel free to take risks and I am happy to allow them to.* (TL=MS teacher participant)

TL=MS emphasized constructivist and inquiry-based instruction to foster students' active learning, learning with understanding, and higher order thinking skills. This constructivist focus is in alignment with National Council of Teachers of Mathematics principles and standards (NCTM, 2000) and other research on practices associated with high academic performance (Wenglinsky, 2000.) Further, the TL=MS professional development model focuses on using frequent and multiple forms of student assessment to provide useful information and to support student learning. Practices associated with these principles include using manipulatives and hands-on activities; asking students to write and talk about their problem solving; and using multiple forms of assessment, including open-ended assessment techniques. In this report, activities in line with TL=MS professional development are referred to as "reform-based" practices or instruction.

A primary goal of the project was to develop teachers' use of reform-based practices. Early in the project, project staff and course instructors found that TL=MS participants demonstrated "scant understanding" of constructivist approach and "weak grasp" of mathematics content and standards-based curriculum and assessment (Lehman College, 2000). Indeed, results of the pre-summer institute survey of cohort-one teachers showed just 17 percent of teachers feeling "very prepared" to use constructivist pedagogy in their classroom and just 19 percent "very prepared" to implement inquiry or discovery learning or to phrase questions to encourage open-ended investigations. Further, several teachers described themselves as "math phobics" at the beginning of the project. After just one year of involvement with the project, post-survey results with the first cohort of teachers showed that teachers' familiarity with inquiry-based learning, standards-based curricula and instruction, and constructivist pedagogy increased substantially (AED, 2000).

By the end of the project, results of the final teacher survey showed that TL=MS teachers used reform-based practices as part of their everyday practices. For example, using a four-point scale (almost every day, 1-2 times a week, 1-2 times a month, never or hardly ever), nearly all responding teachers (97%) reported asking students to discuss solutions to math problems with other students and talk to the class about their math work almost every day or one to two times a week. According to the survey, the majority of teachers attributed an increase in using reform-based practices, such as having students explain how they arrived at answers, write about problem solving, write in journals, and make literature connections, to TL=MS (see table below).[4] The practice for which the greatest proportion of teachers reported an increase due to TL=MS was asking students to discuss solutions to mathematics problems. At the same time, the majority of teachers reported that, as a result of participation in TL=MS, their use of traditional practices in the classroom—such as lectures and worksheets—decreased or stayed the same.

[4] For each practice, teachers were asked to report the frequency with which they used the practice (using a four-point scale) and separately to self-report if the practice increased, stayed the same, or decreased as a result of their participation in TL=MS.

Table 8: Changes in Instructional Practice as a Result of TL=MS Participation				
		As a result of participation in TL=MS, percent of teachers reported their practice has:		
How often do students in your class do each of the following?	**Almost every day/1-2 times a week**	**Increased**	**Stayed the Same**	**Decreased**
Reform Practices				
Explain how they arrived at their answers.	99.2%	78.3%	21.7%	0%
Discuss solutions to mathematics problems with other students.	96.8%	81.0%	19.0%	0
Talk to the class about their mathematics work.	96.8%	75.7%	23.4%	0.9%
Solve mathematics problems in small groups or with a partner.	96.0%	76.3%	23.7%	0%
Work and discuss mathematics problems that reflect real-life situations.	91.9%	71.7%	28.3%	0.0%
Work with manipulatives (e.g., color tiles, pattern blocks, multilink cubes.)	90.3%	74.3%	24.8%	0.9%
Investigate problems that have multiple solutions.	89.5%	73.7%	25.4%	0.9%
Write at least a few sentences about how to solve a mathematics problem.	83.6%	69.6%	29.5%	0.9%
Write in mathematics journals or logs.	79.7%	67.0%	31.3%	1.8%
Provide extensions to mathematics problems.	74.6%	61.9%	37.2%	0.9%
Use literature connections.	69.4%	76.6%	21.6%	1.9%
Create rubrics to score their work.	47.1%	56.8%	42.3%	0.9%
Other Practices				
Use a computer.	50.4%	30.3%	64.2%	5.5%
Use a calculator.	34.4%	26.5%	69.0%	4.4%
Traditional Practices				
Complete worksheets for drill or practice.	53.2%	14.9%	49.1%	36.0%
Listen to a lecture from the teacher.	48.3%	12.7%	50.9%	36.4%
Note: The number of teachers responding to each item ranged from 120 to 125. Respondents used the following four-point scale: almost every day, 1-2 times a week, 1-2 times a month, never or hardly ever.				

Teachers also reported frequently using multiple forms of assessment consistent with reform practices (such as portfolios, peer evaluation and rubrics). For example, over half of TL=MS teachers (54.4%) reported an increase in the use of projects or presentations, and almost 70% used short written responses to assess students' progress at least once or twice a week. In the case of assessment, one-half to two-thirds of responding teachers also attributed increases in their use of reform-based assessment practices to TL=MS. In contrast, 83% reported their participation in TL=MS resulted in no change in, or a reduction in, their use of multiple choice tests (see table 9).

Table 9: Changes in Teacher's Assessment of Student Progress as a Result of TL=MS Participation				
Assessment practice	**How often do you use the following?** **Almost every day/1-2 times a week**	**As a result of participation in TL=MS, percent of teachers reporting their assessment has**		
		Increased	**Stayed the Same**	**Decreased**
Individual or group projects or presentations	45.5%	54.4%	42.1%	3.5%
Multiple choice tests	20.7%	16.8%	64.6%	18.6%
Short written responses	68.4%	65.5%	32.7%	1.8%
Portfolios	43.5%	51.8%	40.9%	7.3%
Peer evaluation——students evaluate each other's work	50.4%	50.0%	47.3%	2.7%
Teachers use of rubrics to score students' work	62.3%	67.2%	31.9%	0.9%
Students apply rubrics to score their own or others' work	46.3%	62.6%	35.7%	1.7%
Note: The number of teachers responding to each item ranged from 115 to 123. Respondents used the following four-point scale: almost every day, 1-2 times a week, 1-2 times a month, never or hardly ever.				

Not only did teachers report that the frequency with which they used reform practices increased as a result of their participation in TL=MS, but comparisons with a national sample of teachers indicated that TL=MS teachers were much more likely to use these reform practices than most teachers nationally. Differences between fourth-grade TL=MS teachers and the national sample of fourth-grade teachers responding to the National Assessment of Educational Progress (NAEP)[5] survey were substantial. For example, as shown in Table 10 below, almost double the percentage of TL=MS teachers reported asking students to discuss solutions to math problems almost every day compared with a national sample of teachers at the same grade level (86% vs. 44%). Over three times as many TL=MS teachers reported asking students to solve math problems in small groups or with a partner (80% vs. 26%), and over four times as many TL=MS teachers reported asking students to write a few sentences about how to solve a math problem almost every day compared with teachers nationwide (59% vs. 14%). Consistent with reform practices, TL=MS teachers were less likely to report using multiple choice exams to assess student learning compared with teachers nationwide (12% vs. 20% reported using the practice 1-2 times a week).

[5] The NAEP is based on a nationally representative sample of students, not teachers. Therefore, the national results presented here pertain to the practices of teachers of a representative sample of fourth-grade students. Percentages reported from the NAEP teacher survey should be interpreted as the percentage of students whose teachers use that pratice.

Table 10: Instructional and Assessment Practices of TL=MS Teachers and a National Sample (NAEP)		
Practice	**Grade 4 TL=MS teachers (n=51)**	**National sample** of grade 4 teachers (year surveyed)**
	Percent reporting "Almost everyday"	
Students discuss solutions to math problems.	86%*	44% (2000)
Use a computer.	35%*	15% (2000)
Work and discuss math problems that reflect real-life situations.	67%*	39% (2000)
Solve math problems in small groups or with a partner.	80%*	26% (2000)
Write a few sentences about how to solve a math problem.	59%*	14% (2000)
	Percent reporting at least "1-2 times a week"[6]	
Use individual or group projects or presentations to assess student progress in math.	51%*	6% (2003)
Use multiple choice tests to assess student progress in math.	12%	20% (2003)
Use short written responses to assess student progress in math.	70%*	32% (2003)
Use portfolios to assess students' progress in math.	33%**	17% (1996)

*$P \leq .001$, 1-tailed ** $P \leq .01$, 1-tailed, based on Z test of sample versus population. Number responding to each item ranges from 48 to 51 for the TL=MS sample.
** Source: National Assessment of Educational Progress (NAEP). Different items were asked in different years of administration. Data are reported for the most recent year item was asked.
Note: Respondents used the following four-point scale: almost every day, 1-2 times a week, 1-2 times a month, never or hardly ever.

The fact that teachers up to five years after first participating in TL=MS continue frequent use of the reform-based approach espoused by the project indicates that reforms related to TL=MS were sustained at least at the classroom level.

Teachers were also asked about their emphasis on different areas in planning their mathematics class. Using a three-point scale (heavy emphasis, moderate emphasis, little/no emphasis), teachers reported placing the most emphasis on developing an appreciation for the importance of

[6] Data on use of assessment practices show the percentage reporting using the practice "at least 1-2 times a week" instead of "almost everyday" because very few teachers used the assessment practices almost everyday.

math, learning how to communicate ideas in mathematics effectively, and developing reasoning and analytic ability to solve unique problems, areas consistent with a reform-based approach (see Table 11). Compared with teachers nationwide, TL=MS teachers were much more likely to report placing heavy emphasis on these areas. Further, as shown in Table 12, the majority (between 72% and 87%) of TL=MS participants reported that their emphasis on these areas increased as a result of their participation in TL=MS.

In addition the data show that TL=MS teachers were able to emphasize these approaches without sacrificing emphasis on learning facts and procedures. The majority of TL=MS teachers also placed heavy emphasis on students' learning mathematics facts and concepts, as well as skills and procedures for solving routine problems, but to a much less extent than teachers nationally (Table 11). These findings illustrate a shift in TL=MS teachers' pedagogy from more traditional mathematics instruction to include more reform-based practices and indicate that TL=MS teachers use reform-based practices to a much greater extent that most teachers nationally, while maintaining an emphasis on facts and concepts and the skills and procedures necessary to solve routine problems. This finding highlights the fact that emphasis on constructivist and traditional approaches do not need to be mutually exclusive but rather can be integrated for a balanced approach.

Table 11: Instructional Emphasis of TL=MS Teachers and a National Sample (NAEP)

Practice	Grade 4TL=MS teachers	National sample** of grade 4teachers (year surveyed)
	Percent reporting "heavy emphasis"	
Developing an appreciation for the importance of math	90%*	65% (1992)
Learning how to communicate ideas in mathematics effectively	90%*	50% (2000)
Learning mathematics facts and concepts	73%*	91% (2000)
Learning skills and procedures needed to solve routine problems	86%**	93% (1992)
Developing reasoning and analytic ability to solve unique problems	88%*	62% (2000)

*$P \leq .001$, 1-tailed ** $P \leq .05$, 1-tailed, Number responding to each item ranges from 48 to 51 for the TL=MS sample.
**Source: National Assessment of Educational Progress (NAEP). Different items were asked in different years of administration. Data are reported for the most recent year item was asked.
Note: Respondents used the following three-point scale: heavy emphasis, moderate emphasis, little/no emphasis.

Table 12: Changes in Teacher's Emphasis as a Result of TL=MS Participation (n=115 to 116)			
	As a result of participation in TL=MS, percent of teachers reporting their emphasis has		
Area of emphasis in planning mathematics class	**Increased**	**Stayed the Same**	**Decreased**
Developing an appreciation for the importance of mathematics	72.4%	27.6%	0
Learning how to communicate ideas in mathematics effectively	87.1%	12.9%	0
Learning mathematics facts and concepts	60.3%	37.1%	2.6%
Learning skills and procedures needed to solve routine problems	65.2%	33.0%	1.7%
Developing reasoning and analytic ability to solve unique problems	81.9%	18.1%	0

Interviews with TL=MS teachers and comments on open-ended survey items revealed parallel findings and illustrated the impact of TL=MS on teachers. More than just increasing or decreasing use of specific practices, teachers described changes in their whole approach to mathematics instruction. The following are several examples of the ways teachers described changes they had made in their practices:

> *Before TL=MS I was much more structured. Now, I let students explore more, and I use multiple approaches to help them learn.*

> *I've learned how to expand lessons and I pay a lot more attention to students. I spend a lot more time listening to what they are saying.*

> *I try to create an atmosphere where it is okay to be wrong. I want to enrich their experiences in math and make sure they've participated inside the process and learned from it.*

> *TL=MS enabled me to look at student work in a different way and to allow students to spend more time looking at their work.*

> *I am much more comfortable facilitating rather than controlling. This* [TL=MS] *has affected my teaching in all subject areas!*

One principal summed up the changes he has seen in the classroom since participating in TL=MS:

> *If you see a good lesson you see a lot of thinking* [among students] *happening. Teachers want students to make sense of math now.*

For many teachers, increased confidence in their understanding of mathematics content translated into increased comfort in teaching it. A typical comment to this effect included:

> *I was always a bad math student. TL=MS has helped me. I wasn't comfortable teaching it. Now, I gravitate to math because I'm so comfortable teaching it.*

> *TL=MS has helped me feel competent in math. The content piece of the summer institute opened some doors in my classroom. It boosted my confidence.*

With a deeper comprehension of the content, teachers were able to extend math problems, allowing students to explore different answers and problem-solving strategies, and present concepts in multiple ways. The following comments from teachers illustrate this point:

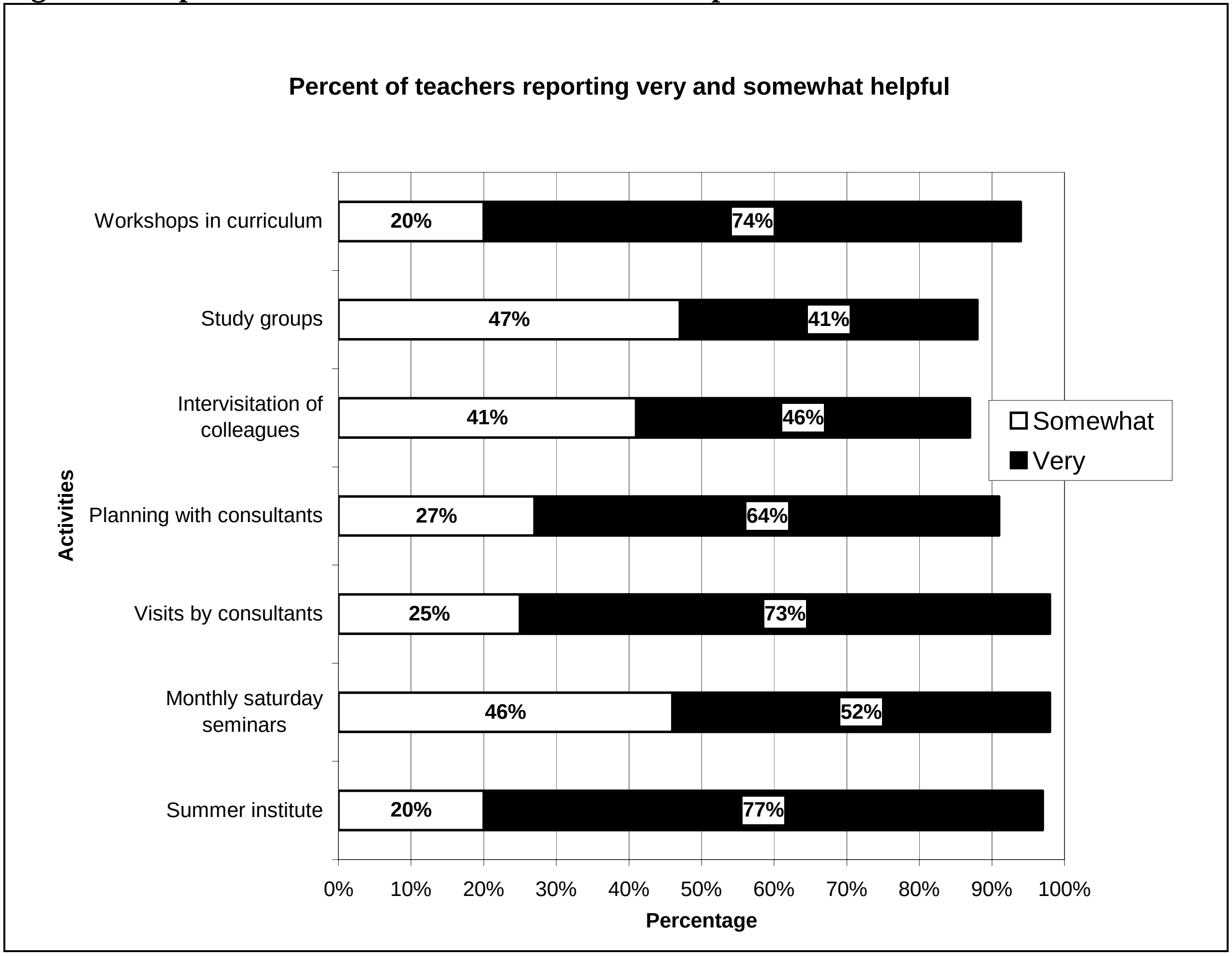
> *If you understand the content, it's easier to teach and give multiple ways of presenting a concept. Otherwise, you're stuck teaching only the way it is presented in the textbook.*
>
> [Through TL=MS], *I've seen so many different people solve things in different ways. That helps me as a teacher because I am able to explain it in more ways to students and to understand what their perspective might be.*

Impact of Individual TL=MS Professional Development Activities

> *TL=MS helped expand my comfort zone and become more confident sharing knowledge and becoming a leader in what I'm good at. It allowed me to share myself. I wouldn't have thought of myself as a teacher leader without TL=MS building that up and increasing my* [content] *knowledge.* (TL=MS teacher participant)

Most (95% or more) of survey respondents participated in all the various professional development activities offered through TL=MS. Reactions to these activities were very positive with surveyed teachers giving very high ratings of helpfulness to the various TL=MS professional development activities (see figure below).

Figure 2: Helpfulness of TL=MS: Professional Development Activities



Note: The number responding to each item ranged from 122 to 124. Respondents used the following five-point scale: very helpful, somewhat helpful, not very helpful, not at all helpful, and I did not participate in this. Respondents reporting they did not participate in the activity were not included in this analysis.

Using a five-point scale (very helpful, somewhat helpful, not very helpful, not at all helpful, I did not participate in this), respondents gave the highest rating of helpfulness to visits from teaching consultants to the school and monthly Saturday sessions closely followed by the summer institute. About three-fourths of respondents described these two activities as very helpful, with the remaining saying they were somewhat helpful.

> *The most helpful aspect of TL=MS project was having a consultant visit and model engaging lessons using manipulatives, game etc. I also learned a lot from the consultant modeling how to question the students on their mathematical thinking.*

> *The most helpful aspect of the TL=MS project was the visits by the consultants in the classroom. It allowed time to reflect on practices and build a shift in my mathematical philosophy. It gave me a chance to conference and assess my own technique.*

Teacher consultants (TCs) visited schools regularly—during the first year of involvement, TCs typically spent a day per week at the school; in the second year, they typically visited a school twice a month; and, by year three, TCs visited schools about once a month. They worked with TL=MS participants individually and in groups. They met with teachers to plan lessons; they observed TL=MS teachers and provided feedback; they counseled teachers on specific problems; and they modeled, demonstrated, and co-taught lessons. TCs met with the TL=MS team as a group to plan curriculum and interdisciplinary units. TCs also served as a resource for materials and ideas, and some provided off-site support through telephone or e-mail contact. Lastly, some TCs also led study groups or conducted professional development for all TL=MS participants.

Teachers also found planning sessions with teacher consultants very helpful, with 64% reporting so. Other TL=MS professional development activities also received high ratings, with no less than two-fifths describing them as very helpful. Teachers were least enthusiastic about the study group. One teacher noted: "The study groups didn't offer the same level of inquiry as the institute and Saturday seminars."

A few teachers struggled with mathematics content that was outside their teaching level, reporting that the least helpful aspects of TL=MS were "…when the lecturers presented concepts that were not geared to elementary school children," and noted that the math content was "too abstract" or "too high a level." As a result, as mentioned above, several teachers suggested grouping teachers by grade level for the courses, seminars and study groups. A few teachers reported otherwise, noting that they appreciated the cross-level groupings: "Teaching elementary school can be limiting; I had forgotten middle school math, which enlightened my understanding greatly."

Another powerful impact of TL=MS was highlighted in participant responses to an open-ended question about the most helpful aspect of TL=MS. Just behind "visits from the TC," the second most frequently cited aspect was "sharing with and learning from colleagues." One teacher wrote that her TL=MS colleagues provided "needed support" and encouragement. Others noted that it was beneficial to "observe that there are multiple ways of solving problems and hearing others share their solutions." Another teacher reported that as a result of TL=MS, teachers at her school developed a math team to plan goals and new initiatives for the coming year. The team met weekly and shared concerns and successes. Several teachers also noted that it was helpful to meet teachers from other schools and districts. Other comments of this vein included the following:

> *The most helpful aspect was listening and learning from the instructors and my colleagues on various different ways to finding solutions for a problem. Also having a math cohort to ask specific questions when perplexed, just as support, or listening to each others' ideas.*

> [In the study groups] *ideas were shared. It was a positive thing to find that others were experiencing similar problems/success.*

> *Being an active participant in learning math concepts and skills caused me to struggle as the students do in my classroom. When you know and feel what the struggle is for our children, then you're better able to deal with their needs. I'm more sensitive and more patient with my students now. My students are more relaxed—they feel free to take risks and I am happy to allow them to.*

Many of these quotes illustrate what Lord (1994) refers to as "productive disequilibrium." TL=MS provided a community of learners or "critical colleagueship" through which teachers could explore, articulate and question their own and others' ideologies and practices: "Instead of relying on routine dissemination of information and techniques to inspire new practices, critical colleagueship turns to increased reflection, informed debate, honest disagreement, and constructive conflict as tools of change (p 194)."

Impact on Teacher Leaders

> *TL=MS empowered me to step out of my safe classroom. I became a staff developer for two schools using many of the techniques I learned through* [TL=MS]. (TL=MS teacher participant)

According to the final teacher survey, 27 percent of respondents reported that their position had changed since they began participating in TL=MS. Of those teachers, 43 percent said the change was to a leadership position (e.g. staff developer), and they believed their promotion to that position was in some way related to their participation in TL=MS. In explaining how TL=MS had an impact on her moving into a teacher leader position, one teacher wrote on the final teacher survey, "I think my principal began to see some leadership qualities in me." A teacher who was already in a leadership position as a staff developer wrote that TL=MS helped develop her leadership skills even further. Specifically, she wrote that TL=MS helped her "develop and focus professional conversations about mathematics and promoted sharing of ideas between math teachers." Other teachers wrote:

> *It* [TL=MS] *empowered me to step out of my safe classroom. I became a staff developer for two schools using many of the techniques I learned through* [TL=MS].
>
> *TL=MS provided me with the skills and experience needed for the staff developer position.*
>
> *TL=MS taught me how to keep an open mind while teaching math and how to turn-key information to other teachers.*
>
> *With the experience I gained as a teacher leader, I was encouraged to do staff development.*
>
> *Not only have I become a more effective teacher of mathematics, but also a more effective staff developer.*
>
> *Participation changed my professional life—I went from class teacher to staff developer. Now I have applied for math coaching jobs . . . I return with a whole new approach.*

Several teachers wrote that, in addition to providing them with better teaching strategies and content knowledge, TL=MS gave them the "confidence needed to be an effective leader." Another stated:

I started this program as a new teacher, not wanting to become a leader—just wanting to learn how to teach better. Boy did I learn a lot of things for which I am very grateful. Most of all I learned that tough I may not be an expert; I do have something to contribute.

This quote illustrates the point that participants may not have viewed themselves as teacher leaders before TL=MS, but that TL=MS helped them develop as leaders and gain the experience, confidence and skills they needed to do so. The next section describes findings from interviews with TL=MS teachers regarding how they share knowledge as teacher leaders to ensure the spread of TL=MS beyond their own classroom.

Ways of Sharing Knowledge

TL=MS teachers described many formal and informal ways in which they shared what they learned from TL=MS. Eight of the 15 interviewed teachers reported that they shared what they learned through formal and informal mechanisms, while seven described more informal ways for sharing information. Formal mechanisms included providing staff development workshops to their school and at the district level, participating in school leadership teams and in district mathematics meetings, presenting at conferences, and leading and participating in grade-level or team meetings during common planning periods. Three of the teachers who described sharing information in these ways were designated staff developers for their school, and one was the staff developer for the district assigned to work with the school. In their roles as staff developers, all the teachers described additional ways of working with teachers, including co-teaching, modeling lessons, mentoring teachers, and providing one-on-one assistance to teachers as needed. Several teachers noted that TL=MS helped increase their confidence as teachers and teacher leaders, as the following quote indicates:

TL=MS helped expand my comfort zone and become more confident sharing knowledge and becoming a leader in what I'm good at. It allowed me to share myself. I wouldn't have thought of myself as a teacher leader without TL=MS building that up and increasing my [content] *knowledge.*

Seven teachers reported that they shared what they learned with non-TL=MS teachers through much more informal mechanisms. This included informal "hallway" discussions with teachers and assisting teachers seeking their help. For two of these teachers, their school already had staff developers in charge of conducting professional development, which they believed limited the extent to which they were encouraged to share what they had learned. As two teachers put it:

The first year I did a lot more sharing and training, but [the staff developer] *does that now.*

I have not been provided with enough rope to go further, either there's not enough time or I am told someone else is already doing that [professional development].

Several TL=MS teachers noted that while some or most of their colleagues were very open to hearing about new ideas and strategies and routinely sought out their advice or assistance, other colleagues were not as open. Some came to professional development workshops conducted by TL=MS teachers only when they were required to do so by administration. One teacher explained:

Some teachers don't like to use math games because they think they're only for fun. I have a hard time convincing them that kids get a lot more out of games than they do worksheets.

Another mentioned that several teachers in her school, "don't like manipulatives. They think the kids are too noisy with them."

Spread Beyond TL=MS Classrooms

Of the five schools visited, three showed evidence that the impact of TL=MS had spread beyond the classroom of the teachers directly involved. TL=MS teachers conducted schoolwide or districtwide workshops. They shared materials, lessons, strategies and techniques with their non-TL=MS colleagues and were seeing changes in non-TL=MS classrooms as a result. Several TL=MS teachers were seen as a resource for all mathematics teachers in the school. Finally, these teachers had an impact at the policy level through leading grade-level meetings, serving on school-based management and curriculum teams, and assisting in writing school comprehensive education plans (CEPs).

In the three schools where the impact went beyond individual TL=MS teachers, at least one TL=MS teacher became a staff developer for the school, which facilitated the spread of TL=MS practices and philosophy. This "spread" also took place in schools where more than one cohort of teachers participated in the project. In these schools, interviewed teachers and principals reported that multiple cohort participation helped create a "critical mass" of teachers involved in TL=MS who shared information they had learned and used a common language around mathematics instruction. In addition, these schools seemed to benefit from the continued presence of a TL=MS teacher consultant. Although the TCs were officially charged with working with cohort-two and -three teachers in 2001-02, they often maintained their contact with cohort-one teachers, offering assistance when at the school.

In two of the five visited schools, teachers reported that the impact of TL=MS remained within their own classrooms. Neither school had a TL=MS teacher in a staff development role. In one school, teachers reported that the opportunities to share what they learned were restricted because their administration did not support such sharing. They also had little time for professional development in the school. One TL=MS teacher from this school had previously been the school staff developer, but, as a result of funding cuts, was now a full-time classroom teacher. She reported that she had no opportunities to share what she had learned on a formal basis:

There's too much to do as a fulltime classroom teacher to do workshops and professional development. I do some informal sharing with teachers, but much less than last year.

For full-time classroom teachers, the general sentiment regarding sharing what they learned is illustrated by one teacher who said:

You have to invest the time in professional development. It can't be another thing to do on top of all the other things to do.

In the other school with limited "spread" of TL=MS, one of the three TL=MS teachers left the school and another became an assistant principal. Although the third teacher shared some information on a very informal basis, she reported that she did not have time to do so more regularly or formally:

> *A teacher leader can only go so far. Carrying a full teacher load, it's impossible to give support to other teachers.*

The assistant principal also reported that her new responsibilities severely limited the extent to which she could focus on professional development in mathematics.

Support and Structures Needed to Be "Teacher Leaders"

Teacher leaders identified three main supports that made it possible for them to share what they learned from TL=MS: time to meet other teachers, the support of administration to be teacher leaders, and the support of colleagues, such as that provided by the TL=MS network.

In terms of time to meet with other teachers and conduct professional development, one staff developer who described meeting with other teachers before and after school for planning said, "The meeting time, talking and sharing across grades has a strong impact." Teachers also indicated that to have a lasting impact, schools must have a structure in place for ongoing professional development, including common planning periods, regular professional development sessions, and release time for teachers to attend professional development outside of school. Two teachers noted:

> *The challenge is finding the time to really work with teachers at their grade levels. Teachers have sought me out and I have shared many different lessons, but a workshop here and there is not enough to make the difference in the classroom, to change the way of thinking teachers are accustomed to.*

> *The support a teacher leader needs is the time to do it. Time has to be built in to provide guidance, support and modeling of lessons. Time is critical; otherwise, it's not going to happen.*

Second, teachers reported that they needed support from their administration to be teacher leaders. Support from principals or assistant principals legitimized teacher leaders in the eyes of their peers and also allowed them to explore new ideas and practices. In talking about the support she received from administrators, one teacher stated:

> *My supervisor encourages me to work with other teachers and share what I know. He sends teachers to me for guidance.*

Another said that she believed she was an effective teacher leader in part because of the support she received from her administration:

> *The administration is very supportive. They take our suggestions and go with our ideas. They come to us and ask for our input.*

Conversely, a lack of administrative support was cited by some respondents to the teacher survey as an insurmountable barrier to effective leading. A few teachers reported that their administration did not support common prep periods for teachers to plan nor provide coverage so that teachers could attend TL=MS events. In addition to lowering teacher morale, one participant wrote that the “lack of commitment of administrators reduces the success and achievements that could have been gained.” Other teachers suggested that administrators should be required to attend TL=MS professional development, or, as one teacher noted, “at least give them an understanding of the project and the great deal of effort and work put forth by the participating teachers.” In one school, the administration and district emphasized reading instruction to the detriment, teachers believed, of mathematics instruction. "To tell you the truth, the focus here is on reading, so not much attention is paid to math," said one teacher.

These findings about the administrative support needed to foster leadership around mathematics corroborates findings reported by TL=MS project staff. Specifically, project staff described the ongoing support and involvement of school principals and other administrators as crucial: “They can influence positively or negatively development of a team’s esprit de corps, the effectiveness of our on-site consultants, other staff members’ awareness of team efforts, availability of designated meeting times among team members and on-site consultants, communication in general and, most importantly, the progress of mathematics education at the school level” (Lehman College, 2000).

Third, teachers reported that the network provided by TL=MS was instrumental in their efforts to be teacher leaders in their school. One principal of a school with teachers participating in cohorts two and three credited the project with helping the school have “one voice when it comes to mathematics.” He noted that the school sees TL=MS as a resource as well. "When we have a problem or need help, we call on them." Several teachers from other schools also reported that they maintained their connection with the teacher consultants and staff of TL=MS and continue to call upon them for assistance when needed. In most cases, these teachers were at schools still participating in the project with cohort two or three teachers, and therefore they still had access to teacher consultants who visited their school as part of their work with subsequent cohorts. For example, one teacher said:

> *TL=MS builds a support group and I call on that group. It gives you a good network. I still e-mail* [our teacher consultant] *if I need help.*

A teacher in a different school reported that her connection with the TL=MS teacher consultant also continued beyond the formal ending of her participation in the project. Describing her relationship with the TL=MS teacher consultant, she said, "I have my TC's phone number on speed dial on my cell phone and I use it!" In contrast, one teacher, whose school did not have cohort-two or -three teachers, lamented that she lost her network when the project ended:

> *I miss the program* [TL=MS] *and going to the meetings at the district. It's hard when there's no one around. The support network is gone. I'm isolated in my room. There was a whole level of enthusiasm last year* [when the school was still part of TL=MS] *that is gone.*

Another said:

> *I was hoping they* [TL=MS] *would offer some activities this year. I felt I could use another year of training and support. Also, they would let us know what was happening nationwide with math. I don't feel that informed anymore.*

In addition to the support provided by the TC, teachers found tremendous support from the collegial relationships developed through the project. As noted earlier, one of the TL=MS aspects most frequently cited by survey respondents as helpful was the opportunity to meet teachers from other schools and to share and learn from each other.

Impact of TL=MS on Students

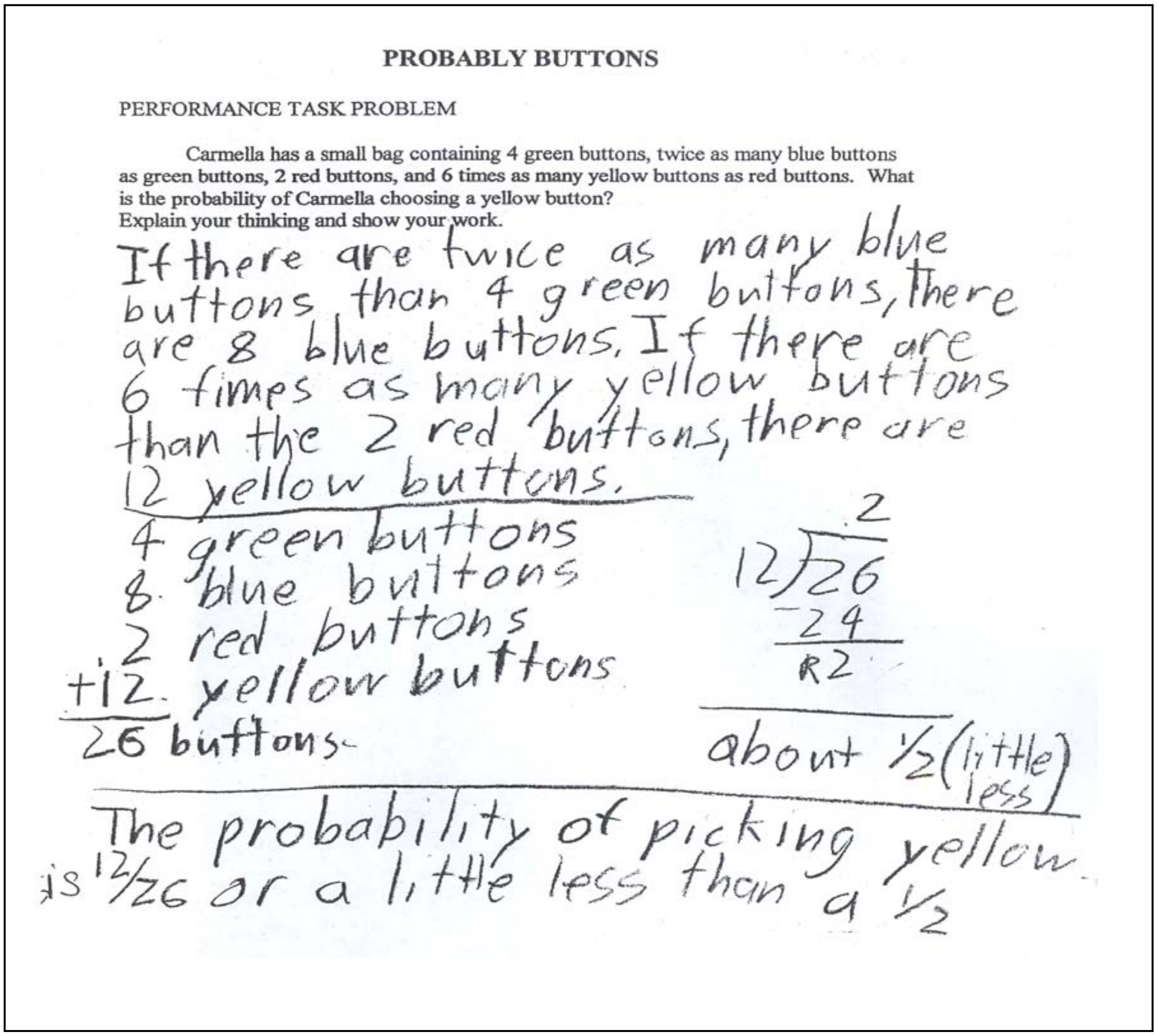
> *Kids get extremely excited about math—it is the highlight of their day! Even those who are struggling with some of the math concepts still get to participate and "play." There are a lot of student-helping-student scenarios.* (TL=MS teacher participant)

The theory of change behind TL=MS holds that high-quality, intensive professional development and support will result in increased teacher comfort level and confidence in teaching mathematics, as well as increased use of reform-based practices and improved teacher effectiveness that will result in greater student achievement. To test the theory that changes in teacher practices results in greater student performance, a sample of TL=MS teachers and comparison non-TL=MS teachers administered a pre-post assessment of students. Teachers were also asked through surveys and interviews to report the impact of TL=MS on their students. Findings from the pre-post assessments and teacher surveys and interviews are summarized in the following section.

Pre-Post Assessment Results

As noted in the methodology section, students completed a pre-post performance assessment designed to measure their mathematical knowledge, strategic knowledge, and mathematical communication (see scoring rubric in appendix). Each response was scored separately for each area on a five-point scale ranging from zero to four. A total score was calculated by summing the scores from the three individual areas (mathematical knowledge, strategic knowledge and mathematical communication). Therefore, the highest possible total score was 12. Figures 3 and 4 are examples of responses to each version of the pre-post assessment. In both examples, students scored the highest possible score on the rubric for each of the three areas.

Figure 3: Probably Buttons



PROBABLY BUTTONS

PERFORMANCE TASK PROBLEM

Carmella has a small bag containing 4 green buttons, twice as many blue buttons as green buttons, 2 red buttons, and 6 times as many yellow buttons as red buttons. What is the probability of Carmella choosing a yellow button?
Explain your thinking and show your work.

If there are twice as many blue buttons than 4 green buttons, there are 8 blue buttons. If there are 6 times as many yellow buttons than the 2 red buttons, there are 12 yellow buttons.

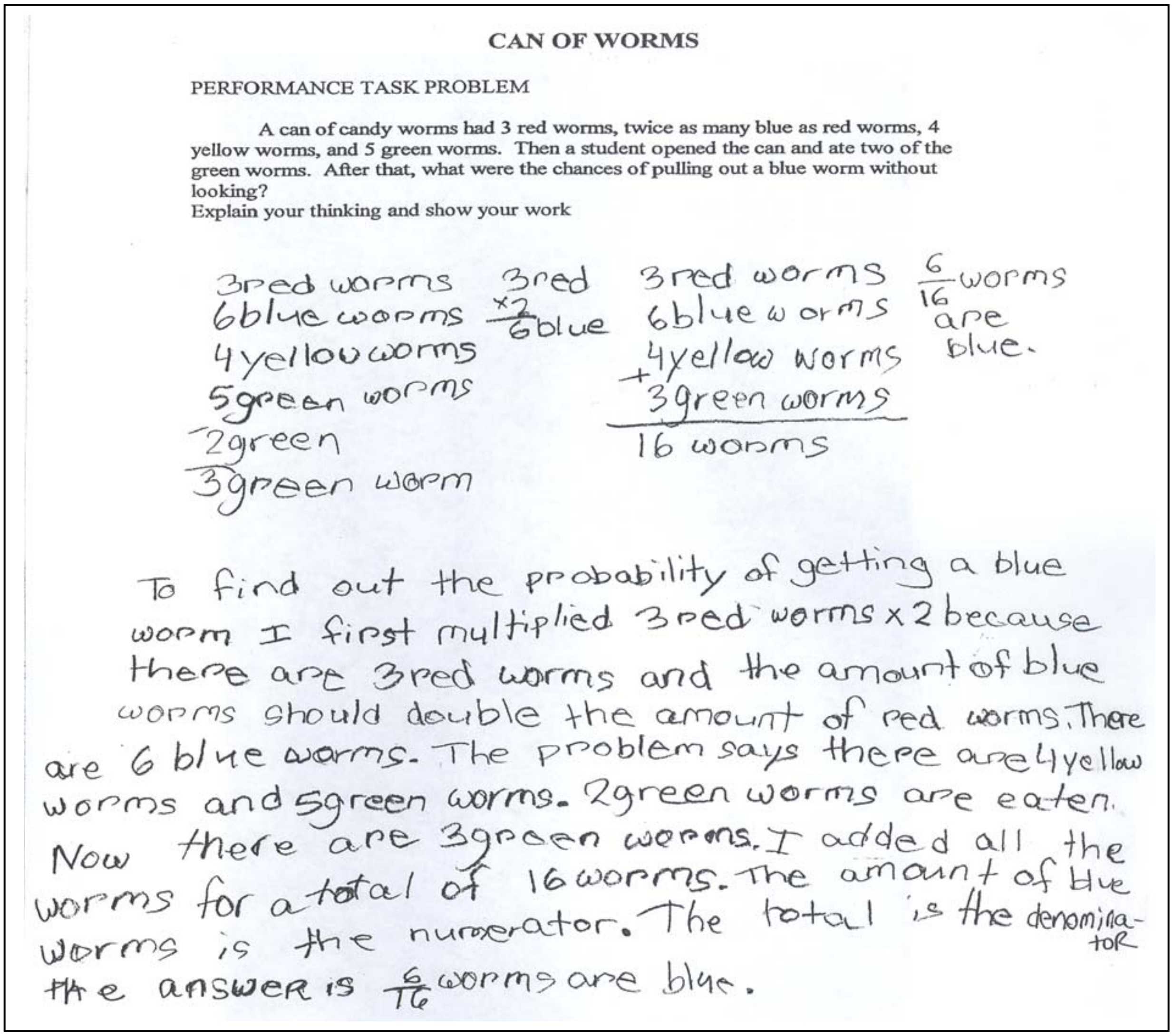
4 green buttons
8 blue buttons
2 red buttons
+12 yellow buttons
26 buttons

12)26 = .2, −24, R2

about ½ (little less)

The probability of picking yellow is 12/26 or a little less than a ½

Figure 4: Can of Worms



CAN OF WORMS

PERFORMANCE TASK PROBLEM

A can of candy worms had 3 red worms, twice as many blue as red worms, 4 yellow worms, and 5 green worms. Then a student opened the can and ate two of the green worms. After that, what were the chances of pulling out a blue worm without looking?
Explain your thinking and show your work

3red worms
6blue worms
4yellow worms
5green worms
-2green
3green worm

3red
x2
6blue

3red worms
6blue worms
4yellow worms
+3green worms
16 worms

$\frac{6}{16}$ worms are blue.

To find out the probability of getting a blue worm I first multiplied 3red worms x 2 because there are 3red worms and the amount of blue worms should double the amount of red worms. There are 6 blue worms. The problem says there are 4yellow worms and 5green worms. 2green worms are eaten. Now there are 3green worms. I added all the worms for a total of 16worms. The amount of blue worms is the numerator. The total is the denomina-tor the answer is $\frac{6}{16}$ worms are blue.

The rest of this section describes the sample of students who took the performance assessment, descriptive results for TL=MS students, and results of multilevel analyses used to determine the effect of TL=MS and to isolate the impact of teacher practices on students' performance.

Demographics of Students Taking Pre-Post Assessment

The sample for the current study consisted of 400 students enrolled in Bronx public schools (District 7, 9, 10, 11 and 12) for at least one year during the TL=MS study. This included 326 TL=MS students who received math instruction from 12 TL=MS teachers during the 2002-03 school year and a comparison group of 74 students in three classrooms who did not receive instruction from a TL=MS teacher. The classrooms invited by project staff to participate in the performance assessment were selected to represent a range of districts and schools and were similar to the overall population of TL=MS students in terms of race/ethnicity, ELL status, free/reduced-price lunch status, and prior academic achievement. Comparison classrooms were invited to participate after project staff identified TL=MS classrooms and were selected because their students' demographic characteristics were similar to the TL=MS students. Despite small monetary incentives, participation by comparison classrooms was disappointing with only three agreeing to participate.

Table 13 describes the characteristics of the 220 TL=MS students who took both the pre and the post-test. About half the students in the sample were girls, and two-thirds were fourth graders at the time of assessment. Nearly 61% of the students were Latino/Hispanic. In terms of prior achievement, half (50.2%) the students scored at proficient or advanced levels (levels three and four) on the spring 2002 standardized math test. The fourth graders' average scale score on the prior year's math test was 606, translating to the upper limit of level two or "basic performance." The fifth graders' average scale score on the prior year's math test was 654, translating to level three or "proficient performance."

Nine percent of the students in the TL=MS group were classified as special education students, and 21% of students were considered English language learners (ELLs). Almost all the students in the sample were eligible for free or reduced-price lunch.

Table 13: TL=MS Student Characteristics
(n=220)

	Percent
Gender	
Female	50.7%
Grade	
4th	69.4%
5th	30.6%
Race/ethnicity	
Latino/Hispanic	60.8%
African American	32.2%
Asian	5.0%
Other*	2.0%
English Language Learners	20.7%
Special Education Students	8.8%
Free or reduced-price lunch	95.9%
Scoring 3 or 4 on prior year (2002) mathematics test**	50.2%

*Includes Native American, Pacific Islander, White, etc.
**21% of TL=MS students were missing prior year's math test score because of ELL status.

Performance Assessment Results for TL=MS Students

The following section presents tables and graphs of performance assessment scores at baseline and follow-up and the changes in scores over time for fourth and fifth graders who were enrolled in mathematics classrooms headed by TL=MS teachers. Overall, the majority of TL=MS students experienced improvement in assessment scores from pre- to post-test. Table 14 displays the mean pre- and post-scores for all TL=MS students taking both tests. The highest possible score on each subarea was four. The highest possible total score was 12. Total scores ranging from 0 to 3.99 were deemed minimal performance, those ranging from 4 to 8.99 were deemed adequate performance or on grade level, and those ranging from 9 to 12 were deemed strong performance or above grade level. Student subscores for mathematical knowledge, strategic knowledge, and mathematical communication all increased, with the TL=MS group mean for total assessment score improving from 4.49 at the pretest, to 8.65 at the post-test (total math scores ranged from 0 to 12). A pretest mean of 4.49 represents the very low end of adequate performance, whereas the post-test mean of 8.65 represents the very high end of adequate performance.

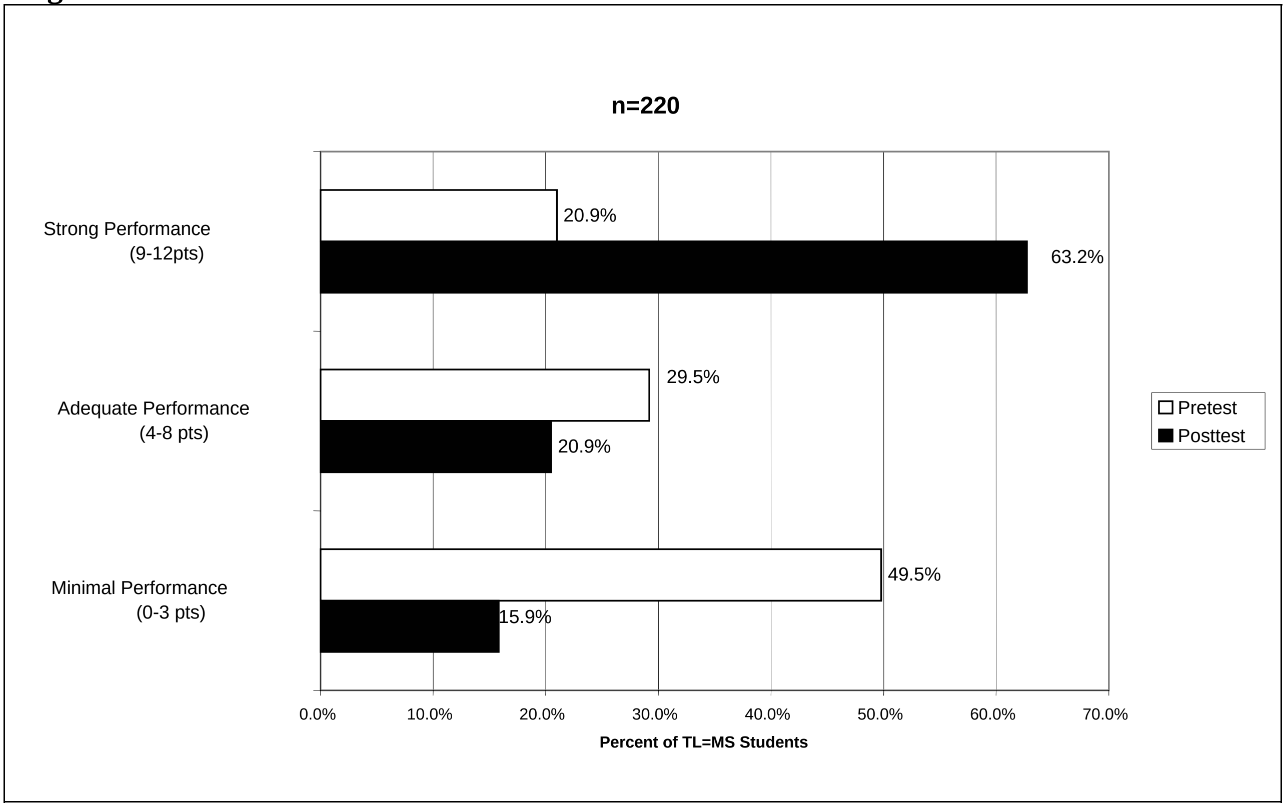
Table 14: Mean Pre- and Post-test Scores for TL=MS Students Taking Both Tests
(n=220)

	Mean Pre Score	Mean Post Score
Mathematical knowledge	1.41	2.84
Strategic knowledge	1.63	2.96
Communication	1.45	2.84
Total score	4.49	8.65

Overall Change in Results from Pre- to Post-test

As Figure 5 illustrates, the majority of TL=MS students reached adequate or strong performance on the math post test (20.9% of TL=MS students exhibited adequate performance, 63.2% of TL=MS students exhibited strong performance).[7]

Figure 5: Total Math Pre- and Post-Scores TL=MS Students



[7] Total score of 0-3.99 = minimal performance, 4-8.99 = adequate performance, 9-12 = strong performance.

The percentage of TL=MS students with strong performance in the assessment tripled between pre- and post-tests. Conversely, the group of students with minimal math performance decreased from half of all tested students to only 15.9 percent of all tested students.

Gains or Losses from Pre- to Post-test

The data in Table 15 illustrate the increases in scores from pretest to post test for TL=MS students. Nearly 9 out of 10 TL=MS students (88.2%) increased their total scores, and between 72% and 79% showed an increase in specific areas.

Table 15: Change in Performance from Pretest to Post-test TL=MS Students
(n=195)

Impact Area	Percent who showed improvement
Mathematical knowledge	79.0%
Strategic knowledge	71.8%
Written communication	75.9%
Total score	88.2%

*Note: The 25 TL=MS students with perfect pretests are not included in this analysis.

Analyses by subgroup (gender, race/ethnicity and ELL status) showed some differences in performance (see table 16). Boys showed greater increases than girls, and Latino and ELL students showed greater gains than African Americans and non-ELL students. However, multilevel analyses revealed that once pretest scores and teacher characteristics were controlled for, the differences were not statistically significant.

Table 16: Mean Differences between Pre- and Post-test Performance for Subgroups of TL=MS Students

		Mean Difference Between Pre-and Post-tests			
			Subscore		
	N	Mathematical Knowledge	Strategic knowledge	Communication	Total Score
Gender					
Female	101	1.60	1.52	1.59	4.72
Male	92	1.78	1.59	1.72	5.10
Race/Ethnic					
African American	54	1.72	1.48	1.42	4.63
Latino	112	1.81	1.75	1.89	5.46
Language					
ELL	43	1.83	1.86	2.18	5.88
Non-ELL	150	1.64	1.47	1.50	4.62
Overall	195	1.67	1.54	1.65	4.82

Note: The 25 students with perfect pretest scores are not included in this analysis. All differences between groups are non-significant as determined by multilevel analysis.

Multilevel Analyses of Student Assessment Data

Similar to most studies of student achievement, data collected for this study are nested. That is, students are located within classrooms and teachers. The teacher characteristics for students in each classroom do not vary. Therefore, data collected from students are not independent from the classroom/teacher within which they are located. Given this inherently nested nature of the student assessment data, we conducted multilevel analyses to further investigate the impact of TL=MS on student performance and the relationship of classroom practices to student outcomes on the assessment. Multilevel analyses are currently the only method that allows us to statistically link the effects of teacher or classroom-level characteristics with the achievement gains of their students (Raudenbush & Bryk, 2000). A summary of the multilevel analyses results is presented below. A full discussion of the multilevel analyses and results is presented in the appendix.

To investigate classroom practices related to student achievement, TL=MS teachers and the comparison non-TL=MS teachers who administered the pre-post student assessment also completed a survey of their practices. From the survey data, we constructed two scales of practice—traditional and reform practices. Three items comprised the traditional practices scale and 18 comprised the reform practices scale.

- The **traditional practices scale** asked teachers to reflect on how often students listened to a lecture, completed drill-type worksheets, or were given multiple-choice tests.
- The **reform practices scale** asked teachers to reflect on how often they used more activity-based instructional strategies: having students discuss solutions to mathematics problems with other students, having students work and discuss mathematics problems that reflect real-life situations, having students solve mathematics problems in small groups or with a partner, using student portfolios to assess student progress, or using students' written responses for assessment purposes.

Differences between the TL=MS and comparison groups on the average frequency of use of traditional or reform teacher practices were not statistically significant, which is not surprising given the small sample size (n=15). However, the direction of the difference for the traditional scale items in particular favors the TL=MS group. That is, TL=MS relied less on these types of classroom strategies relative to the comparison group. This finding supports findings from the comparison of TL=MS teacher survey data to the NAEP national survey data.

Summary of Multilevel Analyses

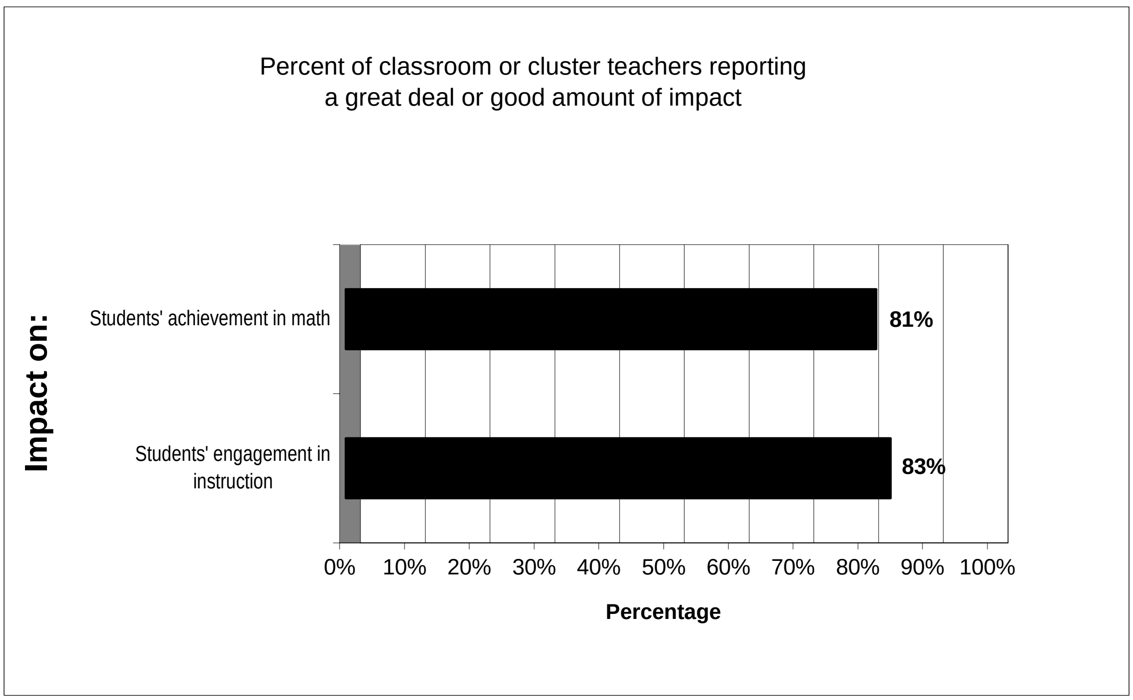
Multilevel analyses showed that TL=MS had an impressive positive and statistically significant impact on students' overall scores and the mathematical communication subtest compared with non-TL=MS students. When controlling for pretest scores and relevant teacher characteristics, TL=MS students outperformed non-TL=MS students by an average of 3.29 points (out of 12). On the communications subscale alone, TL=MS students outperformed non-TL=MS students by an average of 1.2 points (out of 4). Given the emphasis of TL=MS on mathematical communication (e.g. using "accountable talk," asking students to explain how they arrived at their answers), it is not surprising that students made the most gain in this area of the assessment. Although the effect of TL=MS was not statistically significant for the other subareas, they showed the same pattern of a positive effect. Non-significance is likely a result of the small sample size.

The impact of specific teacher practices on student outcomes is much less clear. Analyses indicated that more frequent use of traditional and reform practices are associated with higher post-test scores. However, much of the variance in post-test scores remain unexplained, indicating that the survey of teacher practices used in this study was either not sensitive enough to reveal differences in frequency of use of strategies for TL=MS and non-TL=MS teachers, or did not measure key teacher-level variables associated with student outcomes. The small sample size (n=15) also contributes to the limitations of this model in determining the relationship between practices and student outcomes (see Appendix A for full discussion of multilevel analyses results).

Teacher Reports of Impact on Students

As shown in Figure 6, using a five-point scale (great deal, good amount, some, a little, none), most teachers reported that TL=MS had a great deal or good deal of impact on their student's achievement (81% of respondents) and engagement in instruction (83% of respondents).

Figure 6: Teacher-Reported Impact of TL=MS on Student Engagement and Achievement



Note: Respondents to each item used the following five-point scale: a great deal, a good amount, some, a little, none. The number responding to each item ranged from 92 to 93.

These same data are disaggregated by grade level in Table 17. Just as middle-grade teachers reported less program impact on their comfort level and understanding of math, they were also more likely to report less program impact on their students' engagement in math instruction and achievement in mathematics; however these differences were not statistically significant. Just under three-fourths (72.2%) of grades 7-9 teachers reported a great deal or good deal of impact on student engagement, whereas 87.8% of K-3 teachers felt their participation in the program would have an impact on their students' engagement. Survey and interview data support the finding of a positive effect of TL=MS on student achievement, as described earlier in this report.

Table 17: Reported Impact of TL=MS on Student Engagement and Student Achievement by Grade Level (Classroom and Cluster Teachers Only)				
		As a result of participation in TL=MS, percentage of teachers reporting a great deal or good deal of impact		
Area of impact	All teachers n= 92 to 93	K-3 teachers n=49*	4-6 teachers n=52*	7-9 teachers n=18*
Students' engagement in mathematics instruction	82.8%	87.8%	84.6%	72.2%
Students' achievement in mathematics	80.6%	79.6%	82.7%	70.6%
*Note: Teachers could have indicated that they were teaching more than one grade level. Differences were not statistically significant using chi-square test.				

Teacher interviews substantiated the finding that TL=MS had an impact on student engagement. Teachers found that strategies such as cooperative group work, hands-on activities, and guided discovery increased student engagement in mathematics and led to greater understanding of mathematical concepts. Observing students succeed in mathematics was very satisfying for many teachers, especially when their success was related to the implementation of these new teaching methods, as the following quotes illustrate:

> *I used to hear kids say, "I hate math!" but I don't anymore. The kids ask me where I was when I miss a day. In one class, they cheer when I walk in the door!*
>
> *The kids really enjoy math and look forward to it. They think they're playing* [during math] *and don't realize they are learning. Other teachers see the reaction and are getting involved because they see how excited the kids get.*
>
> *I think kids are much more excited—they don't fear math because there are so many little tricks that I now have in my hat and I always pass them along. For example, I showed them how learning their multiplication facts for three will help them with the facts for six and nine. They love coming here.*

Comments on the final teacher survey also supported these reports. For example, one teacher wrote that TL=MS "helped me to explain and analyze math problems in various ways instead of just one. I believe this helped my students understand more."

One principal also noted an increased engagement in mathematics in his school:

> *The kids are more into it* [math]. *They argue over the best way to do it and defend their way of doing it.*

Summary and Conclusions

> *The most helpful aspect was listening and learning from the instructors and my colleagues on various different ways to finding solutions for a problem. Also having a math cohort to ask specific questions when perplexed, just as support, or listening to each others' ideas.* (TL=MS teacher participant)

> *I started this program as a new teacher, not wanting to become a leader—just wanting to learn how to teach better. Boy did I learn a lot of things for which I am very grateful. Most of all I learned that although I may not be an expert; I do have something to contribute.* (TL=MS teacher participant)

Ball and Cohen assert that teachers need serious and sustained learning opportunities at every stage in their career (Feiman-Nemser, 2001). **Findings from this study show that teachers at all levels of their career (novice, experienced and veteran) benefited from those types of opportunities through TL=MS**. Periodic surveys of teachers over the five years of the evaluation, classroom observations, teacher interviews, and results of a pre-post student assessment paint a powerful picture of how teachers transformed their practices and approaches to mathematics instruction and the positive effect it had on student achievement.

The impact of TL=MS on teachers was deep—affecting their comfort level with teaching mathematics as well as their content and pedagogical knowledge around teaching mathematics. Many TL=MS participants entered the project as self-described math "phobics" and with limited knowledge both of mathematics content and constructivist approaches. They emerged from the project more confident and competent in their understanding of mathematics and their skills as a teacher leader. Teachers also reported increasing the frequency with which they used reform-based practices shown to be effective in improving student achievement (Hamilton, et al., 2004). They also reported decreasing the use of traditional practices and attributed these changes in practice to their participation in TL=MS.

These finding were supported by comparisons with mathematics teachers nationwide, which showed that TL=MS teachers were much more likely to use reform practices emphasized by TL=MS and less likely to use traditional practices. The differences between TL=MS and comparison teachers were statistically significant. Further, longitudinal survey data show that teachers—up to five years after first participating in TL=MS—continued frequent use of the reform-based approach espoused by the project, indicating that TL=MS was sustained—at least at the classroom level.

More than just increasing or decreasing use of specific practices, teachers described changes in their whole approach to mathematics instruction. Teachers attributed much of this shift to the powerful impact of the summer institute and the value of the work conducted by teacher consultants (TCs). TCs supported teachers and schools in a variety of ways, including conducting observations and providing individual feedback to teachers, modeling and demonstrating lessons, and participating in team meetings and planning sessions.

In addition, many teachers spoke to the benefit of working and learning with a collegial group of TL=MS participants. TL=MS gave teachers a community and network through which they could share ideas, learn from each other, problem-solve, and support each other. These positive findings held for teachers at different levels of their career (novice, experienced and veteran), different grade levels, and for each of the projects' three cohorts.

The impact of TL=MS on students was equally impressive. **The majority of teachers and administrators reported that they saw a great deal or good amount of impact on students' engagement in mathematics instruction and achievement**. These findings were substantiated by an open-response, pre-post performance assessment completed by students in a sample of TL=MS and non-TL=MS classrooms.

TL=MS students saw large gains between pre- and post-assessments of their mathematical knowledge, strategic knowledge, and mathematical communication. Nearly nine out of 10 TL=MS students showed increases in their total scores, and between 72% and 79% showed an increase in a specific sub-area (mathematical knowledge, strategic knowledge and mathematical communication). Multilevel analyses showed that, controlling for teacher characteristics and pretest scores, TL=MS classes gained an average of 3.29 points more (out of 12) than non-TL=MS students on the post-test. This is a very strong finding, especially given the small sample size (12 TL=MS classrooms and 3 non-TL=MS classrooms). In addition, the positive effect of TL=MS held for a diverse group of low-income students, showing no differences by gender, race or English language learner status.

In conclusion, findings from multiple sources, including multilevel analyses of pre-post student assessment data, converge to provide strong evidence that TL=MS provided teachers with a deep and lasting professional development experience, which had an impact on teachers' practices and attitudes about teaching mathematics as well as on student achievement. This study is one of a small number studies that have looked beyond impact of professional development on teachers to investigate the impact on student outcomes (Killion, 1998), and supports findings from other studies that have shown a connection between high-quality, sustained professional development opportunities, reform-based practices, and student achievement (Garet et al., 2001; Hamilton et al., 2004; Wenglinsky, 2000).

We were not able to determine which specific teacher practices had an impact on student performance. This was likely due to the small sample of classrooms and survey measurement error. It is possible that the success of TL=MS teachers in raising student performance is related to the interactive process characteristic of effective teaching and student learning—teachers determine which practices to use and how to use those practices based on the needs of their students and the context of the instruction. Further, the quality or successful use of any of the particular strategies may vary across teachers.

Our survey did not capture these nuances of teacher practice. These issues warrant additional study to further our knowledge about the relationship between teacher practices and student achievement and how professional development models can most effectively support teacher development and ultimately, student achievement.

References

Feiman-Nemser (2001). From Preparation to practice: Designing a continuum to strengthen and sustain teaching*, Teachers College Record*, Vol 103, No 6, pp. 1013-1055.

Fennema and Frank (1992.) *Handbook of Research on Mathematics Teaching and Learning.*

Hamilton, L., McCaffrey, D., Stecher, B., Klein, S., Robyn, A., and Bugliari, D. (2004). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Education Evaluation and Policy Analysis.*

Heibert and Carpenter (1992). Learning and teaching with understanding. In *Handbook of Research on Mathematics Teaching and Learning*, edited by Douglas A. Grouws, Reston, VA: National Council of Teachers of Mathematics.

Killion, Joellen (1998). *What Works in the Middle: Results-Based Staff Development*. Oxford, OH: National Staff Development Council.

Lehman College (2000). *Teacher Leaders for Mathematics Success (TL=MS). Annual Report to the National Science Foundation* (July 1, 1999- June 30, 2000). New York: Author.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics.* Reston, VA: Author.

Wenglinsky, Harold (2000). *How teaching matters. Bringing the Classroom Back into Discussions of teacher quality.* Princeton, NJ: Educational Testing Service.

Appendices

Appendix A: Description of Multilevel Analyses

The purpose of this portion of the TL=MS study was to a) investigate differences in traditional teacher practices and reform practices in the classroom between a sample of 12 TL=MS teachers and three comparison teachers; b) attempt to link these differences in classroom practices to changes in student outcomes on a mathematics assessment; and c) investigate the overall impact of TL=MS on student outcomes. Due to absenteeism or other factors during the evaluation phase of the project, not all students took both the pre- and post-test assessments. Thus, the multilevel analyses presented here focused only on those students who had complete data on both administrations of that assessment.

Multilevel analyses allow us to statistically link the effects of teacher- or classroom-level characteristics with the achievement gains of their students. Although these analyses are statistically sophisticated, their interpretation can be quite straightforward. Multilevel analysis in studies of children within classrooms or other natural groups involves the notion of variability. For all students within the same classroom, teacher characteristics linked to those students never vary; they all have the same teacher. However, between classrooms, teacher characteristics do vary; one of these characteristics is, of course, whether a particular classroom teacher participated in the TL=MS program or was part of the comparison group. Another characteristic is the frequency with which that classroom's teacher reports using particular types of classroom practices or activities during mathematics instruction. The multilevel models let us look for overall differences in student outcomes across the 15 teachers in this study. Although multilevel models are at heart simple regression models, they are referred to as multilevel, or nested, because the student-level data is nested within the teacher-level data.

Multilevel approaches are currently the only method for statistically connecting student data to teacher practice, and they adjust for the inadequacies of traditional statistical methods where the naturally occurring hierarchy is ignored (Raudenbush & Bryk, 2000). The sample size of teachers for the current study is small ($n = 15$), but if teacher classroom effects exist, the multilevel approach is the best available strategy for recognizing them. The power of all statistical tests rests in part on the size of the effect and is influenced by sample size; in the current analyses, only large differences across teachers are likely to be discerned. The validity of these findings or our ability to attribute identified effects to the TL=MS program is based on the assumption that any differences between TL=MS and comparison teachers has been adequately captured by the measured teacher variables. It remains possible, as with any statistical analysis, that potential external factors not identified from the data collection instruments used in this study may also exert an influence on students' performance in mathematics. Thus, these results are not causal, but informative of general directions or trends suggested by the design of the study and the collected data.

Clustered data inflates the statistical significance levels for those variables that are inherently at a higher level of analysis in the model (Raudenbush & Bryk, 2000). For the aggregate analysis, those variables include TL=MS and the version of the performance test given, as well as their interaction; all three of these were found to be significant in the aggregate analysis. With clustered data, there is no methodology in an aggregate single-level model for adjusting for variability in scores *across teachers*. A more reliable approach toward investigating an overall TL=MS effect would be to adjust the analysis for the fact that students are clustered within their teachers. Further, since the TL=MS program is directed towards teachers rather than individual students (i.e., the unit of assignment of the intervention is "teacher"), investigating stability in the TL=MS effect across *classrooms* (rather than across all students aggregated together) becomes critical to our understanding of how specific teacher characteristics or practices emphasized through the TL=MS intervention might be contributing to student outcomes. Multilevel methods allow us to adjust for clustering and focus on teacher effects and their impact on student outcomes.

Two versions of the pre-assessment were given to students by classroom; assignment was reversed at the post-test. In order to control for any potential differences in difficulty due to the version of the test, the pre-assessment version was included in the statistical analyses.

Possible interaction effects between the covariate pretest scores, the version of the test taken, gender, racial identity, and ELL/LEP with the treatment variable of TL=MS were investigated; interaction effects between each of the categorical variables and the covariate of pretest score were also tested. A significant interaction effect was found between TL=MS and the version of the test. No other interaction effects were found between TL=MS and any of the independent variables or the covariate pretest score, nor between any of the demographic variables and the covariate pretest; thus only the interaction between TL=MS and the version of the tests were included here, as well as in subsequent models.

Thus, these results seem to suggest that, after controlling for all demographic characteristics and pretest performance, TL=MS students who took performance version B at the pretest (thus, A at the post-test) tended to outperform the comparison group on post-test performance. For those students who took version A at the pretest (thus, B at the post-test), TL=MS students tended to perform less well on the post test relative to the comparison group. However, it should be noted here that only one teacher in the comparison group used version A at the pretest compared with five of the TL=MS teachers.

As mentioned previously in the introduction section, one problem with the aggregate regression approach to the analysis of the post-test scores is that students are clustered or nested within their teacher.

Teacher Characteristics

Prior to conducting the multilevel models, differences between classrooms were investigated solely at the teacher level. Teacher characteristics important to the TL=MS intervention involve mathematics instruction practices in the classroom. The TL=MS study used a teacher survey to identify variation in classroom practices between TL=MS teachers and the comparison group of teachers. Of most importance was the development of two scales of teacher practices. Teachers were asked the frequency with which they used different strategies for teaching or assessment of student progress on 24 key items. Previous factor analyses, reliability assessments, and scale derivations led to the development of two scales reflecting (a) traditional practices and (b) reform practices. Scale scores were determined as the average of the item scores for the items included in each scale.

The traditional practices scale asked teachers to reflect on how often students listened to a lecture, completed drill-type worksheets, or were given multiple-choice tests. Responses were selected as 1 = almost every day, 2 = one to two times per week, 3 = one to two times per month, and 4 = never or hardly ever. The reform practices scale asked teachers to reflect on how often they used more activity-based instructional strategies. These included, for example, having students discuss solutions to mathematics problems with other students, having students work and discuss mathematics problems that reflect real-life situations, having students solve mathematics problems in small groups or with a partner, using student portfolios to assess student progress, or using students' written responses for assessment purposes. Several items showed no variability and thus were not included in the scale.

The final scales for these analyses thus included three items on the traditional scale, and 18 items on the reform scale (refrmtB). Means and standard deviations for these two scales across the TL=MS and comparison groups are provided in Appendix Table 1. Both scales were determined to be reliable with alpha levels of .65 and .85, respectively.[8] The scales created for this study are similar to scales constructed to investigate the relationship between instructional practices and student achievement in the National Science Foundation's Systemic Initiatives program (Hamilton, et al., 2004).

Differences between the TL=MS and comparison groups on the average frequency of use of traditional or reform teacher practices were not statistically significant. However, the direction of the difference for the traditional scale items in particular favors the TL=MS group. That is, TL=MS teachers had a *higher* mean on the traditional scale, implying *less* reliance on these types of classroom strategies relative to the comparison group. It appears that from a solely quantitative assessment of the data (i.e., t-tests, means, and standard deviations) both the TL=MS and comparison group teachers tended to use reform-based practices with relatively equal reported frequency.

Appendix Table 1: Means (Standard Deviations) and Results of t-tests for the Tradition and Reform Practices Teacher Scales (n=15)

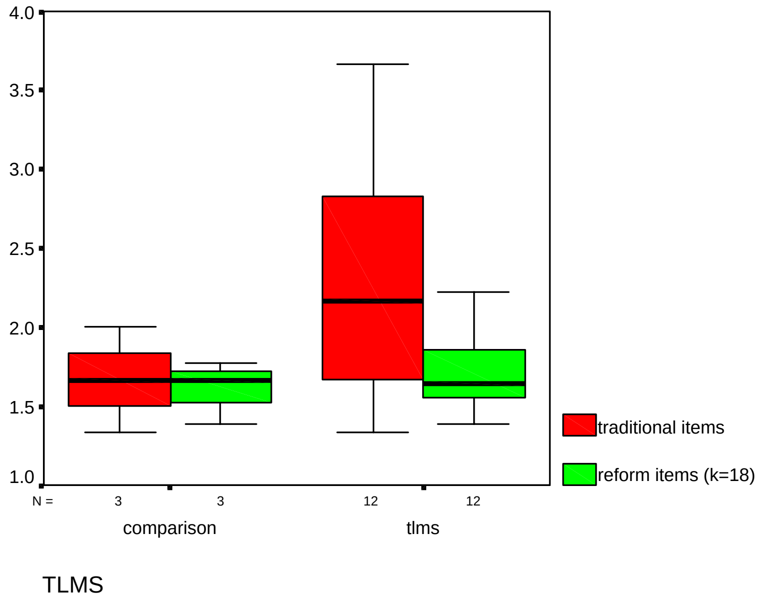
Group	Traditional (k = 3)	Reform (k = 18)
Comparison (n = 3)	1.67 (.33)	1.61 (.20)
TL=MS (n = 12)	2.28 (.69)	1.73 (.26)
sig.	.170	.468

However, the data for these comparisons is limited in terms of sample size. A graphical approach to investigating differences between the two groups may bring to light potential differences in a way that is not captured by statistical significance tests for t-tests with these sample sizes. Figure 1 provides a boxplot comparison of the two teacher practice scales between the comparison and the TL=MS teachers. From these boxplots, it is evident that, on the traditional practices scale, much greater range of frequency of use is observed for the TL=MS teachers. This can be interpreted to mean that there is a tendency for TL=MS teachers to rely less on daily use of traditional practices, relative to the comparison group teachers. From these boxplots we also see a slightly larger spread of frequency of use for the reform-based items for the TL=MS teachers, but not as large differences as on the traditional scale. One plausible explanation for this graphical depiction of the data is that TL=MS teachers may be using *several* reform strategies across a week or a month, in order to supplement their lessening reliance on traditional strategies. Thus, while the overall frequency of use is still somewhat similar to that of the comparison group (which is quite high), there may be more *exploration* of a variety of strategies

[8] Due to the small sample size for teachers who administered the pre-post assessment (n=15), the Chronbach's Alpha test was conducted on the entire sample of teachers who took the survey of instructional practices (n=129).

within TL=MS classrooms. Indeed, classrooms observations of TL=MS teachers did reveal a wide variety of reform strategies being used. However, we did not conduct observations of non-TL=MS classrooms and therefore cannot confirm that TL=MS teachers use a wider variety of strategies than non-TL=MS teachers. This possibility deserves attention in future studies.

Appendix Figure 1: Scale Score Differences for Traditional and Reform Classroom Practices between TL=MS (n = 12) and Comparison Group (n = 3)



Based on frequency distributions for the two scales across the TL=MS and comparison groups, 50% of TL=MS teachers and 100% of comparison group teachers had an average frequency of use of strategies on the traditional scale of at least once or twice per week (average ≤ 2 on the traditional scale). For the reform-scale strategies, 83.3% of the TL=MS teachers and 100% of the comparison group had average values indicating use of these strategies at least once or twice a week (average ≤ 2 on the reform practices scale). Thus, supporting the visual depiction of the data in the boxplots, there is more variability among the TL=MS teachers regarding average frequency of use of traditional approaches, but the two groups appear to be somewhat similar on the reform scale.

From this descriptive data, it is clear that TL=MS teachers are not relying on traditional classroom practices, such as having students listen to lectures, complete worksheets, and take multiple choice exams, as frequently as the comparison group teachers. However, from this data, it is not possible to discern what strategies these teachers might be relying on more frequently to replace the traditional practices. Further, the quality or successful use of any particular strategies may vary across teachers but was not measured in this survey; this element of classroom instruction cannot be identified through the frequency items and may be an area for additional investigation in later studies.

Statistical tests were also conducted to discern if there were particular items that contributed to differences in classroom practice between TL=MS and comparison group teachers, rather than averaging items on a frequency scale. To account for differing and sometimes limited dispersion among the items, both t-tests and chi-square analyses were conducted on an item-by-item basis. Four items were identified as statistically different in frequency of use between the two groups. Descriptive information on these items is provided in Appendix Table 2.

Appendix Table 2: Item Differences across TL=MS and Comparison Teachers (n = 15)			
Item	**TL=MS**	**Comparison**	
Traditional Practices			
Students complete worksheets for drill or practice	2.42 (1.24)	1.33 (.57)	p=.060[a]
Reform Practices			
Students discuss solutions to mathematics problems with other students	1.00 (.00)	1.67 (1.15)	p=.038[b]
Teacher uses short or long written responses to assess student progress	2.00 (.74)	1.00 (.00)	p=.04[c]
Teacher uses portfolios to assess student progress	3.33 (.80)	2.33 (.58)	p=.06

[a] Based on t-test for heterogeneous variances; [b] based on chi-square test statistic; [c] (chi-square result, p=.060)

Multilevel Models—Baseline Analyses

Initial analyses indicated that there is an observed positive TL=MS effect on students' post-test mathematics assessment scores for those students taking version B at the pretest and version A at the post-test, and that there are evident differences in classroom practices between comparison and TL=MS teachers that might contribute to understanding variability in these student scores. A series of multilevel analyses were conducted in an attempt to further clarify these findings and to discern if the TL=MS effect was stable when accounting for teacher variability. These analyses can also be used to help explain variability in adjusted post-test scores for students within comparison or TL=MS classrooms.

When data are nested in inherent clusters or groups, such as students nested within classrooms or within teachers, one natural measure of variability in the data is determined by the proportion of total variance in the outcome that is *between teachers*. This effect is often called the clustering effect, or the intraclass correlation coefficient (ICC). Regression analyses assume that all students in the sample are independent of each other and thus ignore the presence of any clustering effect. Yet in general, students within the *same* classroom tend to have correlated observations leading to dependence in the data and a positive non-zero value for the ICC. When the ICC is small, variability between classrooms or teachers would tend to be similar to the overall variability in scores. However, even when the ICC is small, significance tests from analyses that ignore the clustering of the data can be severely affected (Goldstein, 1995; Murray, 1998; Raudenbush & Bryk, 2002). When the ICC is large (greater than zero), the variability between classrooms or teachers can be modeled through multilevel designs. Multilevel analyses take into account the correlations among student scores induced by such natural groupings as teachers or classrooms and allow partitioning of the total variation so that student-level effects on the outcome can be isolated from teacher-level effects on the outcome.

For the current study, “teacher” serves as the clustering variable. The analysis has two levels. The level-one or student-level regression model identifies how the student variables of interest affect the post-mathematics-assessment outcomes. The independent variables here include the pretest score (covariate), and student demographic characteristics include gender, racial identity and identification as ELL/LEP. The level-two or teacher-level models examine how specific teacher characteristics impact on the intercept and slope estimates derived from the student-level regression model—that is, how teacher characteristics contribute to the effects observed at the student level. The level-two variables of interest here are TL=MS (intervention or comparison group), the scores on the traditional and reform practices scales, number of years teaching mathematics, the version of assessment given at the pretest (all teachers gave the opposite version at the post-test, and, as a result, only the nature of the pretest version needs to be controlled for), and the interaction between version of the assessment and TL=MS group.

In all the models described below, a level of significance of .10 is used for identification of variable effects. The sample size of 15 teachers is small enough to warrant such a consideration; a strict .05 level of significance might mask some important tendencies or trends.

The results of initial analyses revealed that none of the demographic variables are related to post-test score, but, as expected, the covariate (pretest score) is strongly related to student outcomes. Given the non-significant fixed effects for the demographic variables (and the fact that none of them contain residual variance worth explaining in further models), these variables were dropped from further analyses. Given the earlier result indicating differences across versions of the assessment, the version of the test and a term for potential interaction between test version and TL=MS was included in the results presented here for clarity, even though the multilevel analysis found that this interaction was not statistically significant.

Multilevel Models—Teacher Effects

The analyses now turn to the impact of teacher characteristics or practices that could impact on the adjusted (for the covariate) post-test scores. A collection of six variables were included as teacher-level predictors of student-level outcomes:

- TL=MS (intervention=1 or comparison=0)
- Years of teaching mathematics (1 = less than one year to 6 = more than 20 years)
- Scores on the traditional scale (average of 1 = almost every day to average of 4 = never or hardly ever)
- Scores on the reform practices scale (average of 1 = almost every day to average of 4 = never or hardly ever)
- An indicator used to control for version of the mathematics assessment that was given to the students at pretest (version A=1, version B=0),
- An interaction term for version of the assessment by TL=MS group.

The pretest was centered at the grand-mean for all analyses. The general model used in these analyses is as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(pretest_{gmc})_{ij} + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(TLMS)_j + \gamma_{02}(tradtot)_j + \gamma_{03}(refrmtb)_j + \gamma_{04}(yrsmath)_j + \gamma_{05}(pfrmcode)_j + \gamma_{06}(tlmsform) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Appendix Table 3 presents results for this model for the total post-test score and for the mathematical communications subscore. Only the area of mathematical communication indicated teacher effects on the respective student post-test scores; thus, only the results for this subscore are presented.

Appendix Table 3: Results for Teacher-Effect Models for Total Post-test and Communications Subtest, with Interaction of TL=MS by Version of Assessment

Model	**Total PostScore**[a]	**Communications Subtest**[b]
Fixed Effects		
For Intercepts		
γ_{00} (intercept)	**12.92*** (p=.027)	**4.11*** (p=.035)
γ_{01} (years teaching math)	-0.61 (p=.292)	-0.24 (p=.226)
γ_{02} (TL=MS)	**3.29*** (p=.100)	**1.20*** (p=.076)
γ_{03} (traditional practices scale)	**-1.69*** (p=.069)	**-0.59*** (p=.073)
γ_{04} (reform practices scale)	-1.15 (p=.611)	-0.20 (p=.787)
γ_{05} (assessment version taken at pre-test)	2.97 (p=.276)	0.31 (p=.702)
γ_{06} (interaction of TL=MS by form)	-2.56 (p=.391)	-0.37 (p=.690)
For Covariate Slopes		
$\gamma 10$ (intercept)	**0.40**** (p=.000)	**0.32**** (p=.001)
Random Effects		
τ_{00} (variance(intercepts))	**3.00**** (p=.000)	**0.27**** (p=.000)
τ_{11} (variance (slopes for pretest))	**0.04**** (p=.034)	0.02 (p=.158)
σ^2 (Variance(r_{ij}))	9.23	1.37

Note: Fixed effects tested with t-test, random effects tested with chi-square test. Hypotheses for fixed effects are parameter=0 versus parameter n.e. 0; for random effects: variance=0 versus variance g.t. 0.

[a] Dependent variable is the total post-test score; covariate is total pre-test score.

[b] Dependent variable is the communications post-test score; covariate is communications pre-test score.

In the results presented in column two of Appendix Table 3, investigating overall average post-test score differences *across n=15 teachers*, there is a significant effect of TL=MS (p=.100) and a significant effect of teachers' frequency of use of traditional strategies (p=.069). Years of experience in teaching mathematics is not statistically significant, nor are any of the other effects in the model. The results also show that TL=MS classes average 3.29 points higher (γ_{02} = +3.29) than non-TL=MS classes on the post-test scores, after controlling for all other effects in the model. Further, as teachers' use of traditional practices within their classroom *decreases* (recall that frequency of use was reported as 1=almost every day to 4=never, so that 4 represents a lesser use of that kind of strategy), the estimate for their adjusted classroom mean *decreases* (γ_{03} = -1.69), controlling for all other effects in the model.

Thus, greater use of traditional strategies is associated with improved post-test scores relative to teachers who use these strategies with less frequency. The same pattern is evident for use of reform-based strategies, but the effect is not statistically significant. The effect of assessment version is positive (version A was coded as "1" and version B as "0") but not statistically significant; and there is no evidence of interaction between TL=MS and version of the assessment when these effects are appropriately treated as teacher variables rather than student-level variables (as they were in the aggregate regressions). That is, once the teacher-clustering effect is accounted for, no interaction between test version and TL=MS is found, and TL=MS has an overall positive effect on student outcomes.

No reliable interaction effects were found between TL=MS and either of the teacher practice scales (not reported in tables). Thus, according to this analysis, the TL=MS intervention is associated with increased scores on the post-test regardless of the version of the assessment, and decreasing reliance on traditional practices in the classroom seems to be associated with lower classroom means on the adjusted total post-test score. In addition, the use of reform practices as measured here does not contribute strongly to understanding variation in the adjusted post-test scores. However, according to the random effects portion of the results, there is considerable variation remaining among the post-test scores that is not captured or explained by the teacher-variables available for inclusion in these models (τ_{oo} = 3.00, p=.000). Overall, 24.62% of the variance in the intercepts (adjusted post-test scores) is accounted for by the collection of teacher variables in this model.

A similar pattern of results is observed when the subtest of mathematics communication is considered separately. Here, the effect of TL=MS is statistically significant (p=.076) at the α = .10 level, and is positive (γ_{02} = +1.20), suggesting that students of TL=MS teachers do tend to perform better on the post-subtest for communication, after adjusting for their pretest communication scores and other effects in the model. The effect of traditional teacher practices is again negative and statistically significant (p=.073); and the effect of reform practices, also negative as expected, does not contribute to the model. There is no interaction found between version of the assessment and TL=MS group. There is, however, considerable variation in the post-assessment scores that remains unexplained by the teacher-level variables available for inclusion in these models (τ_{oo} = 0.27, p=.000).

Appendix B: Survey Items on the Traditional and Reform Practices Scale

Reform Practices Scale (k=18)

How often do students in your class do each of the following? (almost every day, 1-2 times a week, 1-2 times a month, never or hardly ever)

1. Discuss solutions to mathematics problems with other students.
2. Talk to the class about their mathematics work.
3. Work and discuss mathematics problems that reflect real-life situations.
4. Work with manipulatives (e.g., color tiles, pattern blocks, multilink cubes.)
5. Solve mathematics problems in small groups or with a partner.
6. Explain how the arrived at their answers.
7. Write in mathematics journals or logs.
8. Investigate problems that have multiple solutions.
9. Provide extensions to mathematics problems.
10. Create rubrics to score their work.
11. Write at least a few sentences about how to solve a mathematics problem.
12. Use literature connections.

How often do you use each of the following to assess student progress in mathematics? (almost every day, 1-2 times a week, 1-2 times a month, never or hardly ever)

13. Individual or group projects or presentations.
14. Short written responses (e.g., a phrase or sentence) or long written responses (e.g., several sentences or paragraphs).
15. Portfolios.
16. Peer evaluation (students evaluate each other's work).

How often do you do the following to assess student progress in mathematics?

17. Use rubrics to score student work.
18. Have students apply rubrics to score their own or others' work.

Traditional Practices Scale (k=3)

How often do students in your class do each of the following? (almost every day, 1-2 times a week, 1-2 times a month, never or hardly ever)

1. Listen to a lecture from the teacher.
2. Complete worksheets for drill or practice.

How often do you use the following to assess student progress in mathematics? (almost every day, 1-2 times a week, 1-2 times a month, never or hardly ever)

3. Multiple choice tests.

Appendix C: TL=MS Scoring Guide to Problems

Score Level	Mathematical Knowledge	Strategic Knowledge	Communication
4	-You show a complete understanding of the ideas in the problem and the mathematics that is needed to solve the problem. -You use the correct mathematical terms and labels.	-It is clear that you understand all of the important parts of the problem. -It is clear that you have a plan for working out the problem and that you are able to work through this plan intelligently.	-You give a complete, well written explanation of the process that you need to solve the problem. -You answer all of the questions completely and clearly. -You include any diagrams, graphs, or charts where necessary.
3	-Your understanding of the ideas in the problems and the mathematics needed to solve it is almost complete. -Your computations are, for the most part, correct but there may be some small errors.	-You are able to identify most of the important parts of the problem and you show a good understanding of them. -Your solution process is almost complete.	-Your explanations and diagrams are nearly complete. There might be some information missing.
2	-You show some understanding of the main ideas of the problem and the mathematics needed to solve the problem. -Your work may contain some serious computational errors.	-You are able to identify some of the important parts of the problem. -You are able to show some evidence of a solution process.	-You are able to give some explanation but your ideas may be hard to understand. -Your work may include a diagram with some explanation of its parts.
1	-You show very little understanding of the main ideas of the problem and the mathematics needed to solve the problem. -You use some incorrect mathematical terms. -Your work contains some major computational errors.	-You use some outside information that is not important to the solution of the problem. -You are unable to identify important parts of the problem. -You use an incorrect strategy for solving the problem. -The process that you use to solve this problem is difficult for the teacher to identify.	-You give very little explanation of the process. -You leave our important parts of the problem. -Your diagram may be incorrect of your explanation of the diagram contains errors or is unclear.
0	-You show no understanding of the problem or the mathematics needed to solve the problem.	-You use unimportant outside information. -You are unable to identify any of the important parts of the problem. -You copy part of the problem, but you show no attempt at a solution.	-You fail to give a written explanation where asked. -Your explanation does not have anything to do with the problem. -Your diagram cannot be understood.

The Academy for Educational Development (AED) is an independent, nonprofit organization committed to addressing human development needs in the United States and throughout the world. As one of the world's foremost human and social development organizations, AED works in five major program areas: U.S. Education and Workforce Development; Global Learning; Global Health, Population and Nutrition; Leadership and Institutional Development; and Social Change. At the heart of all our programs is an emphasis on building skills and knowledge to improve people's lives.

The AED **Center for School and Community Services** is part of AED's U.S. Education and Workforce Development Group. The Center uses multidisciplinary approaches to address critical issues in education, health, and youth development. To achieve its goals, the center provides technical assistance to strengthen schools, school districts, and community-based organizations. It conducts evaluations of school and community programs while striving to provide the skills and impetus for practitioners to undertake ongoing assessment and improvement. The center also manages large-scale initiatives to strengthen practitioner networks and accelerate systems change and uses the knowledge gained from this work to advocate for effective policies and practices and disseminate information through publications, presentations, and on the World Wide Web.

AED is headquartered in Washington, DC, and has offices in 167 countries and cities around the world and throughout the United States. The Center for School and Community Services is in AED's office in New York City. For more information about the Center's work, go to the Center's website at www.aed.org/scs or contact Patrick Montesano or Alexandra Weinbaum, co-directors, at 212-243-1110, or e-mail sweinbau or pmontesa@aed.org.

Principal Offices

1825 Connecticut Avenue
Washington DC 20009-5721
Tel: 202-884-8000
Fax: 202-884-8400
www.aed.org

100 Fifth Avenue
New York, NY 10011
Tel: 212-243-1110
Fax: 212-627-0407
www.aed.org/scs

AED

Academy for Educational Development