1-1-2012

Making mathematics matter: professional development improving outcomes in high-poverty environments

Carolyn A. Siebers
Wayne State University,

MAKING MATHEMATICS MATTER:
PROFESSIONAL DEVELOPMENT IMPROVING OUTCOMES IN HIGH-POVERTY ENVIRONMENTS

by

CAROLYN A. SIEBERS

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2012

MAJOR: Educational Leadership and Policy Studies

Approved by:

____
Advisor Date

____

____

____

ACKNOWLEDGEMENTS

I would like to thank my family, friends, and colleagues AND all pair-wise combinations of the aforementioned groups who supported me as I travelled this path. All of the teacher data analyzed in this research study were collected during the four-year existence of a federally-funded, state awarded Mathematics and Science Partnerships grant titled, *“Project: Making Mathematics Matter”* (PM^3). Therefore, of special note are friends and colleagues who collaborated in the creation and implementation of PM^3 including my RESA colleagues Libby Pizzo, Frederica Frost, Nanci Mitchell; the STEM faculty from the University of Michigan-Dearborn including Roger Verhey and Adele Sobania, and PM^3 coaches Rosanna Formaro and Marie Vitale. As much as this PM^3 project was the work we do as mathematics professional developers, it was also based upon our shared belief that ALL students deserve equal educational opportunities – so let us work collaboratively in the highest-poverty areas to help the belief become more real.

Also, in my mind are the wonderful educators I met and learned from in our two treatment districts. The world in which they committedly work on a daily basis is so different from my educational experience. During our project, two of our participants died of a rare form of cancer and we wondered about health risks. While many of these teachers may have originally been somewhat lacking in mathematics background, they had an abundance of flexibility and resilience.

I would also like to thank my mother, Dorothy D. Siebers, for being my best cheerleader for 89 years

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Background to the Study

Much has been written about the importance of a good teacher in every classroom to give students equal educational opportunities to achieve their goals (Sanders & Horn, 1998; Darling-Hammond, 1996; Darling-Hammond & Barnett, 2006). Not only has research revealed that many teachers lack the needed expertise to guide good instruction, it also demonstrates great educational damage can occur to a student who has two unsatisfactory teachers in successive years (Sanders & Rivers, 1996; Sanders & Horn, 1998). While new and/or ill-trained teachers can be working in any district, research shows greater proportions of these teachers in high-poverty schools (Darling-Hammond, 1998; Darling-Hammond, 2000; Borman & Rachuba, 1999; Laczko-Kerr & Berliner, 2002; Lankford, Loeb & Wyckoff, 2002; Harris & Ray, 2003; Hill, Rowan, & Ball, 2005; Flores, 2007; Jacob, 2007). If the educational opportunities of students in these high-poverty schools are to be realized, it becomes vital for all teachers currently in these schools' classrooms to have the necessary, on-going, professional learning opportunities to gain the necessary expertise to become good teachers.

While the concept of continuing professional development for educators is not new, the thinking about such professional learning activities has changed dramatically. In the 1980s a teacher might be expected to annually file a document recording the attainment of sixteen hours of professional development outside of the school day. The required sum might come from a variety of unconnected activities such as taking a university course, attending an educational conference, or hearing a noted motivational expert. The goal of this exercise was to achieve the necessary hours; common wisdom held that measuring impact of such educational practices was not realistic due to the complexity of American schooling with its many intervening variables.

Over the last twenty years or so, a body of careful educational research within ordinary school settings has been growing. While the diversity among teachers and students makes discovery of straightforward, causal relationships in education unlikely, researchers have increasingly made use of large, national databases and well-designed studies to provide evidence about the impact of many educational practices on teaching and learning. Researchers have found strong evidence that professional development programs that result in teachers using effective teaching strategies are associated with improved student achievement (Slavin, Lake, & Groff, 2009). Professional development is the most effective tool for supporting in-service teachers in learning how to implement these focused practices (Shulman, 1986; Borman & Rachuba, 1999; Fetler, 1999; Loucks-Horsley & Matsumoto, 1999; Darling-Hammond, 2000; Harris & Ray, 2003; Loucks-Horsley, Love, Stiles, Mundry & Hewson, 2003; Hill & Ball, 2004; Hill, Schilling & Ball, 2004; Hill, Rowan & Ball, 2005; Miles, Odden, Fermanich, & Archibald, 2005; Hanushek, 2006; Darling-Hammond, Williamson, & Hyler, 2007; Blank, de las Alas, & Smith, 2008).

Among researchers, there is a growing consensus about characteristics of professional development that are associated with positive gains by teachers and/or students. Elements of these effective programs include being long-term and sustained, focusing on effective instructional strategies, probing student thinking and reasoning, and increasing teachers' content knowledge (Yoon, Duncan, Lee, Scarloss, & Shapely, 2007; Wei, Darling-Hammond & Adamson, 2010). Studies have begun to differentiate among particular content areas to study strategies that successfully help teachers learn about instruction in that discipline—such as reading, writing, science, social studies, or mathematics (Ball & Bass, 2000; Loucks-Horsley et al. 2003; Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009).

Current research on high-quality mathematics instruction and professional development can be traced to new constructs of mathematical content knowledge proposed by Lee Shulman in the late 1980s. Shulman and others began questioning the long standing dichotomy between subject matter professional development as represented by taking higher level college content coursework and pedagogical professional development as seen in participating in a content-free study of topics such as classroom management (Shulman, 1986, 1987; Wilson et al. 1987). In their work, a new type of mathematical content knowledge for teaching emerged which united both content and pedagogy and was referred to as pedagogical content knowledge (Shulman, 1986, 1987; Wilson et al.). This type of mathematical knowledge is different from the content knowledge required for a mathematician. One body is not superior to the other as offered in the infamous quote, "He who can, does; he who cannot, teaches." The type of deep mathematical knowledge required for each occupation is different. Just as an engineer needs unique mathematical knowledge and skills for solving complex engineering problems, a mathematics teacher requires unique mathematical knowledge and skills to facilitate the mathematical understanding of diverse learners in a classroom.

Subsequent researchers built on this work about the mathematical content knowledge for teaching and identified as crucial relationships between teachers' content knowledge, pedagogical knowledge, and high quality classroom instruction (Ball & Bass, 2000; Ball & Bass, 2002; Borko et al. 1992). Once such links were identified, investigators studied professional development programs designed to increase in-service teachers' content knowledge and to improve their classroom instruction (Wilson et al. 1987; Ball, 1992; Eisenhart et al. 1993; Loucks-Horsley & Matsumoto, 1999; Ma, 1999; Ball & Bass, 2000, 2003; Gusky, 2003; Loucks-Horsley et al. 2003; Borko, 2004; Hill, 2004; Banilower, Heck, & Weiss, 2007). Researchers have found characteristics of professional development for mathematics teachers to be linked to increased teacher knowledge. Three crucial elements identified for effective mathematics PD are: 1). project is long term, sustained, collaborative, school-based, linked to curricula, focused on student learning; 2). project targets teachers developing deeper understanding of math they teach; 3). project provides coaches to support teacher growth, implementation, and collaboration (Loucks-Horsley & Matsumoto, 1999; Hill, Schilling, & Ball, 2004). Analysis of research performed in large urban districts suggests another element whose importance may increase in settings with high rates of staff and student mobility—having whole school or whole district teacher participation in the professional development (Odden, Archibald, Fermanich & Gallagher, 2002; Miles et al. 2004).

The ultimate discrimination between effective professional development programs as opposed to ineffective ones must be the extent to which researchers are able to link a professional development program to increases in student achievement. Unfortunately, in mathematics there is a dearth of such studies. The Institute of Educational Sciences (IES) in the U.S. Department of Education (US DOE) recently performed an extensive review of studies about mathematics professional development and student achievement and found:

> Of the more than 1,300 studies identified as potentially addressing the effect of teacher professional development on student achievement in three key content areas, nine meet What Works Clearinghouse evidence standards, attesting to the paucity of rigorous studies that directly examine the link (Yoon et al. 2007).

To add to the limited research base about effective professional development in mathematics, the federal government through the U.S. Department of Education established a Mathematics and Science Partnerships (MSP) program. In 2003, each state began receiving annual federal funding to award competitive MSP grants to projects designed to examine the relationships between professional development strategies that effectively increase teacher content knowledge for teaching, improve classroom instruction, and ultimately raise student achievement. The stated focus of these MSP grant projects was improving mathematics achievement in low-performing, high-needs schools. This study examines data collected from the MSP professional development program, "Project: Making Mathematics Matter" which examined program impact on two, urban, high-poverty districts in Wayne, Michigan compared with a demographically similar district in another county that did not receive the PD.

Statement of the Problem

Good mathematics teachers delivering high-quality instruction are necessary for every mathematics classroom. This need is particularly great in high-poverty areas where sub-groups of students are inordinately impacted by inexperienced and/or ill-trained teachers. Unfortunately, the research base for aiding in the design of mathematics professional development programs linked to increased teacher content knowledge and improved classroom instruction culminating in higher student mathematics achievement is limited. There is a great need to increase the literature about effective mathematics professional development—especially programs for teachers in high-poverty schools where students' mathematics achievement is significantly lower than their peers—both within the U. S. and around the world.

A number of recent findings create new opportunities to contribute to this gap in the literature. While the results from the international mathematics comparisons might be seen largely as identifying a U.S. mathematics problem, they can also be viewed as giving more insight into possible system corrections for achieving better results. The video case study in the Third International Mathematics and Science Study (TIMSS) reports that U.S. teachers resort to giving students hints when they struggle with difficult tasks—rather than giving their students the tools they need to develop independence, persistence, and confidence in their mathematical reasoning ability (NCES, 1999). This finding indicates a need for professional development in which teachers learn about the nature of high-quality mathematics instruction and how to maintain high levels of student thinking in their classrooms. In addition, the disaggregation of the international mathematics comparison data reveals important information for narrowing the scope of efforts aimed at improving U.S. student performance. Researchers found that non-disadvantaged students perform comparably with high-performing countries, but that average U.S. achievement was lowered by scores of students from inadequately funded, high-poverty schools (Payne & Biddle, 1999). From this result, there is an obvious need to provide the requisite professional development for teachers in these low-scoring districts to help them acquire the mathematical content knowledge and instructional skills necessary to support the learning of their students.

Perhaps the most relevant additions to the literature are the more fully developed definition of the construct of mathematical content knowledge for teaching and a valid and reliable instrument to measure it. Studies of these elements provide further guidance about effective mathematics professional development designs and accurate tools to measure potential links between teacher professional development and student mathematics achievement. Without further evidence about effective professional development that helps inexperienced or ill-trained mathematics teachers in high-poverty schools increase their content knowledge for teaching and improve their classroom instruction, it is unlikely the persistent mathematics achievement gaps between groups of students based on racial and economic differences will be eliminated.

Purpose of the Study

Students in high-poverty schools have reduced access to high quality instruction and associated resources and often fail to realize the higher levels of mathematics achievement of their counterparts in wealthier districts. The purpose of this study is to add to the literature that examines potential links between teacher participation in intensive, sustained, well-designed mathematics professional development (PD) programs and increased teacher content knowledge, improved classroom instruction, and increased student mathematics achievement in disadvantaged districts. The ultimate goal is to increase the access of students in these high-poverty settings to higher quality instruction and to reach higher levels of mathematics achievement—to narrow the achievement gap.

"Project: Making Mathematics Matter" (PM³) was the name of the MSP grant from which the data in this study were collected. PM³ partnered with university mathematics faculty and high poverty districts to support in-service teachers of students in grades four through eight with an intensive program of mathematics professional development designed to deepen teachers' pedagogical content knowledge of mathematics, to improve classroom mathematics instruction, and to increase the mathematics achievement of students. The next section gives a brief description of PM³ design and content as background for the review of literature chapter.

Description of "Project: Making Mathematics Matter"

Project: Making Mathematics Matter was conceived in response to a call from the U.S. Department of Education's Mathematics and Science Partnerships (MSP) department to award funding through state agencies for research projects which might add to the literature about effective mathematics professional development programs for teachers in high-poverty settings. PM³ was a four-year, sustained, professional development program implemented in two, disadvantaged urban districts beginning in the fall of 2004 and continuing through the summer of 2008. The PM³ PD model was designed around research literature findings about critical characteristics for effective mathematics professional development—with special note to research within high-poverty environments. The first of the three major components selected from the research for inclusion in PM³ was mathematics institute courses focusing on teachers increasing their pedagogical, content knowledge for teaching; the second was content-focused, classroom, on-site coaching; and the third was the establishment of a whole-district, collaborative, professional learning culture for participants. This construct of collaborative practice began at mathematics institute sessions where teachers worked together in small groups to find solutions for rigorous tasks and was further established with regular, whole-district, grade specific meetings. These grade-level after school meetings were facilitated by a coach and colleagues had the opportunity to discuss instructional issues, student learning issues, and work together to identify instructional and/or assessment strategies to increase understanding of mathematics for their students.

The PM³ project chose a whole-school, whole district model and requested target districts allow all teachers who provide mathematics instruction for 4th through 8th grade students within the schools and districts—including special education, general education, and English Language Learners teachers—to participate in the program. Between the two districts, approximately fifty-five staff members taught mathematics in grades 4 through 8 and these teachers became the treatment group for Project: Making Mathematics Matter. With a quasi-experimental, matched comparison group design, the PM³ project identified about the same number of comparison teachers in a demographically similar district in another county. In the fall of 2004, PM³ initiated its intensive mathematics professional development project and intervention teachers began attending mathematics institutes, met their district mathematics coach at the first institute, returned to school with the on-site support of their coach, and started gathering monthly after school with the coach to have grade band discussions.

Teacher participants in PM³ attended two mathematics institute courses of thirty-hours each during the school year and another during the summer for a total of ninety hours per year of institute course-work focused on deeper understanding about the mathematics they teach. For development and oversight of this course work, mathematics faculty from the Center for Mathematics Education located in the University of Michigan-Dearborn took major responsibility. A sequence of mathematics institute classes was established to help participants deepen their mathematical content knowledge for teaching the important mathematics concepts within the various strands of learning identified by the state for students in their classrooms. Each institute course included at least thirty-hours of professional development and carried optional graduate credit to help participants work toward "highly qualified" status under NCLB. Mathematics Institute course work included the following:

1. Number, Operations, and Proportional Reasoning
2. Proportional and Algebraic Reasoning
3. Conceptualizing and Teaching Linear Functions
4. Geometric Reasoning
5. Implementing Standards-Based Mathematics Lessons
6. Data Analysis and Statistical Reasoning

Each of these classes examined one of the mathematics strands in the Michigan Grade Level Content Expectations for mathematics and the intent of the course work was to deepen the participants' knowledge of the mathematics they teach in each of the major mathematics strands.

Institute instructors modeled a constructivist approach and employed rich, mathematical tasks that allowed small groups of teachers to experience multiple ways of thinking about and solving conceptually important mathematical problems. Additionally, instructors called on groups to share their solution which facilitated teachers seeing and understanding a variety of solutions to the task in preparation for teaching the lesson in their own classroom. Institute instructors also debriefed their instructional decisions to make the teaching process used more transparent. The discussion following the completion of the task includes the types of questions that follow. Why do you think I chose these numbers? Why do you think I had this group share their solution before this group?

In addition to ninety hours of large group, collaborative, institute time throughout the year, participants also had a content-focused, mathematics coach with whom they had individual professional development time supporting classroom work with students. The coach attended the institute with the participants and worked with each person according to needs and requests to consider implementation of new instructional strategies. This coaching support took various forms including planning before instruction, modeling or co-teaching parts of a lesson, or discussing the implementation of a strategy. In this individual setting, participants received a minimum of another twenty hours of professional development support over the course of the year.

The third leg of the PM^3 professional development model was the monthly, collegial, after school meetings. While this component of PM^3 may have only added fifteen hours to reach an annual sum of about 125 hours of professional development per year for PM^3 participants, this feature provided the forum where whole system change could occur. The mathematics coach for the district called and facilitated monthly, collaborative, whole-district meetings for small groups of grade specific teachers. These grade level meetings served an extremely important function in the project model; they provided an opportunity for teachers to meet regularly over a sustained period of time to address issues of importance to their grade's curriculum, instruction, and student learning. A group of grade 4 teachers may have needed to have very different discussions than middle school teachers. Issues addressed include curriculum alignment and pacing guides, examining student work, writing common assessments, examining common assessment results, MEAP analysis, challenges posed by district circumstances, and more work on specific mathematics topics. The intention was for each of the grade level teams to develop into a collegial, collaborative, professional learning community.

Study of PM³ Data

This research study will be mixed-methods, ex-post-facto evaluation of data collected by the PM³ mathematics professional development project during its years of operation from fall 2004 through summer 2008. Quantitative analyses will be used to look for any differences in mathematical pedagogical content knowledge for teaching between treatment and comparison teachers, to search for any changes in instructional practices in treatment teachers' classrooms, and to study student scale scores from the treatment districts and comparison district. Qualitative analyses of focus group transcripts will be performed to add further depth and/or evidence for research findings. Evaluation tools for PM³ included the Learning Mathematics for Teaching (LMT) assessment for measuring teacher content knowledge for teaching, the Science and Mathematics Program Improvement (SAMPI) classroom observational rubric to rate teachers' instructional practice, the mathematics portion of the state Michigan Educational Assessment Program (MEAP) to measure student mathematics achievement, and focus groups to gather participants' perceptions about the PM³ program. These data will be analyzed to address the following three research questions.

Research Questions

- Are there significant differences in mathematics content knowledge for teaching between PM³ treatment group teachers and comparison group teachers as measured by pre- and post-LMT assessments over the period of the PM³ project?
- Are there significant changes in classroom instruction and practice of treatment teachers as measured by the SAMPI classroom observation protocol during their participation in the PM³ professional development project?
- Are there significant differences in grades 4 through 8 MEAP mathematics performances between students in treatment districts and students in the comparison district following participation in the PM³ project?

Limitations of the Study

The PM³ project and the data collected from it had a number of possible internal and external threats which could limit the ability to suggest the treatment was responsible for the results and/or the capacity to generalize its results to the larger population. The threats to the internal validity of the study include: selection; experimental mortality; history; and compensatory treatments. Although the project did not use random sampling, it also did not differentially select participants. Once the two treatment districts were selected, the districts were asked to have all teachers who work with students for mathematics in the targeted grades 4 through 8 participate. A factor that should also reduce the selection threat was the use of a matched comparison group from another county with approximately similar demographics to the treatment districts. Another concern at the beginning of the study was whether the treatment group or comparison group would differ in loss of participants in the project. Perhaps the most pressing internal validity worries were history and compensatory treatments. In the real-life world of education, it is not possible to make sure nothing—other than the PM³ treatment—happened to the treatment group throughout the course of the study. Similarly, with the pressure on all low-scoring districts to improve their mathematics state scores, there was no way to make sure the comparison group teachers did not receive mathematics professional development.

Some potential threats to the external validity of the project must also be considered. These include sampling bias, reactive arrangements, and/or interaction of history and treatment. The possible sampling bias is a consideration since the sampling was not random and there is a question about whether this project sample is representative of the population of grades 4-8 teachers in high-poverty, low-performing schools. Again, the fact that all teachers in two different high-poverty districts were selected might reduce this threat. When deliberating about the impact of the treatment, might it be possible that the positive results are simply a matter of the teachers' reactions to being studied? Such a result is also called a "Hawthorne Effect" after a study examining the impact of high or low lighting on work output found the workers produced more as a result of participating in the study—not related to the amount of lighting (Adair, 1984). Might such a "Hawthorne Effect" impact a four-year study? Probably the greatest threat to the PM³ program and the data it collected was the fact that the project had no control over many of the events that impacted the teachers and students in these districts during the research period. Districts have the authority to hire, fire, and lay-off teachers—as well as assigning teachers to classrooms, requiring teachers to fulfill district obligations, and re-organizing or closing buildings. In high-poverty, low-performing urban districts with declining enrollments, such changes are quite common as the following district information shows.

Table 1. Demographic information for the two treatment districts

District T1	District T2
Both districts 90 percent socio-economically disadvantaged students	
3 elementary schools	3 elementary schools
1 middle school	1 middle school
1 high school	1 high school
62 percent English Language Learners	99 percent African American
	2 homeless shelters near one elementary

Table 2. Challenges for treatment districts during PM³

Year 1	Year 2
*Closed middle school *District re-organized in November, lay-offs *4 new principals *Substitutes in many classrooms, many without teaching certificates	*District starts a new school for ELL students *School closing in January *4 new principals *Superintendent leaves *Assistant/curriculum director leaves
Year 3	**Year 4**
*Close school for ELL students *Re-organize grade configuration from 1-5 and 6-8 to 1-6 and 7-8. Sixth grade self-contained *2 new principals *Teacher lay-offs	*Absorb 300 students from a disciplinary alternative charter school during school year *Curriculum director leaves *4 new principals *Superintendent leaves

Although treatment districts T1 and T2 both had over 90 percent of their students identified as socio-economically disadvantaged according to free and reduced lunch statistics, there were many differences between these two, small urban districts as seen in Table 1. While one district served primarily African American students, the other was more diverse and a majority of their student population was English Language Learners. Table 2 shows the many challenges endured by these districts' staff and students over the four year duration of "Project: Making Mathematics Matter." These challenges underscore the importance of the whole-district approach adopted by the PM³ program.

Delimitations

This study focuses on mathematics professional development for teachers in two, low-performing, high-poverty, urban districts in Wayne County, Michigan. This research is intended to enrich the literature about characteristics of a professional development program that is effective in increasing teacher mathematical pedagogical content knowledge, improving the quality of classroom mathematics instruction, and ultimately increasing student mathematics achievement in such challenged districts.

Definitions of Terms

Adequate Yearly Progress (AYP) is an accountability measure required by the NCLB legislation that mandates states create rigorous and challenging performance standards for all students in the state and use valid and reliable measures to assess continuous growth in student proficiency to assure the proficiency of all students within the allotted time frame of the NCLB legislation.

Analysis of variance (ANOVA) is a statistical technique that considers data from two or more groups on one scalar dependent measure and analyzes these data to find if the means of these two or more groups are significantly different. There are a number of assumptions to be met by the data prior to application of any of the several forms of an ANOVA.

Common Core State Standards (CCSS) is a document designed to serve as the new, first-time, national set of standards for mathematics curriculum and instruction in the U.S. The CCSS document also includes literacy, writing, and reading standards. The CCSS has currently been adopted by the vast majority of U.S. states and territories and the CCSS is to serve as the basis for national mathematics testing. The CCSS has eight student Standards for Mathematical Practice that overarch the Standards for Mathematical Content for each grade level, kindergarten through grade 8 and then high school conceptual categories. The Standards for Mathematical Practice state what mathematically proficient students in all grades need to be able to do with the content they study and call for a huge paradigm shift in mathematics instruction toward deeper conceptual understanding. These eight practices are: 1). Make sense of problems and persevere in solving them; 2). Reason abstractly and quantitatively; 3). Construct viable arguments and critique the reasoning of others. 4). Model with mathematics; 5). Attend to precision; 6). Use appropriate tools strategically; 7). Look for and make use of structure; 8). Look for and express regularity in repeated reasoning.

Content Knowledge for Teaching (CKT) or Mathematical Knowledge for Teaching (MKT) refers to a combination of the deep content knowledge about the mathematics being studied, the pedagogical content knowledge about students and how to help them learn the mathematics, and knowledge about the sequencing of topics and the use of resources for an effective mathematics curriculum. This new construct of mathematical content knowledge—specific to teaching—was introduced by Lee Shulman in the late 1980s and further refined by researchers including Deborah Ball in the late 1990s and beyond.

Focus Groups are smaller groups of participants assembled to gather more in-depth information about their experiences in the program, in their classrooms, and with their students.

Institute of Educational Sciences (IES) is the U.S. Department of Education's agency for gathering relevant research and evidence of best practices in the areas of education research, evaluation, assessment, development and statistics.

Learning Mathematics for Teaching (LMT) assessment was designed by Heather Hill and Deborah Lowenberg Ball to measure the unique pedagogical content knowledge important in teaching mathematics, a construct of mathematical knowledge first proposed by Lee Shulman. The LMT uses items that pose tasks or situations that might arise in a classroom such as diagnosing student thinking, identifying student misconceptions, and flexibility in representing numbers and operations using materials or stories.

Mathematical Knowledge for Teaching (MKT) or *Content Knowledge for Teaching (CKT)* refers to a combination of the deep content knowledge about the mathematics being studied, the pedagogical content knowledge about students and how to help them learn the mathematics, and knowledge about the sequencing of topics and the use of resources for an effective mathematics curriculum. This new construct of mathematical content knowledge—specific to teaching—was introduced by Lee Shulman in the late 1980's and further refined by researchers including Deborah Ball in the late 1990's and beyond.

Mathematics and Science Partnerships (MSP) program is in the U.S. Department of Education and is administered by the Academic Improvement and Teacher Quality Program (AITQ) in the Office of Elementary and Secondary Education (OESE) under the No Child Left Behind Act of 2001, Title II, Part B. The goal of the MSP program is to improve academic achievements of elementary and secondary students in mathematics and science by increasing instructional quality.

Michigan Educational Assessment Program (MEAP) is the annual state test for Michigan students in grades 3 through 8 as required by the No Child Left Behind legislation. Before the fall of 2005, the state MEAP test was administered in February of the school year and mathematics was tested in two grades. Since the fall of 2005, the state test in mathematics is administered in October for students in all grades 3 through 8

National Center for Educational Statistics (NCES) is the primary U.S. Department of Education entity for collecting and analyzing data related to education.

National Assessment of Educational Progress (NAEP) is an assessment of what students in grades 4^{th}, 8^{th}, and 12^{th} grades know and can do in various subjects including mathematics, reading, writing, science, and social studies. The NAEP in its subject areas is administered periodically to samples of students across the nation and these results are often referred to as "The Nation's Report Card."

No Child Left Behind (NCLB) is the 2001 version of the reauthorization of the Elementary and Secondary Education Act passed by the federal legislature. The intent of NCLB is to improve elementary and secondary education with new accountability measures and consequences for schools and districts failing to meet the measures.

Profound Understanding of Foundational Mathematics (PUFM) is the essential knowledge required for effective mathematics teaching as defined by Chinese researcher Liping Ma. This is a similar to the CMT or MKT construct of mathematical knowledge described by American researchers Shulman and Ball.

Project: Making Mathematics Matter (PM^3) is the intensive, sustained, mathematics content-focused, professional development grant project which began implementation in the fall of 2004. This project was a federally-funded, state of Michigan administered Mathematics and Science Partnerships (MSP) grant awarded to Wayne Regional Educational Services Agency (Wayne RESA) with a Wayne RESA mathematics consultant as the principal investigator. The data collected during the PM project will be analyzed to test the research hypotheses.

Science and Mathematics Program Improvement (SAMPI) classroom observation tool asks a trained, certified outside evaluator to use the instrument's rubric to rate classroom instructional practices. Each sub-scale area, such as mathematics content, has 7 to 10 items to be rated before arriving at a rating level for quality of mathematics content in the classroom. The ratings are from 1 to 7 with a 7 representing the highest quality of instruction. The internal consistency and reliability of the SAMPI observation system show all Cronbach's alphas above .74. Face validity has been established by the work of the state of Michigan, higher education, and the statewide network of mathematics and science centers identifying a set of classroom performance objectives to be assessed.

Science, Technology, Engineering, and Mathematics (STEM) faculty are professors from higher education in the content fields listed above and are required partners in each MSP project. Although a project may also include higher education faculty from education departments, the US DOE MSP project seeks to widen the conversation about these content fields to departments outside of the school of education.

Second International Mathematics Study (SIMS) is the second large scale study completed in 1981 in which the mathematics performance of thirteen year-olds and students completing their last year of secondary education were assessed and compared to students around the world.

Third International Mathematics and Science Study (TIMSS) is the third large scale study completed in 1995 in which the mathematics and science performance of U.S. students was assessed and compared with similar age students from other countries.

Wayne Regional Education Services Agency (Wayne RESA) is the intermediate school district for Wayne County in southeastern Michigan. Wayne RESA offers technical, purchasing, and instructional support aimed at improving student performance in its 34 public school districts and nearly 100 public charter schools.

Organization of the Document

Chapter 2 of this document will review the literature providing the background and context for the development of the "Project: Making Mathematics Matter" model and the importance of analyzing its data. Included in this review will be a historical perspective of the United States' concern about the science and mathematics performance of K-12 students, comparisons between US students and students in other countries including achievement and poverty status, researchers' identification of a new construct of mathematical knowledge needed by teachers of mathematics for effective instruction, research which identifies features of effective programs of professional development which result in teachers increasing their mathematical knowledge for teaching, and ultimately some new research linking mathematics professional development and student mathematics performance!

Chapter 3 will be a two-part chapter with section "Chapter 3.0" focusing on all of the information about "Project: Making Mathematics Matter" for possible replication of its intensive, sustained, mathematics professional development model. Included in this first section will be the

PM³ research design, methodology, procedures, the independent and dependent variables in its study, the instruments used to measure these variables including information about the validity and reliability of these tools, the way participants were selected, and the project's data collection timeline. The second section of this chapter, referred to as "Chapter 3.1," will concentrate on the research design of this study which used participant data collected during PM³ and performed appropriate analyses to respond to the first two research questions about differences in teachers' content knowledge for teaching mathematics and improvements in teachers' instructional practices. Additionally, in this second section, this research will gather student scale scores from the treatment districts and comparison district to respond to the third research question about the impact of increased teacher content knowledge and improved instruction on students' scores on the state of Michigan's annual assessment, the MEAP.

In Chapter 4, the fun will really begin as analyses of these data finally get started! This chapter will consider each research question, analyze data pertinent to that question, and give an interpretation of the results.

Chapter 5 will summarize the results found during this research study and reflect upon interesting questions that arose along the way, but were beyond the scope of this study.

CHAPTER 2

Review of the Literature

Introduction

The mathematics achievement of American children is a continuing national and state concern. When Russia was first to launch a rocket into outer space in the 1950s with Sputnik, there was an immediate call for more study of mathematics and science to successfully compete in the space race. This cry to improve our students' mathematical and scientific knowledge has been repeated over the last twenty-five years as international studies have reported our student mathematics performance lags that in other countries. In addition to distress about our performance relative to that in other countries, many researchers have documented persistent achievement gaps between groups of students in the United States based upon racial and economic differences (Payne & Biddle, 1999; Fryer & Levitt, 2006; Peterson, 2006).

Relative to white students, black students enter school about .6 of a standard deviation behind in achievement scores and the gap becomes even larger for middle school students (Fryer & Levitt, 2006; Haskins, 2006; Neal, 2006). Perhaps in response to such comparisons and concerns, the 2001 United States legislature enacted the No Child Left Behind Act (NCLB) which mandated all students and student sub-groups achieve proficiency on state tests of reading and mathematics by 2014. This same desire to educate students to successfully compete in the global marketplace led the 2006 Michigan legislature to pass new high school graduation requirements that include four years of mathematics— including second-year algebra. Both of these measures seem to imply the answer to our nation's mathematics woes is to simply legislate more mathematics and punish those who do not meet the increased requirements.

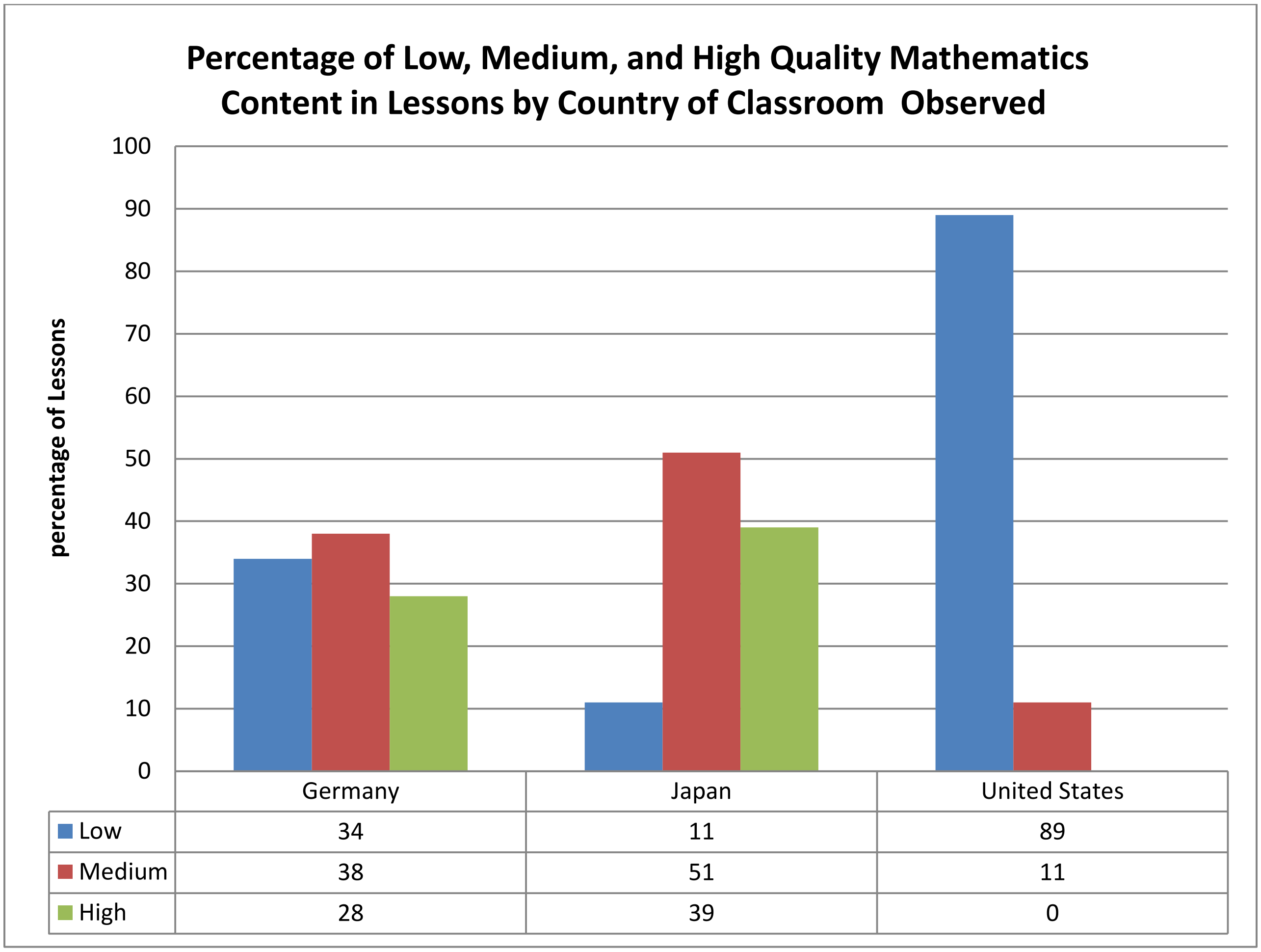
In a closer examination of the findings from international studies that preceded such federal and state legislation, questions arise about the adequacy of this response for solving our perceived mathematics achievement problems. The Second International Mathematics Study (SIMS) was conducted in the 1980s and compared the mathematics performance of eighth grade students from eighteen countries and students in their final year of secondary school from thirteen countries (NCES, 1992). When Payne and Biddle (1999) analyzed eighth grade scores from SIMS, their findings suggested that poor school funding and child poverty play important roles in the lower mathematics achievement for American students. Based upon curriculum, school funding, and demographic information for the US eighth grade SIMS participants, their research model showed students from advantaged districts with high funding and low poverty rates scored favorably with top-performing countries, while students in districts with low funding and high rates of child poverty fell into the lowest scoring tier of countries with Nigeria and Swaziland.

Although it is undoubtedly true that all participating countries have schooling differences, these researchers cited data suggesting that America was unique among the SIMS countries in its localized method of funding schools and in its large percentage of students living in poverty. These two characteristics combined to result in American students experiencing far greater differences in resources available per student based upon the state and community in which that student lived. In the SIMS comparison countries, school funding was more uniform throughout the country. Although there has been historical disagreement among school finance researchers about the role of increased funding in educational improvement, there is some recent consensus (Payne and Biddle, 1999; Addonizio, 2004; Darling-Hammond, et al. 2007) on a positive relationship between school funding resources and student achievement. If the United States' below average international mathematics performance is related to low scores in poorly funded schools, what does the new legislation do to provide adequate resources to address this problem?

An even more comprehensive, comparative research study of international mathematics and science, the Third International Mathematics and Science Study (TIMMS), was conducted in 1995. TIMSS compared the mathematics performance of forty-one countries at various grade levels—including fourth, eighth, and the final year in secondary school. While U.S. fourth grade students scored slightly above the international average, our eighth grade student performance slipped to several points under the international average, and by senior year our students were nearly the lowest performing within an elite study of sixteen countries (NCES, 1999). While the TIMMS fourth and eighth grade assessments included diverse student samplings in all countries, the final year comparison used only students in the top ten to twenty percent in each of the countries studied. America's top group of students enrolled in pre-calculus and calculus courses ranked fifteenth out of the sixteen countries in this study of elite students. What is at the core of America's comparative decline in student mathematics performance as children reach higher grades? Does requiring all students complete the mathematics study currently elected by our top students adequately address the problem?

When the National Center for Educational Statistics (NCES) studied videotaped lessons from the eighth grade TIMSS study, they discovered a striking difference between the level of study in German, Japanese, and American classrooms. A panel of respected mathematicians and mathematics educators viewed lessons from many classrooms in these three countries and gave each a rating of high, medium, or low depending upon the observed quality of its mathematics content. A compilation of the results showed great differences by country in the level of mathematics being studied as seen in the figure below. The percentages for high quality lessons in terms of mathematical content are shown in green, while low quality lessons are in blue.

Figure 1. TIMMS study in three countries comparing quality of instruction.



Note: Source of figure NCES, 1999

While Japan has the highest percentage of high quality mathematics lessons at 39 percent and Germany shows 28 percent lessons rated as high quality, but 0 percent of the American lessons received a high rating. At the other end of the rating scale, 89 percent of US lessons were rated as having a low level of mathematical content, compared to 11 percent in Japan and 34 percent of the Germany (NCES, 1999). If the quality of mathematical content in classroom lessons impacts US student learning, this nearly universal percentage of low quality lessons would seem to indicate a systemic rather than an individual problem in the content of American mathematics lessons. Perhaps the low level of mathematics lessons would help explain the gradual comparative US decline in scores as students reach higher grades. While test items in early elementary grades focus on acquisition of basic skills, questions in later grades expect students to combine operations as they apply their mathematical reasoning to solve more difficult problems. Perhaps a continued low level of mathematics lessons does not give our students the opportunity to develop a deeper conceptual understanding of mathematics that would enable them to solve more complex questions.

Analyses of these two international mathematics studies can lead to differing conclusions about America's mathematics problems and subsequent solutions. Payne and Biddle's research (1999) suggested that students in advantaged schools score quite well, but that US scores suffer because of the considerably lower scores of students living in poverty and attending poorly funded schools. Therefore, a solution would be to provide the needed resources to eliminate this achievement gap between rich and poor. The NCES videotape study (1999) proposed that the low US scores were related to the low quality of mathematics content in classroom lessons and found few instances of high quality mathematics content in any type of classroom. In this case, a solution would be to work with teachers to increase the quality of mathematical content in classroom lessons.

These solutions differ from the current national and state legislation demanding all schools and students perform better on tests of mathematics or receive punishment. There is little support or guidance about the resources to accomplish this particularly daunting task for schools in which student performance is low and poverty is high. One group of researchers reports that the pressure on low-performing schools to increase student performance is so intense that students in such schools are increasingly receiving low-level, rote instruction as teachers work to prepare students for the test that determines whether or not the school receives further state sanctions (Darling-Hammond et al. 2007). If, as the NCES found, this type of low level of instruction prevents our students from gaining higher levels of mathematics performance, such reactions to NCLB and other legislation could result in worsening rather than improving the educational opportunities for students in low-performing schools.

Although it may seem obvious that better learning is related to better teaching, there is little evidence in the literature that directly correlates student mathematics learning and classroom instruction. Many of the studies employing student achievement variables have used teacher background measures such as certification, level of the class, mathematics coursework after calculus, and years of experience—rather than measures of effective classroom instruction—as their independent variables. Such studies have come to differing conclusions with some finding links between some teacher measures and student performance (Hill et al. 2005; Darling-Hammond et al. 2007) and others finding little relationship between the two variables (Borman & Rachuba, 2000). Even though such demographic studies do not quite address classroom level interactions, some studies do suggest that classrooms in schools with high rates of poverty have higher proportions of inadequately prepared teachers resulting in fewer opportunities for these students to learn rigorous mathematics (Loucks-Horsley & Matsumoto, 1999; Loucks-Horsley et al. 2002; Hill et al. 2005; Darling-Hammond et al. 2007).

Recently, some studies have shown a relationship between improved teaching and increased mathematics learning (Mittag, Shoho, & Lien, 2004; Hill et al. 2005). Based upon this emerging evidence, the mechanisms for improving teaching become crucial for improving US mathematics performance and reducing mathematics achievement gaps between advantaged and disadvantaged students. Although this is an important issue for teacher preparation programs, it is perhaps more essential to current students that in-service teachers be given the support to improve their classroom instruction. This need is particularly great for disadvantaged students if achievement gaps are to be reduced.

While the research base relating teacher improvements to increased student mathematics achievement is weak, there is a larger body of evidence linking certain types of professional development initiatives to increased teacher knowledge (Loucks-Horsley & Matsumoto, 1999; Hill et al. 2004). Based upon the strength of these findings and the emergence of research linking increased student knowledge to improved teacher knowledge, this paper will review the literature on important components of a high-quality mathematics professional development program for teachers working in communities with high rates of poverty. The ultimate goal of such a program is to increase the mathematics performance of students in disadvantaged districts.

The balance of this chapter will examine the research on three elements of effective professional development (PD). The first section will trace an evolution of thought about the body of mathematical knowledge needed by teachers and a resulting emphasis on professional development in the form of mathematics institutes that focus on increasing the teachers' knowledge of the mathematics they teach. The second part looks at the research on professional development through the utilization of mathematics coaching. The final section considers the impact of professional development using collaborative professional learning communities and the choice of using volunteer participants or taking a whole-district involvement approach. Each of these three examinations will include an analysis of studies using low-performing, high-poverty communities.

Institutes Focused on Mathematical Knowledge for Teaching

The literature about professional development through mathematics institutes that focus on teachers gaining a deeper pedagogical content knowledge about the mathematics they teach will be the first of three important components examined. Although the idea of teachers participating in professional development is not new, there is a substantial difference between traditional professional development and the longer-term, content-focused professional development discussed in recent literature (Borman & Rachuba, 1999; Loucks-Horsley & Matsumoto, 1999; Farmer, Hauk, & Neumann, 2005).

Most current teachers have been participating in either pedagogical or content-focused professional development sessions for years through required school-wide meetings and/or by electing advanced university mathematics courses to obtain or maintain their certification. Often such professional development occurs at the beginning of the school year and consists of bringing in motivational speakers or offering multiple general sessions from which teachers can choose what to attend. Although these pedagogical in-service opportunities can help school staffs address important issues, many have instead been perceived as one-time, non-contextual lectures to attend before being able to set up the classroom for the real work of teaching. In addition, these one-time professional development sessions have little chance of significantly sustaining any improved classroom content instruction. For improving mathematical knowledge, the main emphasis has been upon teachers returning to universities to continue their study of college mathematics. For continuing professional growth, a teacher could either take higher-level mathematics content courses or elect workshops on research-based, pedagogical strategies that were not related to the particular content being taught.

As noted earlier, researchers in the 1980s such as Lee Shulman and his colleagues (Shulman, 1986, 1987; Wilson et al. 1987) began questioning this long standing dichotomy between subject matter knowledge as represented by higher level college content coursework and pedagogical knowledge as seen in content-free study of topics such as classroom management. As they studied the nature of mathematical knowledge, they proposed a new construct of mathematics content knowledge necessary for teachers that combined content and pedagogy—pedagogical content knowledge (Shulman, 1986, 1987). This type of mathematical knowledge is different than the content knowledge required for a mathematician. While a mathematician may work in isolation to find and justify a solution to a difficult and vexing mathematical problem, a mathematics teacher in a classroom may need to simultaneously anticipate student conceptual difficulties, watch for common student misconceptions, and evaluate the correctness of different student solutions. The type of deep mathematical knowledge required for each occupation using mathematics may be quite different and unique to that occupation.

These mathematics researchers (Shulman, 1986, 1987; Wilson et al. 1987) identified three competencies a teacher must possess before having the necessary knowledge for teaching—content knowledge, knowledge of pedagogical content, and knowledge of curricular content. For content knowledge, the teacher must have a deep understanding of the mathematical underpinnings of the concept being addressed including its part in the structure of mathematics and its connections to other concepts. The necessary pedagogical knowledge base for teachers includes how to convey the content to students, how to devise appropriate representations of problems, and how to anticipate and address common student misconceptions. For curricular knowledge, teachers need the ability to appropriately sequence learning within the larger structure of mathematics.

Subsequent researchers (Ball, 1990, 1991; Ball & Bass, 2002; Borko et al. 1992) built on Shulman's work about the mathematical content knowledge needed for teaching, and identified as crucial the relationships between the teachers' content knowledge, pedagogical knowledge, and curricular knowledge. In her research, Ball (1991) placed increased emphasis on the combination of content and pedagogical knowledge that allows the teacher to examine and gain insight into student thinking. Various studies suggested that student learning depended on the interplay between the teachers' content knowledge and their knowledge of students' thinking about the content (Ball 1990, 1991; Borko et al. 1992). Ma (1999) called this necessary combination of knowledge for teachers a profound understanding of fundamental mathematics or PUFM. In her comparison of U.S. teachers and Chinese teachers, she found most Chinese teachers exhibited PUFM, while most U.S. teachers did not. Ma suggests that this deep understanding of mathematics in Chinese teachers begins in their own elementary schooling, continues in university preparation, and is solidified in professional development contexts where teachers examine mathematical ideas and student thinking with colleagues. By contrast, the early schooling and university mathematics for most mathematics majors in the U.S. has not stressed understanding mathematical concepts or student thinking as much as learning the procedures involved in applying rote algorithms for solving problems (Borko et al. 1992; Eisenhart et al. 1993).

Following up on this work, Deborah Ball and her colleagues (Ball and Bass, 2000, 2002; Hill et al. 2005) delved more deeply into the relationship between knowledge about content and knowledge about student thinking that mathematics teachers need to know for teaching—especially when using a constructivist approach in the classroom. They discovered teachers need a deep understanding of the mathematics they are teaching in order to be able to evaluate unusual algorithms students might use to solve a problem. It is easy to see whether the answer is right or wrong, but far more difficult to determine if the procedure the student used would always yield the correct answer—or if it just coincidentally worked this time. This type of difficult, demanding work to evaluate student work and student thinking is often required in constructivist settings on the spur of the moment. To be successful in such settings, teachers must have the opportunity through appropriate professional development to experience rich, mathematical tasks that give rise to numerous solution strategies and must have confidence in their ability to determine the validity of various solutions. This ability to entertain and evaluate multiple solution paths to problems is a very different way of teaching mathematics than the more historical mathematics instruction that focused on demonstrating how to use particular formulas and procedures to solve certain types of problem. It also requires the teacher have sufficient content knowledge and confidence to guide student thinking in this more unpredictable setting.

Despite an increasing consensus among researchers about the increased mathematical knowledge needed by teachers to improve student performance, there was no way to test such hypotheses until 2004. In that year, Hill, Schilling & Ball (2004) published their research in which they used an instrument called the “Learning Mathematics for Teaching” (LMT) test to evaluate a large-scale professional development initiative in the state of California. The LMT was designed to measure the unique pedagogical content knowledge important in teaching mathematics by use of items that posed tasks that might arise in a classroom. Through the use of the LMT in California’s Professional Development Institutes, researchers were able to detect a relationship between certain types of professional development and an increase in teachers’ content knowledge for teaching. While there were differential results based upon the institute in which teachers were enrolled, the study showed a significant increase in teacher knowledge. In general, they found greater increases in content knowledge for teaching when teachers' professional development was of longer duration and focused on student thinking (Hill & Ball, 2004). Although this initial work with the LMT did not attempt to correlate increases in teacher knowledge for teaching with student achievement gains, a subsequent study by Hill, Rowan, and Ball (2005) reported a positive relationship between a teacher's content knowledge of mathematics for teaching (as measured on the LMT) and student achievement.

With a demonstrated relationship between this specialized knowledge required for teaching AND student performance, researchers considered how to insure the teaching force has this needed knowledge. Ma (1999) describes the elementary mathematics instruction in China that stresses concepts and meaning. Chinese teachers demonstrate a deep, conceptual understanding of mathematics and have acquired the content knowledge for teaching even though they have fewer years of formal education than U.S. teachers (11 years versus 16 years). In contrast, the U.S elementary and secondary mathematics instruction is quite procedural, stressing memorization of algorithms and most U.S. elementary teachers do not acquire the mathematics knowledge for teaching. If elementary teachers do not have a deep understanding of the mathematical concepts they are teaching, how do they help their students master them? It looks like a vicious cycle!! Researchers (Eisenhart et al. 1993; Ma, 1999; English, 2003; West & Staub, 2003) suggest the issue be addressed in the United States at a number of levels including university teacher training for pre-service teachers, mathematics professional development institutes for groups of practicing teachers, and mathematics educational research to inform the teacher training and professional development.

In summary, a number of recent and current researchers have proposed a link between increased student mathematics achievement and high-quality, professional development programs that focus on teachers gaining deeper pedagogical content knowledge about the mathematics they teach in their classrooms. This specialized body of mathematical knowledge required for successful teaching includes content knowledge, pedagogical content knowledge and curricular knowledge—a trio of cognitions Liping Ma (1999) captured with her acronym PUFM. These profound understandings are necessary if a teacher is to provide appropriate frameworks and connections for new mathematical ideas, to anticipate student difficulties and misconceptions, and to evaluate correctness of diverse student thinking.

In light of the significant black-white achievement gap and the requirements of NCLB for all student groups to become proficient on mathematics achievement tests, it becomes critically important to provide such professional development programs aimed at current teachers in low-performing schools gaining PUFM. Further, if it is true that the ability to achieve in mathematics is critical for social and educational progress of at-risk students (Ball, Goffney, & Bass, 2005; Moses & Cobb, 2001) and if other researchers replicate the findings of Hill, Rowan, and Ball (2005) showing an inverse relationship between percentage of minority students in a school and their teachers' mathematical knowledge for teaching, then it becomes exceedingly important for our democracy that we provide appropriate professional development to teachers of these at-risk students to increase teacher knowledge for teaching. Such professional development must include sustained, mathematics institutes that focus on teachers engaging in rich, mathematical tasks with other learners to provide opportunities for them to deepen the knowledge of the mathematics they teach.

Mathematics Content-Focused Coaching

Historically, teaching has been a fairly solitary profession. After a few months of student teaching, a new teacher receives certification and is on his/her own in the classroom. There were increased efforts starting in the 1990's to provide new teachers with experienced mentors to help with various teaching responsibilities, but this support is often not content-specific. Until recent accountability measures such as state standards-based tests and NCLB, teachers had a great deal of autonomy about what went on in their classrooms. This laissez-faire attitude shifted first in low-performing schools as state test results were published in local newspapers. Without much research available about how students learn mathematics, such schools and districts tried to teacher-proof the classroom by using scripted, low-level instruction that focused on rote learning (Darling-Hammond et al. 2007). The TIMSS video study results (NCES, 1999) indicated that such low-level lessons may play a role in the unfavorable international comparisons of U.S. students' mathematics achievement. These findings, combined with researchers' identification of mathematical knowledge necessary for teaching (Shulman, 1986, 1987; Wilson, Shulman, & Richert, 1987; Ball 1991; Ma, 1999; Ball & Bass, 2000, 2002; Hill, Schilling, & Ball, 2004; Hill et al. 2005), suggest a need to help teachers increase the quality of classroom instruction to raise student mathematics achievement.

A knowledgeable coach can support teachers as they gain mathematical knowledge and work to provide students with higher-level lessons. A mathematics coach plays a role similar to that of a sports coach. After teachers and coaches work together on new learning at an institute session, the coach can help teachers practice and use their new mathematical knowledge in their own classrooms. An essential role of coaching is to encourage and facilitate implementation of what the teachers are learning at institutes and grade-level meetings. The coach works individually with the teacher to assist in the task of transforming the recently acquired knowledge into action within the teacher's practice. Millard Fuller, founder of Habitat for Humanity has said, "It is generally easier to get people to act their way into a new way of thinking than it is to get them to think their way into a new way of acting" (Wiliam, 2007). This challenge of working side by side with a teacher to try a new approach and reflect on the results falls to the coach. In a study of resources spent on professional development in five urban districts, researchers suggest that resources invested in school-based instructional coaching may have greater impact than investments in other professional development strategies (Miles et al., 2005).

West and Staub (2003) refer to the idea of content-focused coaching as a professional development tool in which the teacher and coach work together to implement and analyze mathematical strategies and tasks intended to increase student learning and achievement. The teacher and coach need to establish a collaborative partnership based upon trust (Loucks-Horsley & Matsumoto, 1999). The success of this important endeavor requires the coach be an experienced teacher with a profound understanding of fundamental mathematics that is being taught in the classroom (Ma, 1999: West & Staub, 2003). This content knowledge is necessary as the coach meets with the teacher prior to the lesson to assist the teacher in identifying the important mathematical concepts in the lesson, as well as anticipating possible student responses and misconceptions. During the lesson, the coach has to have the knowledge and experience to recognize situations that require intervention and those that can wait until the post conference following the lesson. Perhaps the coach needs the most skill and knowledge at the meeting following the lesson when teacher and coach reflect on student learning and consider possible next steps. Researchers have also noted the importance of such job-embedded coaching support for teachers working to improve their practice as they attend professional development sessions that deepen their knowledge of the mathematics they teach (Stein, Smith & Silver, 1999; West & Staub, 2003).

Coaching is a second important component of a sustained, high-quality mathematics professional development program that increases teacher knowledge and student achievement. While institute sessions present a setting for teachers to collaboratively experience rich tasks designed to deepen their mathematical knowledge for teaching, until that knowledge is transformed into action in the classroom it has no ability to increase student achievement. An experienced, knowledgeable, content-focused coach can provide the support to link a teacher's increased mathematical knowledge to higher-quality lessons that result in higher student achievement.

Whole-District, Collaborative Learning Communities

A third important element to sustain a high-quality professional development initiative is the creation of collaborative professional learning communities among the participants. Many researchers have noted the importance of a collaborative culture within schools to facilitate teacher growth and improvement (DuFour & Eaker, 1998; DuFour, 2004; Hill, 2004; Loucks-Horsley et al. 2003; Loucks-Horsley & Matsumoto, 1999; Ma, 1999; Miles et al. 2005). Richard DuFour was a high school principal who developed a particular model for teachers collaborating to improve student achievement and referred to this as a professional learning community (PLC). DuFour, Eaker, and DuFour (2002) describe three characteristics that distinguish a PLC from any group of teachers meeting. A PLC must have common goals, must work interdependently to meet their shared goals, and must be focused on continuous improvement of their student results.

While some researchers suggest that any collaborative teacher PLC group has the needed knowledge to improve student achievement, Stein, Smith, and Silver (1999) propose that for mathematics groups, it may be important to have an external expert meet with the group. This outside person would be someone with knowledge and understanding about the recent research in mathematics teaching and learning. Given mathematics research (Ball & Bass, 2000, 2002; Hill, 2004; Hill & Ball, 2004; Loucks-Horsley et al. 2003; Loucks-Horsley & Matsumoto, 1999; Moses & Cobb, 2001; NCES, 1999; Stein, Smith & Silver, 1999; West & Staub, 2003) that suggests a dramatic mathematics teaching paradigm shift is required—from teachers presenting the single algorithmic solution for a problem to teachers providing rich mathematical tasks that promote students developing multiple solution paths for solving problems—the need for a skilled mathematics educator to support the group becomes obvious. Much of the research on the effectiveness of PLCs without using an external resource person has taken place in stable, higher performing schools. However, the mobility and instability of schools in high-poverty districts present another reason to provide teacher groups in such districts with a mathematics specialist who meets with them on a long-term, sustained basis.

When researchers discuss ways for professional developers to help create a collaborative culture among participants, they reiterate the importance of groups working interdependently to achieve a shared vision of improved student achievement. They stress the importance of student thinking and learning being the focal point of the collaboration (Ball & Bass, 2000, 2002; DuFour & Eaker, 1998; Eaker, DuFour & DuFour, 2002; Loucks-Horsley et al. 2003; West & Staub, 2003). Although the group may choose to consider what actions, or teacher moves, a particular teacher made to achieve better student learning, the emphasis on the diversity of student thinking and learning is intended to create an environment in which teachers can openly consider changes in practice with the support of others and without disciplinary repercussions.

DuFour and Eaker (1998) and DuFour (2004) insisted that involvement in such professional learning communities should not be voluntary—everyone is required to attend and participate with their group. This idea of whole-school participation is quite different than the traditional professional development where a teacher chose to attend a workshop of interest or take university classes (Loucks-Horsley & Matsumoto, 1999). Staff developers using the volunteer model sought out willing participants to learn together and hoped these early adopters would influence others. While much of the professional development research refers to voluntary participation of teachers (Hill, 2004; Hill, Rowan, & Ball, 2005; Loucks-Horsley et al. 2003), recent researchers evaluating district resource investments in professional development in urban districts for The Finance Project reported a growing consensus that programs that included all teachers—not just volunteers—have a greater potential for significantly impacting achievement in a low-performing district (Miles et al. 2005). Researchers Loucks-Horsley and Matsumoto (1999) in their meta-analysis of professional development, point out the critical role systemic professional development played in the improvement of District #2 in New York City—another urban district. District #2 is the district in which Lucy West developed and used her model of content-focused coaching. They concluded that such local, systemic initiatives that work with all of the teachers within some unit have shown promising improvements in mathematics teaching. Unfortunately, there is not as much careful research examining the impact of these whole system programs on student achievement—especially in high-poverty districts.

Support for organizing professional development around groups of teachers and whole faculties is also found in the analysis of effective PD features by Odden, Archibald, Fermanich &

Gallagher (2002). These researchers worked on developing a method for analyzing the cost effectiveness of various components of professional development and whole school participation was one of the six elements of effective PD they chose for inclusion in their framework. In a later study, a research group applied this method in their examination of professional development dollars expended in five, large, high-poverty, urban districts from this finance perspective. What features of professional development give the best return within these urban, high-poverty settings? These researchers noted an important policy distinction within these districts as to whether the major focus of the PD is on building individual or group capacity. With higher rates of student and teacher mobility, working with all within the system helped maintain the coherence of instruction for students. Their meta-analysis of effective professional development in these urban settings suggested that an emphasis on increasing group capacity with whole school, whole district participation might be a critical feature in such highly mobile settings (Miles et al. 2004).

Conclusion

Concerns about student mathematics achievement—particularly in high-poverty communities—have prompted research into areas that are associated with professional development that enhances the teaching and learning of mathematics. This review focused on examining the research about three components of effective professional development initiatives: teachers acquiring the necessary mathematical knowledge for teaching by participating in mathematics institutes; coaches supporting teachers in their classroom implementation of new learning; and groups of teachers coming together to analyze student work and plan instruction in systemic, whole-school professional learning communities.

There is a fairly substantial body of research about the specific substance of the deep content knowledge teachers need to have to successfully prepare high-quality lessons, evaluate the correctness of multiple solutions to rich tasks, and diagnose student difficulties or misconceptions (Shulman, 1986, 1987; Wilson et al. 1987; Ball, 1991; Eisenhart et al. 1993; Ma, 1999; Ball & Bass 2000, 2002). In addition, Hill, Schilling, and Ball (2004) have recently published a reliable instrument for measuring this specific type of mathematics pedagogical content knowledge for teaching, the LMT. Many researchers have supported mathematics professional development institutes as the setting for teachers to collaboratively participate in rich tasks designed to increase their mathematics content knowledge, pedagogical knowledge, and curricular knowledge (Loucks-Horsley & Matsumoto, 1999; Ma, 1999; Ball & Bass, 2000, 2002; Loucks-Horsley et al. 2003; West & Staub, 2003). Based upon research suggesting students in high-poverty areas tend to have teachers who are less qualified and have fewer opportunities for professional growth (Borman & Rachuba, 1999; Darling-Hammond, 1996; Darling-Hammond et al. 2007), such mathematics professional development institutes are critical for providing these students with an opportunity to learn and achieve.

While such institutes may be a necessary ingredient for increasing mathematical pedagogical content knowledge and changing teacher thinking, they are probably not sufficient to transform the new learning into improved classroom instruction. Although a teacher has gained knowledge about a particular task or strategy to increase student learning, s/he may still be hesitant to act on it in the classroom. Helping bridge this gap between knowing how to teach a concept and actually teaching it in this new manner is the mathematics coach. The mathematics coach must be a skilled mathematics educator who can accompany the teacher back to the classroom and collaborate with her to apply the newly acquired knowledge to improve the quality of lessons (Loucks-Horsley & Matsumoto, 1999: West & Staub, 2003; Miles et al. 2005). Only through improved instruction can students in low-performing, high-poverty schools be provided with the opportunity to gain a deeper understanding of mathematics that will result in increased student achievement.

A new consideration of teachers functioning interdependently in professional learning communities is a promising change from the traditional model of teachers working in isolation (DuFour, 2004). A PLC reinforces the value of collaboration and the strength of the professional knowledge shared among its members. They empower teachers to look within for the answers to the challenges they face in improving student mathematics achievement. PLCs are one model of whole-district systemic change initiatives. Recent studies of effective professional development models for urban areas suggest that there is greater potential for student impact when all of the teachers within a unit participate in a whole-district, systemic program (Loucks-Horsley & Matsumoto, 1999; Miles et al. 2005).

Given the punishments applied to schools and districts that fail to meet state and federal requirements such as NCLB, it is unconscionable to implement such consequences upon the lowest performing schools without providing the needed resources to succeed. One resource that is sorely needed is a robust body of research performed in similar at-risk districts that provides further information about initiatives that are linked to improved student achievement in mathematics. Students in such schools will benefit from resources that promote stable, thoughtful, sustained, research-based approaches to improved classroom instruction. This approach is contrasted with the current programs of state and federal consequences for low-performance of schools which include changing administrators, changing teachers, reconstituting schools and closing schools—strategies that result in students—especially those in high-poverty areas—having little stability of teachers or schools during a year or from year to year. Many studies have noted that contrary to the idea of teacher-proofing instruction, the teacher is the key ingredient for student learning (Borman & Rachuba, 1999; Darling-Hammond, 1996; Darling-Hammond et al. 2007; Shulman, 1996).

With current research suggesting that the knowledge for teaching mathematics is different than the mathematical knowledge emphasized in university teaching preparation (Ball & Bass, 2000, 2002; Loucks-Horsley et al. 2003; Shulman, 1996, 1997; Stein et al. 1999), a research-based model for effective professional development for practicing teachers is critically needed. Much of the currently available research about mathematics professional development has dealt with volunteers from various districts—possibly including participants from high poverty areas. Although it is interesting to learn that teachers participating in high-quality professional development programs have generally shown increases in their mathematical content knowledge for teaching, the more important dimension of knowledge about these programs must be their impact upon classroom instruction and student mathematics achievement. More research is needed in high-poverty communities to examine the important relationships between increased teacher knowledge gained in sustained, high-quality professional development programs, enhanced classroom instruction as a result of interactions with a mathematics coach and colleagues in a whole-district professional learning community, and improved student mathematics learning and achievement.

If our nation is serious about providing equal mathematics educational opportunities for all and comparing more favorably in international mathematics assessments, it is essential to help low performing, high-poverty districts produce effective, high-quality, sustained professional development programs. These programs need to focus on helping these teachers gain a deep, pedagogical content knowledge of the mathematics they teach in order to improve the quality of classroom instruction that results in the increased mathematics achievement of their diverse students.

CHAPTER 3

Research Design and Methodology

This research study has a two-part research design and methodology chapter. The first part of this chapter (3.0) examines the research methodology and design for the PM³ mathematics professional development program. This section will include the PM³ design, hypotheses, sample selection, instrument selection, descriptions of the PM³ treatment including its three crucial components of mathematics institutes, content-focused coaching, and collegial, after-school, grade-level meetings, and the procedures during the implementation of PM³. The second part of this research and methodology chapter (3.1) focuses on this research study's ex-post-facto evaluation of the data collected during the four-year course of PM³ and will include research design and methodology, hypotheses, and analyses of the data. In addition, the information about collection and discussion of student data will be part of section 3.1.

CHAPTER 3.0

Research Design and Methodology for PM³

Research Design

Project: Making Mathematics Matter used a quasi-experimental, matched comparison group design to compare measures of mathematics teachers' pedagogical content knowledge, teachers' classroom instructional practices, and students' mathematics achievement in high-poverty districts. PM³ was an intensive, professional development program funded through the US DOE and the State of Michigan with a Mathematics and Science Partnerships (MSP) grant award to Wayne RESA in partnership with STEM faculty from the University of Michigan-Dearborn. In its request for proposals, the State of Michigan encouraged projects with a rigorous research design using control or comparison groups. Based upon concern about two low-performing, high-poverty districts in Wayne County and on emerging research evidence that a whole district approach in such challenged districts had a higher probability of success, use of random selection was not an option for PM³. However, the grant funding did allow PM³ to include a rigorous evaluation component for the project including money to attract teachers from a similar district to serve as comparison group teachers for testing, money to pay for experienced educators to become certified on the use of the SAMPI classroom observation protocol, and to also pay them to observe all treatment group teachers at the beginning, middle, and end of the project. The project also had funding for an evaluator to plan and coordinate all of the above activities, as well as to conduct focus groups to formatively guide the project leadership planning team.

Research Hypotheses for Project: Making Mathematics Matter

The null hypotheses for PM^3 follow:

1. There will be no significant differences in mathematics content knowledge for teaching between PM^3 treatment group teachers and comparison group teachers as measured on pre and post scores on the Learning Mathematics for Teaching (LMT) assessment tool. The independent variable will be the participant's membership in the treatment or comparison group and the dependent variable will be the participant's score on the LMT.
2. There will be no significant changes in classroom instruction and practice of treatment teachers during and following their participation in the PM^3 professional development project. The independent variable is the length of time the teacher has participated in the PM^3 professional development program and the dependent variable is score on the Science And Mathematics Program Improvement (SAMPI) classroom observation rubric.
3. There will be no significant differences in student mathematics achievement in grades 4 through 8 as measured by state of Michigan MEAP mathematics scores between students in treatment districts and students in the comparison district prior to, during, and following treatment teacher participation in the PM^3 project. The independent variable is whether students come from the treatment districts or the comparison district and the dependent variable is the student mathematics scores on the state of Michigan standardized annual test, the Michigan Educational Assessment Program (MEAP).

Sample Selection

In 2004, the State of Michigan authored a request for proposals (RFP) for Michigan's allocation of federal funds through the U.S. DOE Mathematics and Science Partnerships (MSP) program. Michigan's funded projects would become part of a nation-wide network of MSP grants designed to add to the literature about the impact of professional development programs on teacher knowledge and student performance and were to be directed at improving the mathematics and/or science performance of students in grades 3-8 in "high needs" districts. "High needs" was defined as having over 35 percent of school students identified as economically disadvantaged, scoring under 50 percent proficiency on the mathematics portion of the Michigan Educational Assessment program (MEAP), failing to make Adequate Yearly Progress (AYP), and having a large number of teachers not meeting the "highly qualified" status under "No Child Left Behind" (NCLB).

After examining the "high needs" data for the thirty-three districts serviced by the Wayne County Math and Science Center, the mathematics consultants at Wayne Regional Educational Services Agency (RESA) and their STEM partner faculty from the University of Michigan-Dearborn, approached leaders in District T1 and District T2 about a partnership in which all of their teachers of mathematics in grades 4-8 would participate in PM³, an intensive, sustained, quasi-experimental, mathematics, professional development study designed to enhance the mathematics performance of their teachers and students.

The leadership team considered the data in Table 1 from schools in these two districts during their needs assessment process. These data represent the overall school summary weighted average of the percentage of students proficient in mathematics that year. This is the percentage used by the state to determine whether the school has met "adequate yearly progress"

(AYP). Each data point includes the scores for all grades of students in that building that took the MEAP mathematics assessment. During the three years of data used for the selection process, mathematics MEAP testing for students only occurred in grade 4, grade 8, and high school. Due to various grade configurations in these districts during these years (K-5, 6-8, K-8), some data points are for a single grade while others include scores for students in grades 4 and 8.

Table 3. T1 and T2 schools' (A, B, C, D) percentages of meeting state expectations 2002-2004

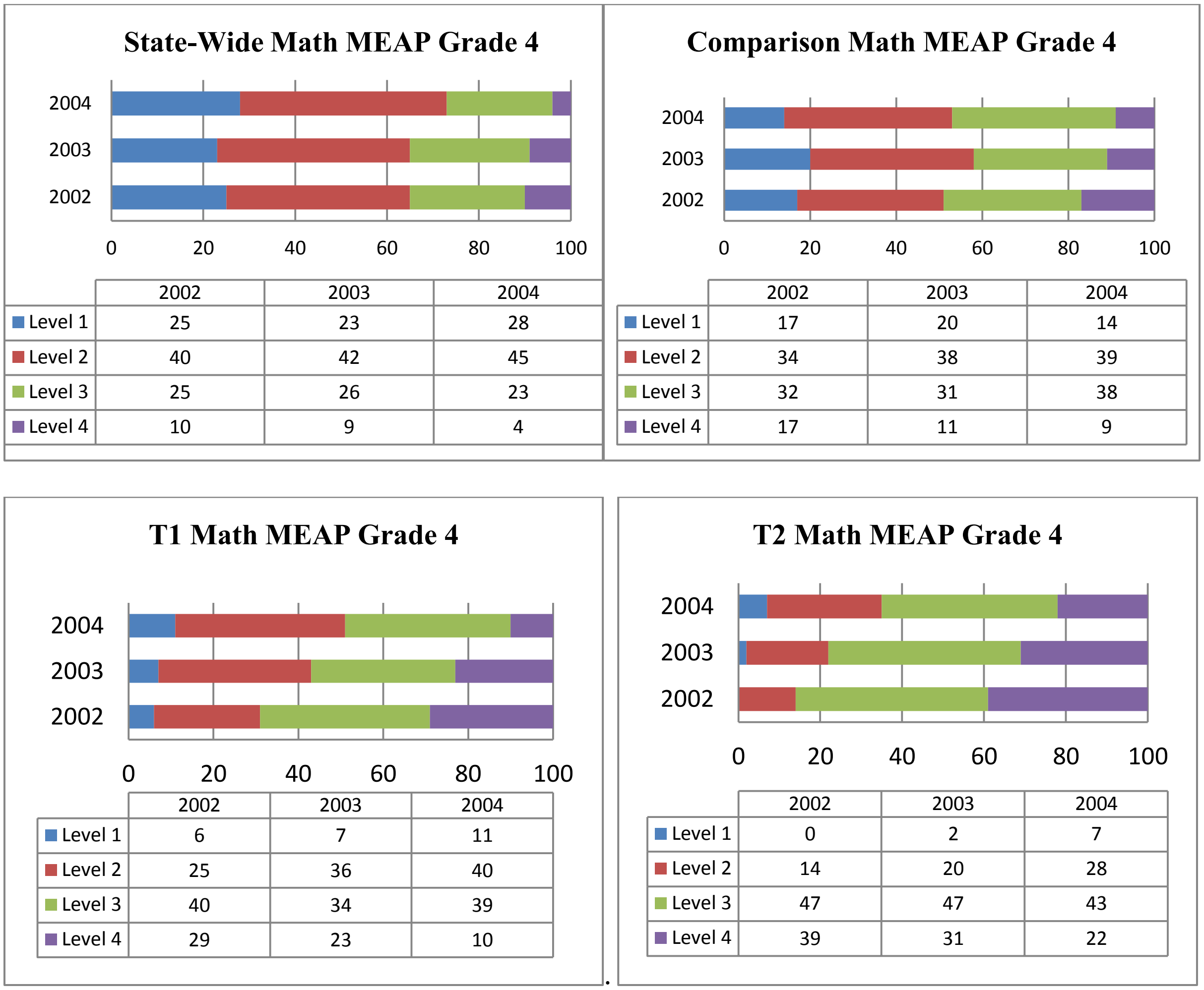
Elementary Schools in Treatment District 1 (T1) and Treatment District 2 (T2)			
District AYP Education YES!	MEAP 2002	MEAP 2003	MEAP 2004
T2 School T2A – Phase 4 (unaccredited)	13.2	11.5	28.0
T2 School T2B – Phase 4 (unaccredited)	10.0	16.3	45.0
T1 School T1A – Phase 4 (Grade D-Alert)	14.3	39.1	43
T2 School T2C – Phase 2 (Grade C)	24.3	31.4	51.0
T1 School T1B – Phase 4 (Grade D- Alert)	51.7	40.8	51.0
Middle Schools in Treatment District 1 (T1) and Treatment District 2 (T2)			
T2 – T2A - Phase 4 (Unaccredited)	n/a	n/a	27.3
T2 – T2C - Phase 2 (Grade C)	n/a	n/a	32.9
T2 – T2D (No Status- Restructured)	11.1	5.2	8.2
T1 – T1D - Phase 4 (D-Alert)	23.4	17.7	39.5

District T1 and District T2 both had elementary schools and middle schools that had failed to make AYP for a number of years and their MEAP scores were among the lowest in the state according to the 2002 MEAP data report. For these two districts in 2002, the grade 4 MEAP proficiency rates were 14 percent and 31 percent, while the grade 8 proficiency rates were 11 percent and 23 percent. Seven of the eight elementary or middle school buildings in the two districts were in Phase 2, 3 or 4 on Ed YES! Five of the schools were in Phase 4 (Restructuring), three had a Grade D "alert" status, two were unaccredited, and one had recently been restructured. Both districts identify approximately 90 percent of their students as economically disadvantaged. It was also discovered that nearly half of the teachers working with students in mathematics in grades 4 through 8 in these districts did not meet "highly qualified" status for NCLB.

In 2002, 2003, and 2004, Michigan's MEAP tests were administered in February and in mathematics only three grades of students were tested— grades 4, 8, and 11. The PM³ project included teachers of mathematics in grades 4 through 8, so the grades 4 and 8 data are shown in figures 2 and 3. These years were selected as they are the three years of MEAP data prior to the fall 2004 start of PM³ and were the baseline years of data considered for "high needs districts" for the MSP grant application. These scores were the years of data requested by the RFP to show districts qualify for funding and are "high-needs" districts due to low performance in mathematics and having large percentages of students identified as socio-economically disadvantaged.

In 2002 through 2004, the MEAP mathematics student scores were identified as Level 1: "Advanced"; Level 2: "Proficient"; Level 3 "Partially Proficient"; and Level 4: "Not Proficient" depending upon which of four ranges the scores were within. When the percent of proficiency of a school or district is reported, this number includes the combined percentages of students scoring in levels 1 and 2. Figure 2 will demonstrate the percentages of students in each of these four categories for the state, the comparison district, and the two treatment districts.

Figure 2. Grade 4 MEAP proficiency comparisons 2002-2004.



State-Wide Math MEAP Grade 4

	2002	2003	2004
Level 1	25	23	28
Level 2	40	42	45
Level 3	25	26	23
Level 4	10	9	4

Comparison Math MEAP Grade 4

	2002	2003	2004
Level 1	17	20	14
Level 2	34	38	39
Level 3	32	31	38
Level 4	17	11	9

T1 Math MEAP Grade 4

	2002	2003	2004
Level 1	6	7	11
Level 2	25	36	40
Level 3	40	34	39
Level 4	29	23	10

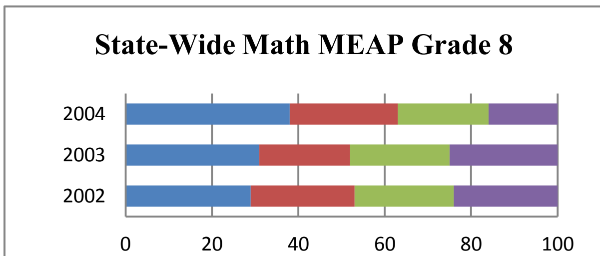
T2 Math MEAP Grade 4

	2002	2003	2004
Level 1	0	2	7
Level 2	14	20	28
Level 3	47	47	43
Level 4	39	31	22

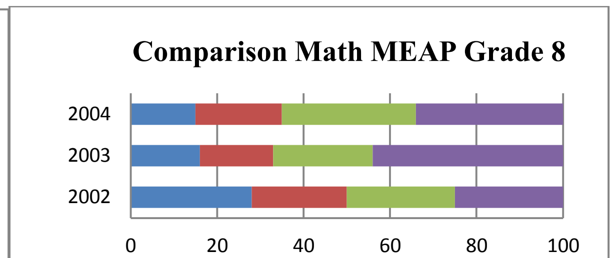
Figure 2 uses bar graphs to provide a picture of student scores in treatment and comparison districts relative to state-wide averages. On the bar graphs in Figure 2, the parts of the bar representing proficiency are the two leftmost sections of the bar, the blue (advanced) and red (proficient) portions. The further the blue and red portions extend to the right of the graph, the greater the percentage of students were rated proficient for that year in that grade. The green part of the graph signifies the partially proficient (level 3) students and the purple part of the bar shows the percentage of students found not proficient (level4). In addition to wanting to increase their percentage of students in the blue and red categories, any school really wants to decrease their students in the purple (level 4) category. Students scoring that far below proficiency will truly be challenged to have full career options.

The state-wide average rate of student proficiency on the grade 4 MEAP tests for the February 2002, 2003, and 2004 administrations was 68 percent and stayed relatively stable for these years. The comparison district average for these same three years was about 14 points lower at 54 percent, T1 was 26 points lower at 42 percent, and T2 was 44 points lower at 24 percent. Furthermore, the purple "not proficient" sections of the bars are large.

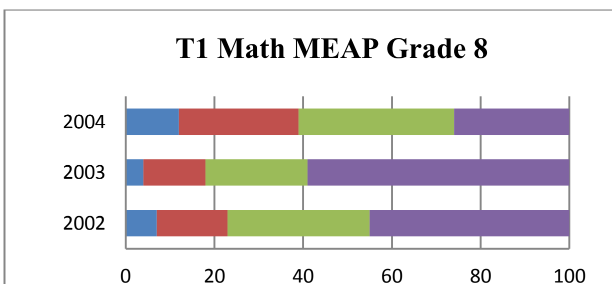
Figure 3. Grade 8 MEAP proficiency comparisons 2002-2004.



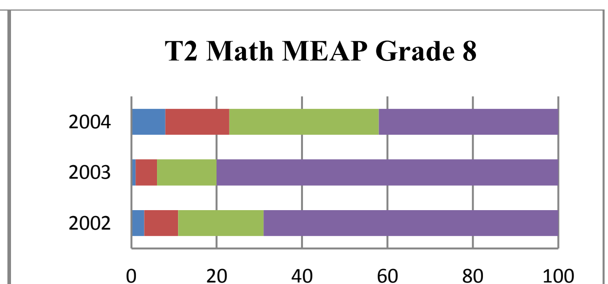
	2002	2003	2004
Level 1	29	31	38
Level 2	24	21	25
Level 3	23	23	21
Level 4	24	25	16



	2002	2003	2004
Level 1	28	16	15
Level 2	22	17	20
Level 3	25	23	31
Level 4	25	44	34



	2002	2003	2004
Level 1	7	4	12
Level 2	16	14	27
Level 3	32	23	35
Level 4	45	59	26



	2002	2003	2004
Level 1	3	1	8
Level 2	8	5	15
Level 3	20	14	35
Level 4	69	80	42

Figure 3 gives a visual representation for comparison of the grade 8 MEAP mathematics scores for the state, the comparison district and the two treatment districts during the same testing administrations and prior to the start of PM³. The 8th grade tests in general have larger purple "not proficient" sections than were seen in grade 4. Generally speaking, the proficiency rates for Michigan students on the MEAP decrease as the grade level increases. Certainly, the test items have more complexity and a greater range of content for higher grades. Another factor that might play a role—especially in challenged environments—could be a compounding effect of years with substitute teachers, years with non-certified teachers, and years of less effective instruction. However, this general decline poses a more universal question about mathematics instruction in our state and in our country as seen in the TIMMS study which found mostly low level mathematics lessons in U.S. eighth grade classrooms. Is a piece of the lower scores shown in Figure 3 a result of our cultural tradition of mathematics instruction? The average rates of proficiency for 2002 through 2004 are a state-wide average of 56 percent; comparison district average of 39 percent, treatment district 1 average of 27 percent, and treatment district 2 average of 13 percent. Even more telling are the average percentages of students represented in the purple portion of the bars rated "not proficient" at 22 percent, 34 percent, 43 percent, and 64 percent respectively. It is difficult to foresee future educational success for students in schools in which such large percentages leave eighth grade with substantial mathematical deficits.

The low student MEAP scores and high rates of socio-economically disadvantaged students in districts T1 and T2 led to discussions with district leaders who indicated their number of teachers involved in mathematics instruction for students in grades 4 through 8 as approximately 55. A PM³ sample size of 50 to 60 treatment participants with an approximately equal number of matched comparison group teachers would result in a larger than needed sample size for robust findings in this type of study. However, the sustained nature of the project along with a concern about elevated rates of change and staff turnover in such high-mobility, high-needs districts prompted the decision to retain this high initial sample size.

The sample size goal of the project was to complete the study with 40 to 50 participants—20 to 25 treatment group teachers and 20 to 25 comparison group teachers. The calculations and computations for arriving at this goal for sample size came from a number of sources. One input about sample size comes from the U.S. Department of Education (US DOE) and the Michigan Department of Education (MDE) in a published request for mathematics professional development proposals to achieve a final treatment group size of at least 25. This US DOE sample size suggestion is in line with a number of other sources for project sample sizes. Another source of information for sample size estimations is the GPOWER computer program published by Erdfelder, Faul, & Buchner in 1996 (as reported by Keppel & Wickens, 2004). Their computations suggest a project have approximately 15 to 40 cases in each group or cell in a study with two groups and a large expected effect size. Another source for sample size estimates is *Multivariate Data Analysis* by Hair, Black, Babin, Anderson, and Tatham (2006). Hair and colleagues suggest at least 20 sets of scores for the teachers who are in the group receiving the treatment and an approximately equal number in the group not receiving the treatment. Since one of the important features of this project concerns the importance of whole district participation for achieving significant improvements for students, random sampling was not an appropriate method for selecting participants.

Once District T1 and District T2 were selected to receive the mathematics professional development intervention, the difficult task of securing an approximately demographically similar district that would agree to serve as a comparison group began. Since the two districts selected for the PM³ project are among the state's lowest performers in mathematics achievement and highest in poverty rates, finding truly comparable districts to provide comparison group teachers is difficult. Since districts T1 and T2 had the two highest rates of disadvantaged students in Wayne County, other districts in the county could not be equally disadvantaged. One district chosen for the treatment consistently rated 34th out of 34 Wayne County districts in the percentage of proficient students on the mathematics portion of the state MEAP test at each tested grade level. In conversations within the county about serving as a comparison group for a research study, districts were hesitant to be compared to districts receiving the treatment, lest their identity be discovered at some point and the results reflect poorly on them. Another issue to be considered for a comparison group is what interventions it will have because of its low mathematics performance. It is virtually impossible to find any low-performing school that is not working on improving its instruction and performance.

Project leaders did secure an approximately demographically similar district in to serve as the comparison group for PM³. The comparison group teachers were from a district in another Michigan county. The comparison teachers volunteered to participate in the mathematics testing for a fee. As it turned out, the emphasis on approximately demographically similar comparison group must be on the word "approximately."

While both treatment districts had approximately 90 percent of their students on free or reduced price lunch, the comparison district had a smaller percentage of their students enrolled in this program, about 60 percent. While both are high percentages, the treatment districts' percentages are higher. In figures 2 and 3, the rates of student proficiency on the MEAP tests for the comparison district were shown to be 14 points and 17 points lower than the state average for grades four and eight respectively. The comparison district met the definition of a "high needs" district by having an over 50 percent free and reduced lunch rate and MEAP scores in some grades under 50 percent. After the first round of testing was completed and demographic data were collected, project researchers found a great difference in teachers meeting the "highly qualified" definition in NCLB. While nearly 50 percent of the teachers in the treatment group did not meet this designation—including teachers without college degrees, without teaching certificates, with no mathematics certification to teach middle school mathematics—inspection of the preparation of teachers in the comparison group indicated a single teacher, a middle school, special education teacher who might not have met "highly qualified" status. Most had mathematics majors or mathematics endorsements on their teaching certificates.

Instrument Selection

Many of the instruments selected to measure the dependent variables in this study were shared by a number of similar Mathematics and Science Partnership projects across the state of Michigan. Following are descriptions of evaluation measurement tools.

Measure of Pedagogical Content Knowledge for Teaching Mathematics

The Learning for Mathematics Teaching Scale (LMT) is an instrument designed by Heather Hill and Deborah Ball at the University of Michigan to measure mathematics content knowledge for teaching as identified by Lee Shulman and his colleagues in the late 1980's. This construct of mathematics content knowledge for teaching was further refined by Deborah Ball, Hyman Bass, Liping Ma and many other mathematics researchers from the late 1990s to the present. LMT items ask teachers to demonstrate their ability to perform many mathematically related decisions they must make during mathematics instruction including judging the correctness of student solutions to problems, diagnosing student misconceptions during problem solving, and choosing a next step to help a student solve a problem situation.

The LMT was developed in the early 1990s and piloted by its authors in a large-scale, state-wide, mathematics professional development program in California in 2001 (Hill & Ball, 2004). Original measures of reliability for the LMT pilot tests include Cronbach's alphas in the 70s and 80s measuring internal consistency and solid Item Response Theory (IRT) reliabilities. The large pool of items piloted in the California project demonstrated average reliabilities in the .80's. Analyses of these items were performed, the test was revised, and the resulting LMT Scale was found to have a reliability of .88 on a successive study (Hill, Schilling & Ball, 2004; Hill et al. 2005). The LMT was not designed to match the curriculum of any particular mathematics professional development program, but validity measures consisted of comparing the scope of the items to the content domains for grade spans as defined by the National Council of Teachers of Mathematics (NCTM). Additional work establishing the validity of the LMT is on-going and researchers are approaching establishing the validity by noting the assertion in Kane's 2004 work that establishing validity is an iterative process. Researchers also note that the strongest case for validity will be made when research demonstrates increases on the LMT are associated with better classroom instruction and student achievement (Schilling & Hill, 2007).

Due to its coherence to recent scholarly research findings about how students learn mathematics combined with resulting implications on mathematics instruction, many of the MSP grants throughout the country are using the LMT as the measure of teacher mathematics content knowledge. The remarkable difference in this assessment of mathematics content knowledge over previous tests was that the LMT items actually were modeled on the types of mathematical knowledge a teacher in the grade span would need when working with students. The item might show a student's solution to a task using base ten blocks and ask what kind of misconception this student was experiencing or which of several different student solutions were using a solution path or algorithm that would always work. To view Learning Mathematics for Teaching assessment item samples see Appendix A.

Measure of Quality of Classroom Instructional Practices

The initial development in 1998 of the Science and Mathematics Program Improvement (SAMPI) classroom observation protocol was funded through the Michigan Goals 2000 grant by Michigan State Board of Education. Revisions in 1999, 2001, and 2003 were funded through competitive grants through the Eisenhower Higher Education Program. The SAMPI classroom observation protocol is a two-step process in which a trained outside evaluator observes and takes notes for an entire mathematics lesson and then leaves the classroom to record the actions witnessed in the classroom onto a rubric.

The SAMPI protocol rubric asks the trained observer to record classroom events on multiple items in three areas: content of the lesson; implementation of the lesson; and classroom culture. Each indicator is rated on a scale of 1 to 7 with a rubric describing the rationale for the measures. In addition, the trained observer establishes a summary or total rating for the lesson quality. The SAMPI observation system was originally developed for the evaluation of Michigan's 33 regional Mathematics and Science Centers and has been used extensively throughout the state for a number of years. Prior to conducting field observations, an external observer is required to attend comprehensive training which includes studying the classroom observation instrument, viewing videotaped lessons, and recording and ranking the lesson on the SAMPI protocol rubric. This ranking calibration process continues until trainees exceed an 80 percent agreement with the rankings from the protocol. Researchers have gathered data which shows Cronbach's alpha reliability scores all greater than .74 on the internal consistency and observer agreement using this SAMPI classroom observation debriefing protocol (Kilday &

Kinzie, 2009). The validity of the SAMPI classroom observation protocol is established through comparisons of the descriptors in the instrument with the national mathematics standards as published by the professional organization for mathematics teaching, the National Council for Teaching Mathematics (NCTM).

Measure of Student Mathematics Achievement

Michigan has had many different state tests of mathematics achievement throughout the history of its Michigan Educational Assessment Program (MEAP). The current test was designed in response to NCLB and is administered in the early fall to all students in every grade 3 through 8. This Michigan MEAP version is based on revised grade-level content expectations written following the 2001 adoption of NCLB and had its first administration in the fall of 2005. The previous state mathematics assessment was administered only to 4th and 8th grades students in February at the beginning of the second semester of the year.

For MEAP measures of reliability, the Cronbach's alphas measuring the internal consistency of the mathematics portion of grades 3 through 8 MEAP are in the lower 90s and the Item Response Theory (IRT) for mathematics in grades 4-8 is also in the 90s. The content validity of Michigan's MEAP is verified through a number of considerations. The first is the fact that the MEAP test itself is based upon the state's content standards by grade and questions are aligned to that document. Face validity is assured through the annual gathering of educators, item writers, and state testing staff to review and field test items.

PM³ Professional Development Treatment

Treatment group teachers participate in all three professional development features selected for inclusion in the model for "Project: Making Mathematics Matter": 1) the mathematics institutes; 2) the content-focused coaching; 3) the collaborative, collegial learning communities. Mathematics institutes focus on deepening participants' mathematics content knowledge for teaching and meet for ten full days throughout the school year with approximately one meeting per month. Intensive institute instruction also takes place during a week in the summer. In addition, participants are assigned a coach who accompanies them back to their classrooms for support in implementing new strategies learned during the institutes. The 25 to 30 participants from each district share a coach and that coach also facilitates the monthly, after school, collaborative meetings for various grade level groups in the district. Each participating teacher received over 100 hours of professional development each year.

Table 4. PD hours per year for PM³ participants

	Fall Institute	Winter Institute	Summer Institute	After-School Meetings	Individual Coaching	Total
PD Hours	30	30	30	15	20	125

Institutes

Each mathematics institute consists of five, full-day sessions that focus on teachers gaining a deeper understanding of an important topic within the mathematics they teach. Since the PM³ model elected a whole-school team approach for two disadvantaged districts as opposed to inviting individual first responders from many districts, the institutes met on school days to make the treatment available to all teachers. Each MSP grantee was required to partner with faculty from the departments of Science, Technology, Engineering, or Mathematics (STEM). For PM³ the STEM partners were from the Center for Mathematics Education at the University of Michigan-Dearborn and the Center provided the coursework and instruction for the five-day, 30-hour institutes. Another advantage for the PM³ treatment teachers was the opportunity to elect earning graduate credits through these courses to work toward becoming “highly qualified” or to add a mathematics endorsement to their certificate. Other participants chose to apply the professional development hours toward becoming “highly qualified” as defined by NCLB. As mentioned previously, nearly half of the original fifty-five PM³ participants from districts T1 and T2 did not meet this standard. The institute courses could be elected for two graduate credits at an additional individual expense. Over the four-year course of the project, participants took up to 12 mathematics content classes. Following is the schedule of institute course-work for the PM³ teacher participants:

Table 5. Mathematics institute course work

Term	2004-2005	2005-2006	2006-2007	2007-2008
Fall	Number and Operations	Geometry and Measurement 1	Proportional Reasoning Cases	Geometric Reasoning Cases
Winter	Number and Proportional Reasoning	Geometry and Measurement 2	Algebraic Reasoning Cases	Action Research Insight into Student Thinking
Summer	Algebraic Reasoning	Using Technology to Study Statistics and Probability	Graphing Technology as a Learning Tool	Implementation of Standards-Based Curricula

In the first year of the project, the focus of study was number, operations, and proportional reasoning, with a two-course sequence, plus a summer course on algebraic reasoning. In the second year of the project, participants examined geometric reasoning, creating intellectual communities in the classroom, and using technology in the study of statistics and probability. Year 3 used case studies and technology to delve more deeply into number and proportional reasoning and algebraic reasoning. The year 3 summer course had participants again working with technology in further study of data, statistics, and probability. In the final year of the project, the fall course used case studies to examine student geometric reasoning, the winter course involves grade level groups performing action research for a problematic concept, and the final course helped district teams create plans for continually addressing issues surrounding implementation of standards-based mathematics curricula after the PM³ grant ended. Institute instructors modeled best practices and facilitated collaborative learning in order to help teachers increase their mathematical pedagogical content knowledge for the mathematics they teach and to improve their classroom instruction.

Coaching

The PM³ coaching model is based on *Content-Focused Coaching: Transforming Mathematics Lessons* by West and Staub (2003). This book documents eight years of research examining the components that result in successful coaching of mathematics teachers. The importance of a coaching component was discovered during offerings of similar mathematics institute courses prior to this project when teachers were hesitant to implement new strategies in their classrooms. These new strategies stressed teaching mathematics for understanding as opposed to teaching mathematics as steps or procedures to follow. Most of the teacher participants had learned mathematics in a procedural way during their own K-12 and college education; they lacked the confidence and/or deep content knowledge to be comfortable allowing students to search out various solution paths to a problem.

For their students to gain a more conceptual understanding of mathematical reasoning, PM³ teachers had a coach to help them successfully facilitate small groups of children working collaboratively on worthwhile mathematical tasks. This classroom learning structure assumed students found different solution paths AND teachers encouraged students to find solutions that make sense to them—rather than use of a teaching structure in which teachers show students the preferred steps in solving that type of problem. Being able to follow the reasoning of others, assess its validity and/or identify students' misconceptions requires far more mathematical content knowledge than teaching the steps in solving particular types of mathematical problems. The PM³ coach supported the teacher as s/he works to create this rich, constructivist, classroom learning structure where students are engaged in doing mathematics—not simply memorizing steps or procedures. Coaching roles may have included helping teachers plan for instruction, debriefing after instruction, or modeling instruction. While early in the project, coaches may tend to do more modeling and co-teaching, later as PM³ teachers gain content knowledge and instructional successes, the coach becomes more like a guide for participants. The classroom, on-site coaching feature of PM³ helped teachers learn how to directly embed the institute learning in their daily teaching and gain confidence in their ability to facilitate divergent thinking with students.

PM³ hired two retired former mathematics teachers with a great deal of experience supporting classroom teachers to serve as project coaches—one coach for each district. Since each coach worked with all project participants from an intervention district, the coaching load was 25 to 30 teachers per coach. In addition to spending time on-site in district schools supporting individual teachers, facilitating after school team meetings, and coordinating with building and district administrators, coaches also belonged to the project leadership team.

The project leadership team included two Wayne RESA mathematics consultants who were the co-principal investigators for the PM³ study, STEM faculty from the University of

Michigan-Dearborn, the project evaluator, the treatment district coaches, and the administrative support staff person who managed records and payments. This leadership group met for planning prior to each of the ten institute sessions and met somewhat more intensely at the beginning of the year to collaborate on the major content and pedagogical goals for the year and at the end of the year in preparation for the intensive summer institute week. The information from the coaches about their teachers' needs and classroom performance played an integral role in program planning. This link between the teachers, their coaches and the PM³ project was important to maintain at the institute meetings, also. It was important for the coaches to attend institute sessions to work with their teachers and to be knowledgeable about the content and strategies discussed at the institute. The coaches used these connections and information gained from institutes when they went on-site to support the teachers in implementation. Table 6 shows the approximate intensity of coaching in the project.

Table 6. Coaching intensity in PM³

Activities	Institute Session Group 1	Institute Sessions Group 2	Summer Sessions	Monthly Coaches Meetings	Project Leader Meetings	On-Site Coaching	Total
Number of Coaching Days	10 Days	10 Days	5 Days	10 Days	15 Days	100 Days	150 Days

After project level responsibilities for coaches were removed from the total of 150 days, about 100 days remained for the on-site activities of the PM³ coaches in this model. One of the major on-site roles of the district coach was communication with and support of building and district administrators concerning the PM³ project. This function was vital as other modes of communication with participants and administration—including electronic—were limited at times. Another allotment of coaching time that was not directly tied to coaching individual teachers in the project was when the coach facilitated the monthly, after-school, grade-level meetings. Each coach held two or three such meetings per month as indicated by the grade configurations within their district. The remainder of the coach's time, about 90 days, can be considered as the time a coach provided individual support for one of the 25 to 30 teacher participants in the district. On average, an individual teacher received about three days, or 18 to 20 hours of individual mathematics professional development per year from the coach. This is in addition to the time the teacher worked with the coach at institute meetings and at the collegial, collaborative, after-school meetings.

Grade-Level Meetings/Professional Learning Communities

The third major component of the PM³ professional development model was having participants gather regularly in smaller, job-specific groups called grade-level meetings. For example, all of the fourth grade teacher participants in a district gathered monthly to discuss their curriculum, make a common assessment, and/or examine samples of student work to gauge student progress. The teachers in a group supported others' work and provided each participant with trusted and known resources. These small collegial, collaborative groups met after school and teachers received a stipend for their time and participation. At the beginning of the project, these groups had major issues to examine including standardizing the content of the district curriculum as enacted in classrooms and determining ways to assess student learning. Over time, the grain size of these groups' discussions became smaller as the teacher teams increasingly focused on gaining greater clarity on student thinking, anticipating student misconceptions, and planning interventions when such misconceptions arose. As teacher participants found value in their collaboration with their colleagues, these small groups developed into professional learning communities similar to those described by DuFour and Eaker (1998) and became an ongoing source for professional learning and continuing instructional improvement. The continuance of these collegial meetings is crucial for teachers as they continue to monitor student thinking and collaborate on strategies to meet the learning needs for students as they enter their classrooms in the future.

Procedures

The PM³ mathematics professional development program started in the fall of 2004 with two sections of 25 to 30 participants—each including about 12 to 15 teachers from each district. While within district collaboration was important, cross-district alliances were also valued; these dual goals resulted in sections where teachers from the same building consistently attended together, but the school groupings rotated so participants had opportunities to work with everyone from both districts. Mixing the district participation in a section also required only half of the substitutes that would be needed if all teachers from a district attended on the same day. Despite the fact that it was difficult on students, teachers, and principals having teachers out of their classrooms to attend institutes, it was deemed necessary to hold the mathematics courses during contractual time to assure that ALL could participate, gaining deeper mathematical understandings, and improving classroom instruction for ALL students. For the "whole-district" approach to be utilized, the institutes needed to be held at a time when participants who really needed the professional development could not opt themselves out of participation for family or coaching responsibilities. The institutes needed to be held during the day for all students to have teachers who had deeper pedagogical content knowledge, demonstrated higher quality instruction, and collaborated with their colleagues about their students' learning of mathematics.

The data collection for PM^3 began at the first mathematics institute in the fall of 2004, when PM^3 participants took the Learning Mathematics for Teaching (LMT), the project's assessment tool measuring teachers' mathematics pedagogical content knowledge for teaching. Within a three-week window, the comparison group teachers also took their LMT pre-test. With this first administration of the LMT, both groups of teachers also filled out forms with demographic information about each participant such as: the grade being taught; their school's name; number of years of experience in the current assignment; total number of years of teaching experience; their mathematics preparation including their "highly qualified" status; gender; ethnicity. Information was also collected about the percentage of students in their schools with applications for free or reduced price lunch. The second administration of the LMT assessment for treatment and comparison teachers was after two years of the PM^3 professional development program in the spring of 2006; the third test administration of the LMT for treatment and comparison group teachers was in the spring of 2008—four years after the start of the PM^3 project.

In addition to testing on the LMT, each PM^3 project participant was observed during a period of classroom mathematics instruction using the SAMPI classroom observation protocol. The intervals for classroom observations were the same as the LMT with participants being observed at the outset of the project in fall 2004, after two years in PM^3 in spring 2006, and at the end of four years in spring 2008. Due to expense, comparison group teachers were not observed in their classrooms. Implementing the SAMPI classroom observation protocol required trained outside observers to complete an intensive professional development course and be certified prior to using the protocol in classrooms.

Gathering information from PM^3 participants about their experiences in the program was very important to the project leadership team. In addition to data collected from the LMT, the SAMPI, and the MEAP, other data collection resources were systematically used. The first of these were the insights gained by coaches out in the field when working in the schools and classrooms of their teachers. The coaches had monthly meetings to collaborate with each other on the progress or challenges of their teachers and also met with the leadership team to bring these experiences to the group for grant and institute planning. Another systematic data collection device was to reserve the last ten to fifteen minutes of each of the one hundred-sixty institute sessions for written evaluations from institute participants. Usually, the grant staff including STEM instructors and advisors would read these session evaluations and discuss the sessions' successes and challenges on that same day after the participants had departed. At every grant leader planning session, the previous sessions' evaluations would be considered as plans for the next session were discussed.

Focus groups were also used as a source for deeper probing into the participants' experiences in PM^3. In the second semester of every year and at other times as seemed warranted, an evaluator would hold a focus group to gather information about participants' views of the components of the professional development program and about their suggestions. Five or six focus group participants were randomly selected at an institute session by the evaluator and were given the time and place the focus group would be held. The grant funded stipends as a compensation for the time of the selected teachers. One additional source of information was from district personnel of the PM^3 districts such as curriculum directors and principals at twice yearly Grant Leadership Team meetings. These meetings were usually held in the early evening and would include a snack or meal.

The focus groups for teachers were found to be quite useful for gaining greater insight into how the program was being received and into what impacts the teachers reported as a result of their participation in PM³. The Wayne RESA mathematics consultants and the evaluator who conducted the focus groups collaboratively wrote the focus group questions. Some questions were part of each spring's focus group questions, while others were written specifically to gather information about something particular to the project at that time.

To collect qualitative data from focus groups, it is important to do everything possible to collect people's true beliefs and opinions. Therefore, only the independent evaluator was allowed to be in the room when the focus groups took place. All coaches and other institute personnel were removed and the groups were conducted away from Wayne RESA. Another step taken to minimize bias in responses was to ask individual participants to write their own responses on their questionnaire prior to any group discussion. Once the participants had completed their own answers, the evaluator gathered participants' responses and then facilitated a group discussion and took notes on participants' oral responses. Therefore, the data gathered included written comments from participants' questionnaires as well as data from notes taken by the evaluator during the subsequent discussions.

Table 7 shows a timeline for the major testing for teachers in the PM³ project, the LMT as the measure for pedagogical content knowledge for teaching mathematics, the SAMPI Classroom Observation Protocol as the measure for instructional quality, and focus groups as a measure of the perceived impact of the project.

Table 7. Participant test administration

Date	10/2004 Pre	2/2005	2/2006	5/2006 Post/Pre	2/2007	2/2008	5/2008 Post
Tests	LMT SAMPI	Focus Group	Focus Group	LMT SAMPI	Focus Group	Focus Group	LMT SAMPI

Table 7 summarizes the timing of the various assessments of the PM³ treatment teachers. Both the treatment teachers and the comparison teachers took the LMT, the assessment of pedagogical content knowledge for teaching, in the windows indicated in 2004, 2006 and 2008. The first administrations were in the beginning of the 2004 school year at the beginning of PM³, but the other administrations occurred in May of the school years. The SAMPI ratings of quality of classroom instruction were only performed on treatment teachers due to expense.

This part of Chapter 3, Chapter 3.0, has focused on a full description of the "Project: Making Mathematics Matter" professional development program. This chapter included the PM³ research design, research hypotheses, sample selection, instrument selection, the major components of the PM³ professional development treatment, and its procedures. The next part of Chapter 3, Chapter 3.1, concentrates on a full description of this study's ex-post-facto evaluation of data collected during the PM³ program as well as of this researcher's collection of student scale scores on the MEAP. Chapter 3.1 includes this study's research design, research hypotheses, sample selection and procedures for each analysis, and descriptions of the quantitative and qualitative data analysis to be performed for this research study.

CHAPTER 3.1

Research Design and Methodology for Dissertation Study

Research Design

This study is an ex-post-facto evaluation using data collected from 2004 through 2008 from "Project: Making Mathematics Matter," a quasi-experimental, matched comparison group research design mathematics professional development program described in Chapter 3.0. This study's purpose is to analyze the data collected during the PM³ program to see whether participation in PM³ had significant impact on measures of mathematics teachers' pedagogical content knowledge, on teachers' classroom instructional practices, and on students' mathematics achievement. The research questions remained the same as those for the PM³ project. The data under examination was collected through a federally-funded Mathematics and Science Partnerships (MSP) grant award to Wayne RESA in partnership with STEM faculty from the University of Michigan-Dearborn. This study will have a mixed-methods design with both quantitative and qualitative analyses of data. While the quantitative analyses will rigorously assess the LMT, SAMPI, and student MEAP scores for statistically significant differences, the qualitative analyses of participants' focus group responses may provide support for findings as well as a richer context within which the results reside.

Research Hypotheses

The null hypotheses for this study are as follows:

1. There will be no significant differences in mathematics content knowledge for teaching between PM³ treatment group teachers and comparison group teachers as measured on pre and post scores on the Learning Mathematics for Teaching (LMT). The independent variable will be the participant's membership in the treatment or

comparison group and the dependent variable will be the participant's score on the LMT.

2. There will be no significant changes in classroom instruction and practice of treatment teachers during and following their participation in the PM³ professional development project. The independent variable will be the length of time the teacher has participated in the PM³ project and the dependent variable will be rating scores on the Science And Mathematics Program Improvement (SAMPI) classroom observation rubric.
3. There will be no significant differences in student mathematics achievement in grades 4 through 8 as measured by state of Michigan MEAP mathematics scores between students in treatment districts and students in the comparison district prior to, during, and following treatment teacher participation in the PM³ project. The independent variable is whether students come from the treatment districts or the comparison district and the dependent variable is the student mathematics scores on the state of Michigan standardized annual test, the Michigan Educational Assessment Program (MEAP).

Sample Selection

Since the PM³ project determined the teachers who participated as treatment or comparison teachers, this study worked with the data already collected. The treatment and comparison teacher participant data analyzed for this study were received from the evaluator of the PM³ project in a variety of files containing some pieces of the needed data for participants by teacher identification number. The challenge for this study was to put together the data from these various files to determine the number of treatment and/or comparison teachers who had sufficient data for inclusion in each of the analyses. The sample selection data analysis for each of the three research questions will be described in the following sub-sections.

Sample for Analysis of Content Knowledge

For the first research question about differences in teacher mathematics content knowledge as a result of participation in PM³, matched pairs of treatment and comparison group teachers were needed. This research question required analysis of changes in scores of treatment and comparison participants on the LMT over the time of the project. Reliability and validity of the LMT instrument was described in the first section of this chapter. The LMT assessment was administered in the fall of 2004, the spring of 2006, and the spring of 2008, resulting in each teacher having three scores across time for data analysis.

Since participants were not randomly selected for participation in PM³, it was important to determine if the treatment and comparison teachers were reasonably well matched before analysis of the collected data could proceed. Success in the matching process would mitigate some of the harm caused by non-random selection. An additional element easing concern about PM³ participants not being randomly selected was that ALL teachers in the treatment districts were selected. As much as teachers in any district represent a cross section of teachers, the treatment group was thus composed. Comparison teachers represented any mathematics teacher of grades 4 through 8 students in the comparison district who were willing to use some of their personal time after school to do some mathematics testing for which they and the district received remuneration.

While the fall 2004 administration of the LMT showed scores for over 100 combined treatment and comparison participants, by the end of the PM³ research study in 2008, about 68 were found to have sufficient data to be considered for inclusion in the analysis of pedagogical content knowledge for teaching mathematics. This attrition represents a possible source of bias (Campbell & Stanley, 1963). Two of our original participants died of a rare cancer in the second year of the institutes. Sufficient information included participation in the first year of PM³, participation in the last year of the PM³, and having LMT scores for at least two out of the three testing cycles. For the 68 participants who met the preceding criteria, ten of 204 (68*3) scores from one of the administrations of the LMT were missing—seven for the comparison group and three for the treatment group. Since the missing data were about 4 percent any imputation method could be employed and the researcher chose averaging.

The treatment group had fewer participants (33) than the comparison group (35) with sufficient LMT information, so the next stage of this study was to look for a good comparison teacher match for each of the treatment group teacher using participants' demographic data. For this matching process, priorities needed to be established about characteristics' relative importance for mathematical reasoning. Data were collected about the school at which a participant works, the school's free and reduced lunch status, the grade the participant was teaching, how long the teacher had been teaching, how long teaching in the current position, the gender, the ethnicity, and the mathematics preparation. The first priority established by the researcher was having all three LMT scores to minimize imputation of data. After that consideration, the grade level of the teacher was deemed the most important component of a teacher's mathematical content knowledge due to the differing depths of mathematics content at each grade level. One might expect the depth of content knowledge of a middle school teacher with mathematics major to be greater than that of an elementary or special education teacher without special mathematics preparation. After considering grade level taught as highly important for matching participants, the prioritization of other characteristics for the matching process were gender, ethnicity, and years of experience.

Looking at the demographic information about treatment and comparison participants, it became immediately obvious that the mathematics preparation would not be useful in the matching process. While all but one of the comparison group teachers reported having a mathematics major and/or having a teaching certificate and teaching in grades 4 or 5, which automatically qualified these elementary, self-contained teacher as “highly qualified,” fourteen of the thirty-three treatment group participants in this research study began participation in PM³ without meeting “highly qualified” status. All but one of the treatment participants met “highly qualified” status after four years in PM³; this one participant still did not have a teaching certificate.

As the data were examined for the matching process, several additional differences between the groups were noted. These included differences in the distribution of grade levels represented in the two groups with the treatment group having a greater proportion of participants teaching grades 7 and 8 and the comparison group having a greater proportion of grade 4 teachers. Another difference between the groups was in participation of general education mathematics teachers contrasted with teachers trained for support of special education students or prepared to work with students who are English Language Learners. While the comparison group had only one non-general education teacher, a middle school special education teacher, over 20 percent of treatment participants were non-general education mathematics teachers with six trained special education teachers and one ELL teacher. Another difference was in the ethnicity of participants in the groups with the comparison group identifying two of their participants as African-American, while the treatment group included twelve African-American participants and one Arabic participant. One other slightly confounding variable for this matching process was that the comparison group teachers had, on average, about two and one-half more years of teaching experience than treatment teachers. The additional teaching experience of the comparison group teachers foreshadowed some challenges for matching. However, years of experience can portend advantages or disadvantages in various cases and differences in experience are probably most meaningful at the beginning of a teaching career (Darling-Hammond, 2000; Rockoff, 2004; Jacob, 2007).

Once the demographics of the two groups had been studied and noted, the initial phase of the matching process could commence. Special care was taken to make sure the only special education teacher in the comparison group was matched with a special education teacher from the treatment group with similar demographics. The treatment participant in this pairing did not meet the NCLB “highly qualified” definition, and the comparison teacher partner was the only teacher from the comparison group that might not meet the definition. This same care was taken for matching the five male teachers in each group, as well as with matching the African American male and female from the comparison group to participants with similar demographics within the treatment group. Once those matches were made, the remaining considerations for the rest of the teachers were grade level and years of experience. Each treatment teacher was matched with a comparison group teacher with similar demographics.

When this initial paring was completed, there were two grade 4 comparison teachers not matched with a treatment participant and a couple of the pairings toward the end of the process deserved reconsideration. Two of the grade 4 comparison group teachers ended up being removed from the sample for data analysis, because of the larger numbers of this grade level teacher in the comparison group. Since the stated priority was to maximize the number of actual scores (rather than imputed scores) in the analyses, reconsideration was given to the pairings to seek to eliminate two grade 4 teachers with a missing score. A few of the pairings were changed to assure each treatment participant had quite a number of commonalities with their comparison group match. The final sample selection of sixty-six participants would enter the analysis phase with 190 actual scores and only eight imputed scores on the 2004, 2006, and 2008 administrations of the LMT assessment of teacher mathematics content knowledge for teaching. This missing data represented about 4 percent of the data.

In this final sample, it was possible to match 32 of the 33 treatment teachers exactly with a comparison teacher at their grade level (26) or within one grade level (6). The thirty-third participant was matched with a comparison teacher two grade levels away of the same gender, ethnicity, general education status, and with each having over twenty years of teaching experience. Part of the issue with exact grade matching was the smaller proportion of comparison teachers of grades 7 and/or 8 teachers and the group's larger number of grade 4 teachers. Since the grade level of a teacher was viewed as the most important criterion, the ability to match these teachers so well on this characteristic was viewed positively. All participants were paired with a person of their own gender and the two African American teachers in the comparison group were matched with two African-American treatment teachers who shared additional similarities. Owing to the average of 2.5 more years of experience among comparison group teachers, in a number of instances, a treatment group teacher would need to be matched with a comparison group teacher with more experience.

While the number of years of experience was deemed an important dimension, it was not considered as important as factors with higher priorities in the matching process. However, due to a number of research findings that teacher experience, certification, and preparation do matter for teachers of mathematics (Fetler, 1999; Goldhaber & Brewer, 2000; Darling-Hammond, 2000; Laczko-Kerr & Berliner, 2002; Jacob, 2007), this study took a closer look at disparities in these areas in the teacher pairings. Interestingly, both groups had eight teachers with twenty or more years of experience and the mean years of experience for each group was approximately equal at twenty-six years. However, while four of these eight treatment teachers in this experienced group did not meet "highly qualified" status, all but possibly one comparison teacher did meet the "highly qualified" requirement.

At the other end of the experience scale, the comparison group contained no teachers with fewer than four years of experience, whereas the treatment group had five such inexperienced teachers. Of the five treatment teachers with little experience, two of them were not "highly qualified" for their assignment. In the final matching, these five participants with two or three years of teaching experience were matched with comparison group participants with four, five, seven, eleven, or fifteen years of experience. The larger differences in some of the pairings were results of higher priorities of grade level taught, gender, and ethnicity in the matching process. The remainder of twenty treatment teachers with between five and nineteen years of experience had eight more teachers who did not meet "highly qualified" status and were matched with teachers who met the definition. Although the overall difference in experience of 2.5 years may not be particularly meaningful in comparing twelve years of experience to ten years, the disparity in terms of teachers just beginning to learn to teach would tend to be in favor of the comparison group teachers.

At the conclusion of the matching procedure, this study had a sample size of sixty-six participants, thirty-three from the treatment group matched with thirty-three comparison group teachers. In this final matching, missing data were minimal with only eight of the 198 data points for LMT unknown. As previously discussed, these eight missing data points were imputed using an averaging method. The scores of these sixty-six participants would be analyzed over the course of the four year time frame of PM³ to address the first research hypothesis about differences in mathematics content knowledge for teaching between PM³ treatment group teachers and comparison group teachers as measured on pre and post scores on the Learning Mathematics for Teaching (LMT) assessment tool. A concern about these data was that some of the differences in "highly qualified" status and inexperience between the two groups might result in the significant differences in scores of treatment and comparison group teachers on the LMT at the outset of the project on the initial 2004 administration of the LMT. Due to the larger average years of experience and the smaller number of teachers not defined as "highly qualified" according to NCLB, might the comparison teachers' LMT scores be significantly higher?

Sample for Analysis of Quality Instruction

The second research question asks about improvements in treatment teachers' classroom instructional practices as the time of their participation in the PM³ professional development program increases. The study will compare the quality rating level of instruction in a participant's classroom at the beginning, midway through, and at the end of the PM³ project. As mentioned earlier, these classroom observations were not performed in comparison teachers' classrooms due to funding constraints. In addition to the expense of paying observers to do the classroom observation in comparison teachers' classrooms, there is also the issue of gaining comparison teachers' approval for such observations. This additional condition might have required additional compensation for comparison teachers and/or have seriously reduced the number of teachers willing to serve in this group. As was seen in the previous section, a smaller number of comparison participants would have been a problem for the matching of separate individuals for the analyses about teacher content knowledge.

For an activity as complicated as teaching to be reliably assessed, there must be agreement on what and how the classroom instruction and practice will be measured. The agreed upon instrument used in the PM³ project was an observational assessment tool used in many research studies across the state of Michigan and the nation (Kilday & Kinzie, 2009). The instrument is called the Science And Mathematics Program Improvement (SAMPI) Classroom Observation Protocol which has a well-defined rubric for each rating level and whose use is restricted to trained, certified observers. Additional information about the SAMPI measurement tool including the training of observers and the reliability and validity of the instrument can be found in Section 3.0 in the sub-section of instrument selection. The independent variable for PM³ participant teachers will be the testing cycle in PM³ (0 years, 2 years, or 4 years), and the dependent variable the participant's scores on the SAMPI summary scale as well as on each of the three SAMPI sub-scale scores.

The scores for this research were those collected by the PM³ professional development project in fall 2004, spring 2006, and spring 2008. Therefore, the sample for these investigations into participants' changes in quality of classroom instruction and practices came from the group of treatment teachers. The reason for reporting that the sample for analysis will come from the treatment group instead of saying it will be the treatment group is because several of the treatment teachers were missing a score from one of the three observational periods. Twenty-four of the participants have all three scores or 72 (24*3) out of 72 possible data points, but nine participants are missing one of their three scores. Although these nine missing data points represent only 10 percent of the 99 data points for thirty-three participants, concerns about imputing a starting or ending level for the quality of lesson requires close consideration including any differences the inclusion or exclusion of these participants' scores would make on the analysis.

The missing scores for the nine participants were at different times in the study. Since two of the teachers were missing their 2006 score midway into the project, the inclusion of these two teachers by imputing a middle score seemed unlikely to pose a risk to the outcome. The most crucial difference for PM^3 treatment participants was the quality of their mathematics instruction at the outset and at the completion of the intensive mathematics professional development program, PM^3. Therefore, the other seven participants were missing a data point from one of these more critical times—two teachers missing the initial rating and five teachers missing the final rating—and replacement considerations will occur during the analysis phase. The data will be analyzed with and without critical imputed data. In looking at the 2008 SAMPI observation data more closely, the quality of instruction had risen and all participants were rated at level 5, 6, or 7 on a seven-point scale. Based upon their ratings in 2004 and 2006, imputation of the final point might not pose much of a risk to statistical outcomes since four of these participants had been rated at level 6 on the seven-point rating scale in 2006.

Sample for Analysis of Student Mathematics Achievement

The third research question asks if there are differences in mathematics achievement between students taught by PM^3 treatment group participants in the two treatment districts and students taught in the comparison group district. The sample will be the students from treatment and comparison districts who took the State of Michigan's MEAP test in any grade 4 through 8 in any year starting with 2005 and ending in 2010. This time period was chosen for a number of reasons. Fall 2005 was the first administration of the new, annual, state assessments required for students in any grade 3 through 8 as required by NCLB. Immediately prior to the Fall 2005 MEAP, mathematics assessments were taken only by students in grades 4 and 8 and the content upon which these tests were based was different than the content blueprints for the new exams. Using scores from the earlier tests would only give more information about two grade levels and these scores could not be connected to the scores on the new MEAP. For further examination of district means prior to 2005, please see Figure 2 and Figure 3 in Chapter 1, pages 51 and 52 in Chapter 4.

Data Analysis

As previously discussed, this ex-post-facto research study uses a mixed-methods design with quantitative analyses of PM³ teachers' scores on the LMT assessment, of treatment teachers' scores on the SAMPI, and of student MEAP scale score from treatment and comparison districts. This evaluation also includes qualitative analyses of information gathered from focus groups conducted in the second semester of the years of study. The sections below will first discuss the quantitative analyses and then the qualitative analyses.

Quantitative Data Analyses

The first research question asks about significant differences in mathematics content knowledge for teaching between PM³ treatment group teachers and comparison group teachers as measured on pre- and post-test scores on the LMT. The independent variable in this measure is the teacher's status as a treatment group participant or a comparison group participant, and the dependent variable is the teacher's score on the Learning Mathematics for Teaching (LMT) scale over the four years of the PM³ study. These mathematics content knowledge for teaching data will be tested for significant differences using a mixed within and between repeated-measures Analysis of Variance (ANOVA). This test is appropriate for this data since the LMT data collected are three scores for the same individual over time—the assessment is repeated three times for each teacher. The within-subject factor for the analysis will be the testing cycle—2004, 2006, or 2008—and the between subjects factor will be the group to which the teacher belongs—treatment or comparison. The dependent variable will be a scalar measure between 0 and 32; the number reports the number of items the teacher answered correctly on that administration of the assessment. Prior to performing the repeated measures ANOVA, the 2004 means of the treatment group teachers and the comparison group teachers will be compared on an independent samples t-test to determine if the treatment group teachers and comparison group teachers scored significantly differently on the initial LMT testing cycle. When using the mixed within and between repeated measures ANOVA, there are a number of assumptions about the data to be met before using this multivariate test. Tests will also be done to look for any violations of the assumptions necessary to use the mixed within and between repeated measures ANOVA. Levene's test will be used to test the assumption of homogeneity of error variance within the LMT scores of the treatment and comparison teacher groups and Box's M will be used to test the hypothesis of homogeneity of the inter-correlation variance between the groups' LMT scores.

To respond to the second research question about the extent to which treatment teachers made significant changes in classroom instruction and practice during their participation in the PM^3, scores on the SAMPI classroom observation protocol from fall 2004, spring 2006, and spring 2008 will be analyzed. The SAMPI Classroom Observation Protocol rates the overall quality of the lesson under observation as well as rating it in three major components of high quality mathematics instruction—implementation of the lesson throughout the session, content of the mathematics discussed, and classroom culture for student engagement and learning.

Participants' scores at the beginning of the project will be compared to their scores after two years as well as their scores after four years. A repeated-measures ANOVA test will be used for this analysis with the number of years in PM^3 as the independent variable (0, 2, or 4) and the participants' SAMPI score as the scalar dependent variable (1, 2, 3, 4, 5, 6, or 7). The summary rating data as well as the data from each of the three major sub-skill areas mentioned above will be analyzed using one-way, repeated measures ANOVA with the testing cycle as the within subjects factor. Paired t-tests will be used to compare means.

The third research question in this study examines any significant differences in student mathematics achievement in grades 4 through 8 as measured by state of Michigan Educational Assessment Program (MEAP) mathematics scores between students of teachers in treatment districts and students of teachers in the comparison district during and following the PM^3 project. The student mathematics achievement data for this project will come from the state grades 4 through 8 mathematics MEAP assessments. The independent variable is whether the student is taught by teachers in the treatment districts or the comparison district and the dependent variable is the student scale score on the mathematics portion of the MEAP. Analyses of these scores will use mixed within and between repeated measures ANOVA. The within-subject factor will be the year of the MEAP administration (2005, 2006, 2007, 2008, 2009, or 2010) and the between-subjects factor will be whether the student is taught in a treatment district or the comparison district. To get a better picture of the shape of students' performance, this section will start with some graphs giving descriptive information about the change in percentages of students meeting state levels of proficiency. Some of this descriptive information will include comparing the lines of best fit for the MEAP proficiency data of treatment districts, the comparison district, and the state. Is the gap in mathematics performance narrowing?

Qualitative Data Analyses

For further substantiation of quantitative results, data gained during annual spring focus group sessions will be analyzed. This examination will concentrate on the prompts that ask participants to reflect on questions pertinent to this study's research questions about changes in content knowledge, in instructional practices, and in their students' mathematical thinking. The following questions represent the whole set of common questions for spring focus groups. This researcher's analyses focused on questions 3, 4, and 5.

1. What one thing did you like best about the Institute?
2. What would you most like to change about the Institute?
3. How have you changed your teaching/classroom practice as a result of the Institute?
4. What kinds of changes/impact have you seen in your students?
5. What effect has the Institute had on your understanding of mathematical content? What has made the most impact on your content knowledge in mathematics?
6. How are you interacting differently with your colleagues? other participants?

Procedure for Analysis of Focus Group Responses

The first step in the analysis process of these focus group data is reading through all of the participants' responses to look for recurring themes. Once such possible themes are identified, they are put on a list and all participant responses are re-read while tallying the number of participants who made similar responses about the major identified themes as well as a sense of the intent of the comment. After all participants' comments were coded by grouping comments relating to an identified theme, percentages of any particular comment on a theme will be determined by taking the number of participants making the comment and dividing this sum by the total number of people involved in the focus group being analyzed.

Table 8. Summary table of research questions, variables, and analyses

Research Question	Variables	Statistical Analyses
Are there significant differences in mathematics content knowledge for teaching between PM³ treatment group teachers and PM³ comparison group teachers as measured by pre and post assessments over the period of the PM³ project?	Independent Variable Teacher PM³ project membership status as a treatment group member or a comparison group member Dependent Variable Score on Learning Mathematics for Teaching instrument	*t-test to compare 2004 LMT means *Mixed within-between repeated measures ANOVA *One-way ANOVA to compare LMT group means 2004, 2006, 2008 *coding and analysis of participants' focus group responses
Are there significant changes in classroom instruction and practice of treatment teachers during and following their participation in the PM³ professional development project?	Independent Variable Time frame within the PM³ project when classroom observation occurs Dependent Variables *Change in classroom practice as measured by SAMPI classroom observation rubric *change in practice as reflected in teachers' focus group responses	*One-Way ANOVA to determine if the means of the SAMPI ratings are significantly different *summary SAMPI measure and subscale measures on SAMPI will both use this test *coding and analysis of participants' focus group responses
Are there significant differences in grades 4 through 8 MEAP mathematics performances between students in treatment districts and students in the comparison district relative to state scores before, during, and following participation in the PM³ project?	Independent Variables Students of teachers in T1, T2, or Comparison district, Year of testing Dependent Variable Student scores on mathematics section of the grades 3-8 state MEAP test (scale scores and descriptive scores).	*Two-Way ANOVA to look for sig differences in means over time between the districts *One-way ANOVA to compare means to find significant differences *coding and analysis of participants' focus group responses

CHAPTER 4

Results

The purpose of analyzing the data collected throughout the four-year duration of "Project: Making Mathematics Matter" was to assess the impact of this intensive, sustained, mathematics professional development program on teachers and students in these low-performing, high-poverty districts. The analysis might also provide additional insight about relationships among teachers' pedagogical content knowledge for teaching mathematics, the quality of teachers' mathematical instruction, and their students' scores on statewide standardized mathematics tests in high-poverty districts. There is a need in the literature for research projects specifically focused on improving mathematics instruction and learning in similarly high-poverty, low-performing districts.

Analyses of Pedagogical Content Knowledge

To assess impact of the four-year professional development program, "Project: Making Mathematics Matter" on the mathematical content knowledge for teaching of its participants, the LMT was administered to treatment and comparison group teachers in fall 2004, in spring 2006, and in spring 2008. While the sample size for the initial fall 2004 LMT testing was quite large with over 100 participants and over 50 from each group, the sample size of 66 participants for this analysis was more than adequate with 33 in each of the two groups. The analysis of changes in mathematics pedagogical content knowledge for teaching over the four year span of PM^3 will compare group means with 33 matched pairs of participants—each treatment teacher matched demographically with a comparison group teacher.

As mentioned earlier, the difficulty in securing comparison districts approximately similar demographically to these high-poverty treatment districts raised a hypothesis that the treatment and comparison groups might begin the study with significantly different mathematics content knowledge. To test the hypothesis that means of the groups were significantly different on the initial fall 2004 administration of the LMT, a one-way ANOVA was conducted on the 2004 treatment and comparison groups LMT scores. Some of the output from this statistical testing is reported in the following two tables, starting with descriptive information about the 2004 LMT scores for the treatment teachers and Table 10 follows with the ANOVA results. Not shown is the Levene's test which showed about a 60% probability that the data did not violate the assumption of homogeneity of variances ($p = .595$).

Table 9. Descriptive statistics for 2004 LMT scores

Groups	N	Mean	Std. Deviation	Std. Error	95 percent Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Treatment	33	15.27	5.137	.894	13.45	17.09	5	24
Comparison	33	16.58	5.420	.944	14.65	18.50	8	28
Total	66	15.92	5.281	.650	14.63	17.22	5	28

Table 10. Comparing group means for 2004 LMT Scores

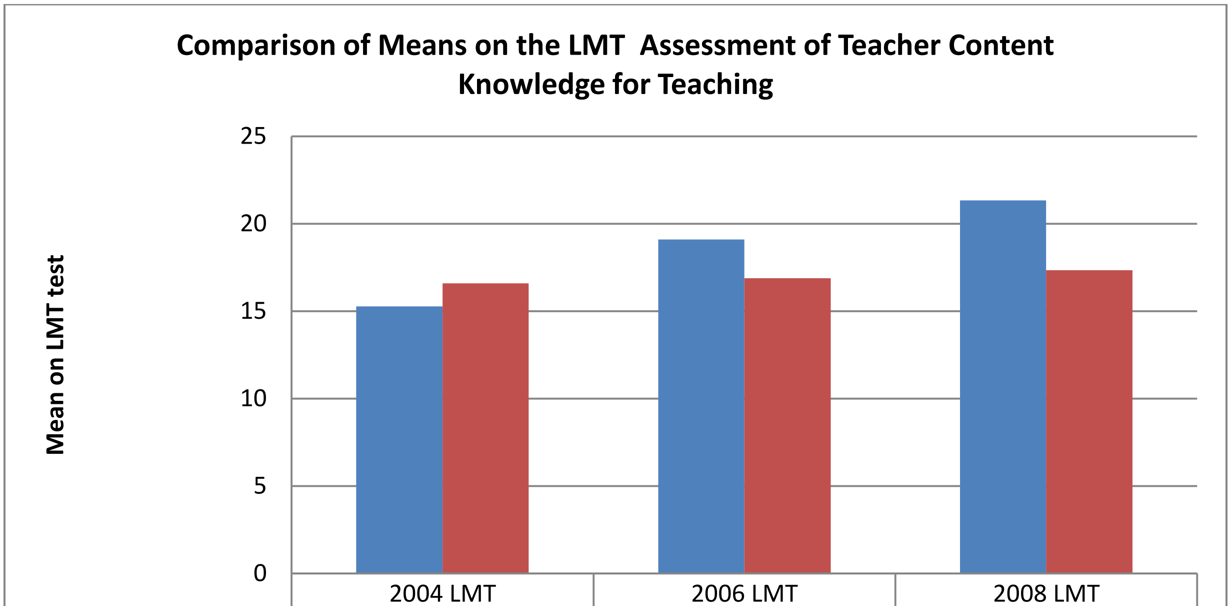
Source	Sum of Squares	df	Mean Square	F	Significance
Between Groups	28.015	1	28.015	1.005	.320
Within Groups	1784.606	64	27.884		
Total	1812.621	65			

Table 9 shows the descriptive characteristics for the treatment, comparison, and combined groups; it reports the means, standard deviations, and standard errors of the 2004 LMT scores for each group. It shows the 2004 treatment group average of 15.27 items correct on the LMT is 1.31 points lower than the comparison group average of 16.58 items, but it does not indicate if this difference is significant. Table 10 shows the results of the ANOVA statistical technique employed to test for significant differences between the treatment group and comparison group teachers' scores on the LMT assessment at the 2004 outset of the PM³ mathematics professional development program. The ANOVA test takes the ratio of the between groups variance estimate (28.015) and the within groups variance estimate (27.884) and reports an F-value of 1.005. This is less than the critical value of approximately 4.000 that would be necessary for these data with 1 degree of freedom between groups and 64 degrees of freedom within groups to indicate a less than 5 percent probability that their difference is due to chance.

In the last column of Table 10, the significance of this F-value is reported as .320 which is greater than this study's criterion significance level of $p < .05$. The .320 significance level indicates that about one-third of the time these differences in scores might happen by chance; this is much too high a probability for a Type 1 error when a significant difference is reported between the groups when in reality none exists. Thus, no significant difference is found between the treatment mean and comparison group mean ($p < .05$) on the 2004 administration of the LMT, an assessment of content knowledge for teaching mathematics. Although the treatment teachers on average scored 1.31 points lower on the 2004 LMT assessment than the comparison group, this difference was not found to be significant.

Before continuing with the quantitative analysis of the LMT data over time, Figure 4 shows general trends in the data for the two groups over the three testing cycles. The blue columns on the left show treatment group means and the red columns on the right show comparison group means. Below the columns, the actual means are given for the "Learning Mathematics for Teaching" assessment on which scores can range from 0 to 32.

Figure 4. Comparison of LMT Scores by group — 2004, 2006, 2008.



	2004 LMT	2006 LMT	2008 LMT
Treatment Group	15.27	19.09	21.33
Comparison Group	16.58	16.88	17.33

Figure 4 gives a visual picture of the changes in means of treatment and comparison teacher groups on the 2004, 2006, and 2008 administrations of the LMT. At the outset of the project, the treatment mean was slightly, but not significantly, lower than the comparison mean, by 2006 the treatment mean grew and was somewhat higher than the mean for the comparison group, and in 2008 the treatment mean increased again. In contrast, the means for the comparison group teachers stayed relatively stable over the 2004, 2006, and 2008 LMT testing. Without the table reporting the group's mean for each testing cycle, it would be difficult to visually judge whether the comparison group's means increased at each cycle.

Table 11. Additional LMT means comparisons

	2004 LMT Mean	2006 LMT Mean	2008 LMT Mean	Total Change in Mean
Treatment Mean	15.27	19.09	21.33	+6.06
Change in Treatment Mean	n/a	+3.82	+2.24	
Comparison Mean	16.58	16.88	17.33	+.75
Change in Comparison Mean	n/a	+.30	+.55	
Treatment Mean – Comparison Mean	-1.31	+2.21	+4	

Table 11 provides an opportunity to take a closer look at the means of the two groups, how each group's mean changes over time, and how the means compare to each other at each of the three testing cycles. In 2004, the treatment group LMT mean of 15.27 is 1.31 points lower than the comparison group mean of 16.58, however, this difference was previously analyzed and found to lack statistical significance. While the means for both groups increased between the start and conclusion of PM^3 from 2004 to 2008, the change is 6.06 points for the treatment teachers and only three-quarters of a point for comparison teachers. The treatment group mean began 1.31 points below the comparison group and ended up 4.00 points higher than the comparison group mean. Do any of these differences represent significant changes? Was the participation of treatment teachers in the intensive, mathematics professional development project PM^3 associated with significant increases in mathematics pedagogical content knowledge for teaching?

To determine if any statistical, quantitative differences do exist, statistical tests were conducted using the treatment and comparison groups teachers' scores over time. An appropriate statistical test to apply to these data to look for significant differences in mathematical content knowledge as measured on the LMT assessment is a mixed-method within-between-subjects repeated measures ANOVA. Since the same thirty-three treatment and thirty-three matched comparison teachers' LMT scores were analyzed over the four year length of the PM³ project (zero, two, and four years), this research study used repeated measures for two different groups—only one of which received the PM³ intervention—from 2004 and 2008. The between-factor for the comparison was whether the score was from a teacher in the treatment or comparison group, and the within-factor is the testing cycle from which the teacher's LMT scores came—beginning, middle, or end.

Testing for Violations of Assumptions

Since the test to be used for this analysis is a multivariate test, the data need to be examined for violations of the assumptions underlying applications of this statistical technique. The assumptions to be met for this multivariate test include many that multivariate tests in general require and one additional more unique assumption. The generally necessary data examinations for multivariate tests include: level or type of the dependent variable; use of random sampling; the independence of observations; normal distribution of data; and homoscedasticity of variance within and between groups. The additional assumption to be met for mixed-methods within- and between-subjects repeated measures ANOVA is the homoscedasticity of covariance within groups. Each of these assumptions will be discussed in turn.

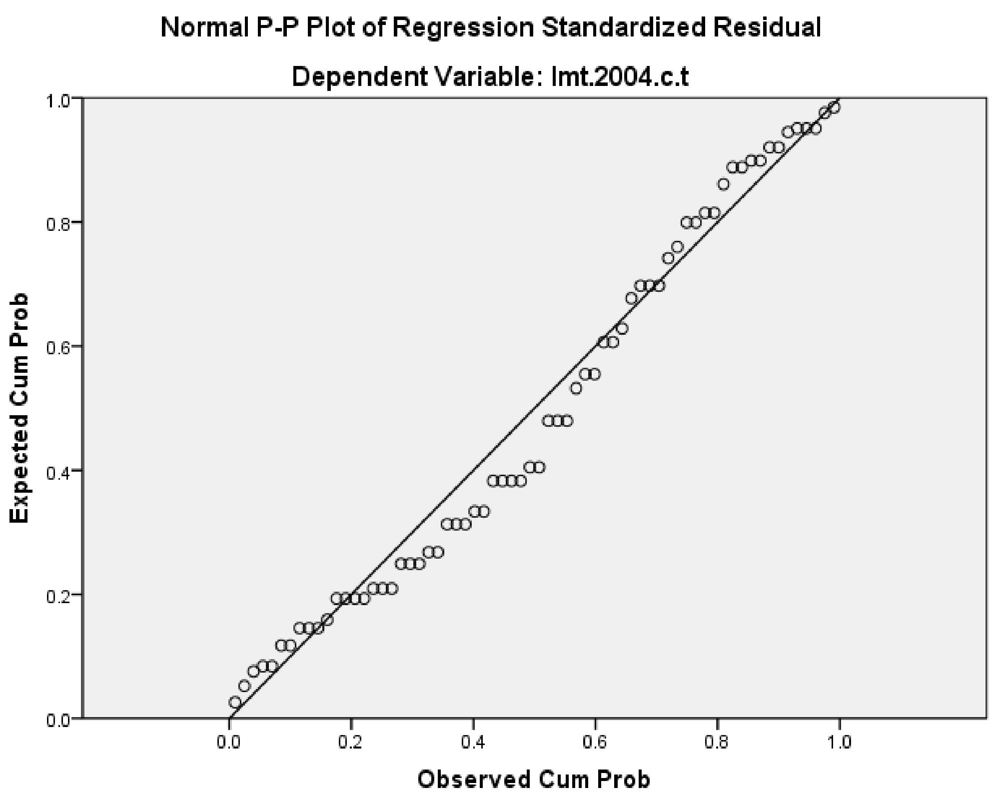
The first assumption, that the dependent variable for the analysis be interval or ratio, rather than nominal, is met. LMT scores can be between zero and thirty-two and the scores represent the number of correct answers on the test. The difference between a score of ten and eleven is one which is the same value as the difference between LMT scores of 25 and 26.

The next assumption is that the sample selection used in the study be random. Although the PM³ sample selection was not random, this research study matched each treatment group participant with a comparison group teacher to reduce bias that might be introduced by the non-random selection. Another factor minimizing selection bias for participants in the study is that in the PM³ design, treatment teachers were not selected. One significant component of the PM³ study was to take a whole district approach due to research findings that heightened mobility in challenged districts might make this a necessary ingredient for successful professional development programs. Once the two treatment districts had been identified, all teachers in the districts who worked with students in mathematics were to participate in the study. There was no bias about who to include and who not to include. As much as the entire faculties of mathematics teachers for grades 4 through 8 students in these two, low-performing, high-poverty districts represent other whole faculties for similar grades in similar districts, concern about selection bias would be minimized.

The third assumption is about the independence of observations and includes how and when the tests were administered, as well as individual learning that might have occurred after taking the LMT. LMT tests were administered by the evaluator under typical standardized testing conditions for quiet, independent work. No items involved participants working together. Treatment group teachers and comparison group teachers took their tests within a three-week window of each other. Since treatment and comparison teachers were from different counties and the treatment and comparison groups did not know the identity of the other group, it is highly unlikely that any communication passed between the two groups. The questions and/or their answers were not discussed with either group. However, it may be that some participants felt more comfortable in the second administration after having an experience with the test; the number of non-responses decreased from the first to second administration. This factor would be the same for treatment and comparison teachers.

Three assumptions remain to be evaluated—normality, homogeneity of variance, and homogeneity of co-variance. The first of these, normality is perhaps the most difficult to address statistically. Fortunately, the sample size for this study is large enough (greater than 30) that violations of this assumption should not cause major difficulties. Normality of the participants scores on the 2004 LMT will be considered by examining a normal probability plot of the standardized residuals and histograms of participants' LMT scores. The probability plot in Figure 5 shows the residuals between participants' expected 2004 LMT scores if the distribution was exactly normal and their actual 2004 LMT scores fall reasonably close to a straight line and this suggests the plausibility that these data are normal.

Figure 5. Normal probability plot for LMT residuals.



While all of these data do not fall exactly on the line, they are fairly close to falling on a straight line which would show the data as being normally distributed. While the pattern suggests the data may be somewhat platykurtic, it does show most residuals fall close to the line and are about equally distributed above and below the line. The following histograms give some more information about the normality of the 2004 LMT data.

Figure 6. Histogram of treatment teachers' 2004 LMT scores.

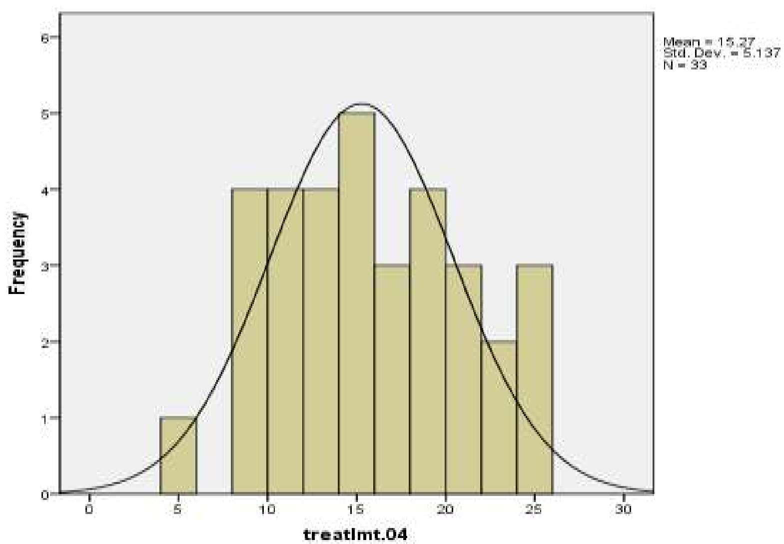
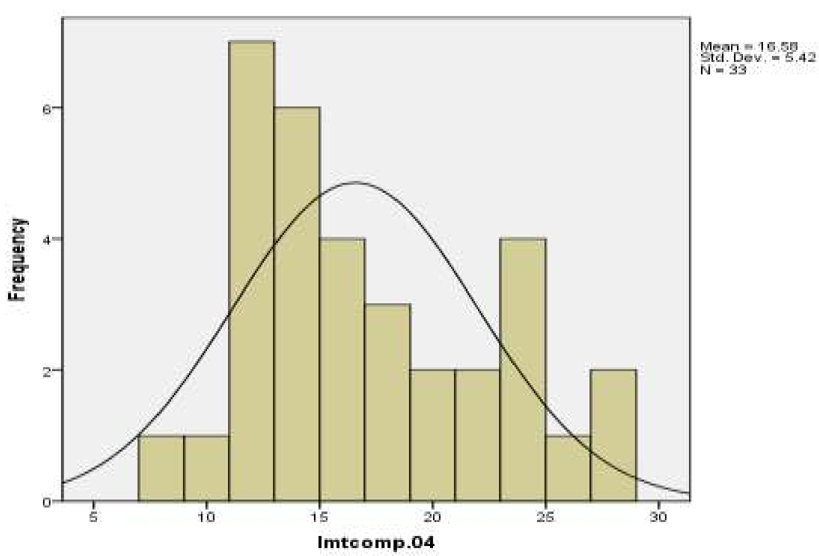


Figure 7. Histogram of comparison teachers'2004 LMT scores.



The histograms with the normal curve superimposed on them give another look at the normality of the data. While the normal probability plot shows the nearly linear residuals for all sixty-six of the participants, the histograms separate the 2004 LMT scores of the thirty-three treatment teachers and the comparison group teachers. The treatment group scores seem to show a slightly more normal distribution than the comparison group scores. The histograms indicate that the treatment group scores are fairly normally distributed, while the comparison group scores are somewhat positively skewed. However, these histograms are slightly misleading as the scale of the axes for the graphs are different; this can easily been seen by looking at the position of the score of 15 on the two x-axes. So although the LMT scores are not perfectly normally distributed, the variance is not large and the multivariate ANOVA is a robust statistical test and the score distribution should not be problematic since the sample size is over 30.

The final two assumptions for these data to meet before interpreting results from the multivariate statistical technique to be applied are univariate and multivariate homoscedasticity. These comparisons of homogeneity of variances within and between groups can be tested by statistical techniques. Both Levene's Test of Equality of Error variances and Box's Test of Equality of Covariance matrices were computed. In this study, Levene's Test was applied for equal variances in teachers' scores in the three testing periods. The Box's M statistic will investigate the co-variances between the groups of teachers and within each testing cycle. For both of these tests of homoscedasticity, the researcher is hoping these variances are not significantly different and is hoping the significance level of the Levene's and Box's M tests will be greater than .05.

Table 12. Box's M test

Box's M	5.199
F	.822
df1	6
df2	29676.679
Significance	.552

Table 13. Levene's test of equality of variances

	F	df1	df2	Significance
2004 LMT scores	.286	1	64	.595
2006 LMT Scores	.582	1	64	.488
2008 LMT Scores	.432	1	64	.514

Box's M results in Table 12 suggests the co-variances within the six groups identified both by group and time are not significantly different (p=.552). Each row in Table 13 contains all 66 of the participants' scores with coding to identify which scores belong to treatment or comparison teachers. Since the significance values for the scores from all three administrations of the LMT are greater than .05 (.595, .488, .514), Levene's test found no significant variances in treatment and comparison teachers within any of the three testing cycles. Additionally, Mauchy's test of sphericity (sig=.222) reports that the variances of the population difference scores across any two time periods are not significantly different than for any other two time periods.

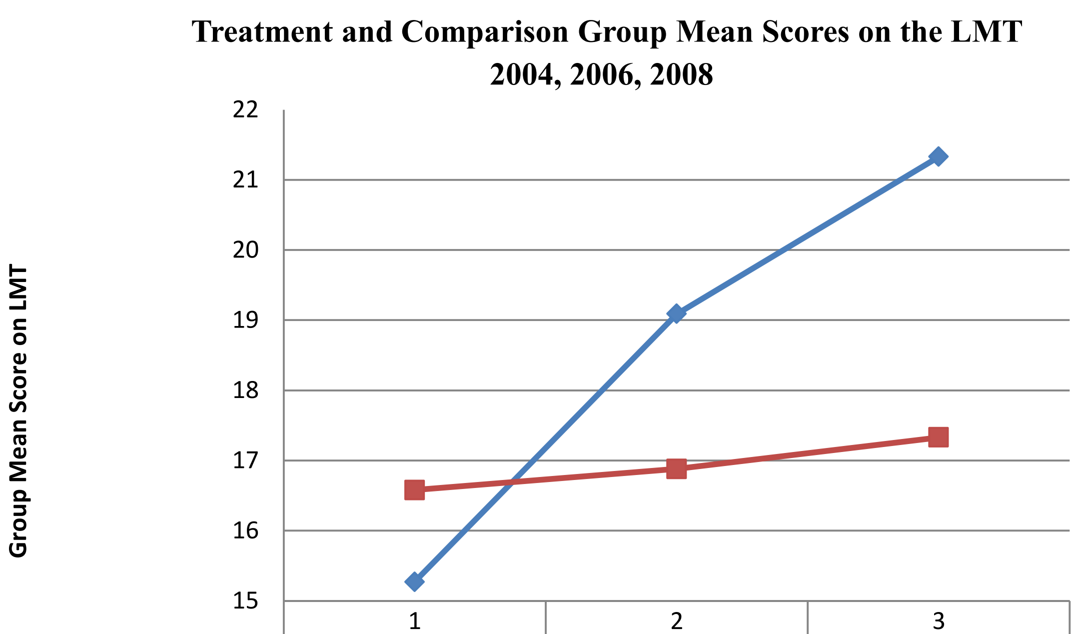
Now that the data have been examined for violations of assumptions and no serious flaws for using mixed between-within-subjects analysis of variance are identified, this statistical test will now be applied to the LMT data. The within-subjects factor will be the testing year— 2004, 2006, or 2008— and the between subjects factor will be group—treatment or comparison. This mixed-methods ANOVA will look for significant differences between LMT scores of teachers who received the PM³ professional development intervention and those that did not. These analyses helped answer the first research question about whether teacher participation in the PM³ professional development model was related to increases in pedagogical content knowledge for teaching. Descriptive information about these data sets were also requested and the means, standard deviations, and numbers of scores analyzed for each group can be seen in tabular form in Table 14 and the means are represented visually with a line plot in Figure 7.

Table 14. LMT mixed between -within-subjects repeated measures ANOVA results

Group		Mean	Std. Deviation	N
2004 LMT	Treatment	15.27	5.137	33
	Comparison	16.58	5.420	33
	Total	15.92	5.281	66
2006 LMT	Treatment	19.09	4.164	33
	Comparison	16.88	4.768	33
	Total	17.98	4.579	66
2008 LMT	Treatment	21.33	4.635	33
	Comparison	17.33	4.973	33
	Total	19.33	5.178	66

In Table 14, the means and standard deviations for the 33 matched pairs of treatment and comparison participants are compared on the 2004, 2006 and 2008 administrations of the LMT assessment for pedagogical knowledge for teaching mathematics. While the total means for the group of 66 increased about two points each testing cycle, from 15.92 to 17.98 to 19.33, the gain was nearly entirely due to the increases in the treatment group teachers' increasing means.

Figure 8. Means of LMT scores by group—2004, 2006, 2008.



	1	2	3
Treatment Group	15.27	19.09	21.33
Comparison Group	16.58	16.88	17.33

As can be seen on the Table 14 or the line plot of the group means on Figure 8, the mean score on the 2004 administration of the LMT for treatment teachers was lower than the mean for comparison teachers by 1.31 points. Two years after the start of PM³, the treatment teachers outscored the comparison teachers by 2.21 points and at the end of the PM³ project in 2008, the difference in scores was even larger at 4.00 points. The comparison teachers' scores increased slightly, but remained relatively stable across the three test administrations. On average, the treatment groups' mean increased 3.03 points per testing cycle, while the comparison teachers' mean increased an average of .375 points per time. Given these two rates of change, 3.03 points/cycle and .375 points/cycle, the rate of increase for treatment teachers was about eight times the rate of increase for comparison teachers (3.03/.375=8.08). Figure 8 also shows that the line plots displaying the interaction effect of group and time of test intersect and cross one another, indicating a disordinal interaction in which a group with a lower ranking at the outset doesn't just improve, it surpasses the group with a higher beginning ranking. The blue line segment displaying the LMT means for the treatment teachers begins lower than the red comparison group segment and ends higher. This was suggestive of a difference, but further significance testing was required and the test used to test for statistical significance was the mixed between- and within-subjects repeated measures ANOVA.

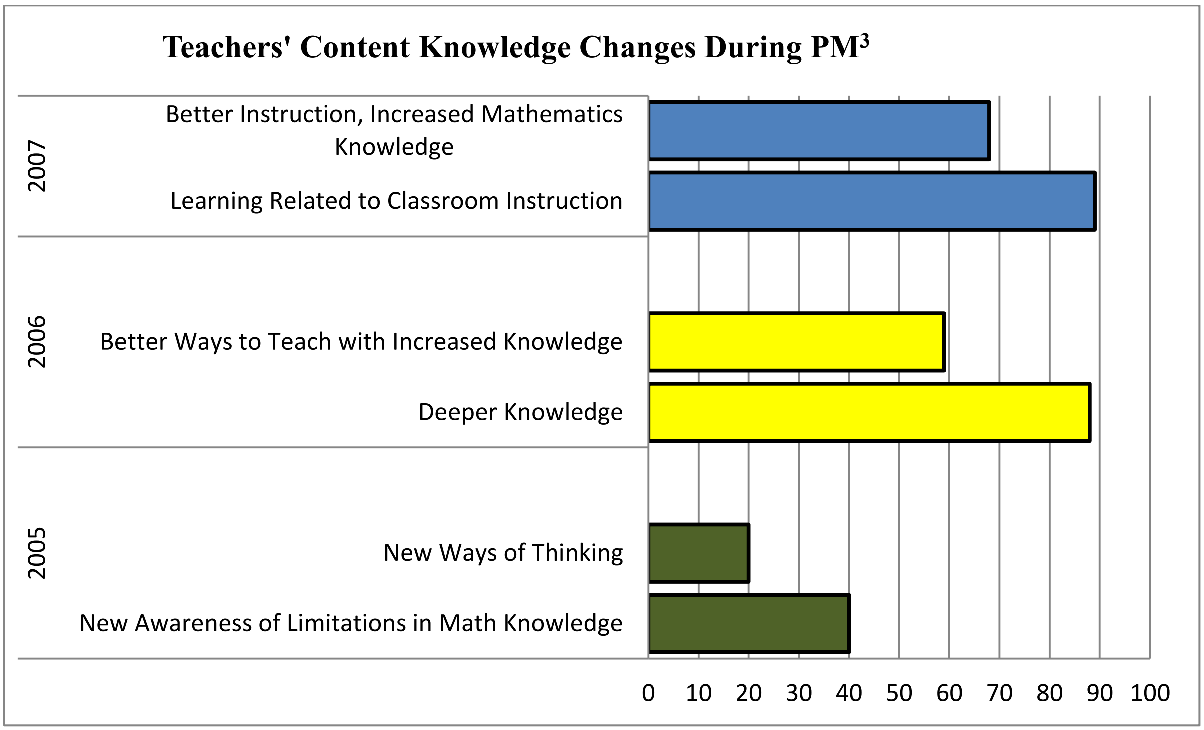
When the mixed between-within subjects repeated measures ANOVA was performed, the Wilks' Lambda multivariate results showed both testing time (2004, 2006, 2008) and the interaction between the group and the testing time were significant ($p < .0001$). While Figure 8 gives us an indication that scores are increasing through the years and that the testing time is a major predictor of LMT scores, the more meaningful effect found significant from this analysis for this research is the interaction between the group (treatment or comparison) and the testing cycle (2004, 2006, 2008) at $p <.0001$. In other words, the LMT scores varied differently over time for the treatment group teachers and the comparison group teachers. Yes, the increase in the treatment teachers' pedagogical, content knowledge for teaching mathematics as measured on the LMT during their participation in PM^3 was shown to be significantly different than for the comparison group at $p<.0001$. The treatment teachers' scores increased more than those of the comparison teachers. The effect size was large with SPSS output reporting a partial eta squared effect size estimate of .305 and Cohen's "d" effect size was calculated by the researcher as .758, and finally the observed power was high at .998. About one-third of the variance in scores can be attributed to the interaction effect of group membership and years of participation in PM^3.

Once this significant interaction between the two groups in the study was found, the researcher again used a one-way ANOVA in which the LMT scores of treatment group participants and comparison group teachers had significantly different means. From previous t-testing, the 2004 LMT means of the treatment and comparison group means were not found to be significantly different. The ANOVA also found the 2004 difference non-significant ($p = .320$), but found both the 2006 and 2008 means of the treatment teachers and comparison teachers to be significantly different with significance levels of $p = .049$ and $p = .001$ respectively. These results follow an intuitive sense that the difference in the 2008 means between the groups is very unlikely to have happened by chance, while the 2006 mean differences are not quite as obviously significantly different at a significance level of $p < .05$. It is also interesting to note that the greatest average gain in LMT scores for the treatment group was between 2004 and 2006 with a 3.82 increase. On the average, treatment teachers were able to correctly answer nearly 4 more of the items on this 32 item test and by the end of the study on average these participants in the PM^3 intervention answered six more items correctly.

The null hypothesis of no difference between the mathematics, pedagogical, content knowledge of treatment teachers and of their matched comparison teachers over the four year course of "Project: Making Mathematics Matter" from 2004 to 2008 is rejected ($p < .0001$). The only testing cycle in which the scores did not show a significant difference was in 2004 at the beginning of $PM^{3.}$ The treatment teachers who received the mathematics professional development intervention named, "Project: Making Mathematics Matter," grew significantly more in their pedagogical content knowledge for teaching mathematics than did the teachers who were in the PM^3 non-intervention, comparison group.

For additional insights into this research question about mathematical content knowledge differences for treatment teachers, qualitative analyses of the annual spring focus group responses will be analyzed on the following question, "What effect has the Institute had on your understanding of mathematical content?" This question asks the teachers to reflect upon their perceptions of what they are feeling or experiencing in regard to their pedagogical content knowledge. Findings of these analyses will be found in Figure 9.

Figure 9. Focus group responses about content knowledge changes.



After six months of participation, the teachers' responses lacked much focus and many reported not knowing what to say. The only two responses that were raised multiple times was a sense of having new realizations about what they did not really know about mathematics (40 percent) and some were noting new ways of reasoning with mathematics. A year later, the types of responses showed more focus on mathematical knowledge and on their beliefs that their content knowledge was growing and deepening (87 percent) and also that this content knowledge combined with new ways of teaching resulted in improved classroom instruction (59 percent). This finding will also be important for the analysis about classroom practice as well. As early as one and one-half years into the program, the treatment teachers were making this critical connection between content and pedagogy in their classroom instruction.

In the analysis of the spring 2007 focus group data, many of the responses were similar to those of the previous year noting a connection between deeper content knowledge and better instruction (69 percent). A new theme also emerged in these 2007 responses with nearly 90 percent of the participants noting that the mathematics content they had been learning was related to the content they were teaching in their classrooms. Both of these themes indicate that treatment teachers felt participation in the PM³ intervention was related to increases in their pedagogical, content knowledge for teaching mathematics in their classrooms. Another strong theme from these focus group responses about mathematical content knowledge was the link many teachers made between their increases in mathematical content knowledge and their improved classroom instruction. This study now moves to quantitative analyses to examine changes in classroom practice and instruction.

Analyses of Instructional Quality

The second research question concerns the classroom practices of PM³ treatment group teachers and asks if significant changes in classroom instruction and practice of treatment teachers occur during their participation in the PM³ professional development project. The assessment tool described earlier for measuring changes in classroom practices is the Science and Mathematics Protocol Instrument, or SAMPI. The PM³ treatment participants were observed during one of their usual, daily mathematics' instructional sessions at the beginning of the project in the fall of 2004, midway through the project in the spring of 2006, and at the end of the program in the spring of 2008. The people observing teachers' classrooms were mostly retired educators who had attended the training and became certified in the use and scoring on the SAMPI rubrics. The ratings awarded on items could range from a low of 1 to a high of 7 based upon comparing the observed lesson to the SAMPI descriptions for each item. The SAMPI protocol has a rubric giving additional information about ratings and as discussed earlier, observers needed to become certified trainers by calibrating scoring so ratings across observers would be equivalent.

For each of the three sub-scale areas of content, implementation, and classroom culture, the evaluator had between seven and ten items to measure which served as the basis for that sub-scale rating. In addition to ratings in these three focal areas, the evaluator awarded a summary rating for the lesson observed on the same 1 to 7 scale. Prior to the observation, the evaluator touched base with the teacher to determine if there are any unique circumstances in the class to be observed—such as special needs students—and to know where the lesson to be observed fits into the unit being studied. While observing in the classroom, the evaluator may take notes, but does not fill out the rubric while in the classroom. The ratings are placed onto the rubric after the observer has thanked the teacher and has found a private place to complete the scoring.

As discussed in Chapter 3.1, the sample for analyzing the SAMPI measures of lesson quality will be performed with different numbers of participants. The sample of 26 does not include the seven teachers missing either the first or last of their three observations. The major reasons a participant missed a SAMPI classroom observation during one of the windows were being out on medical leave or being temporarily laid off. While the background information for these data will use the sample of 26, tests for significance will also be run with the seven scores imputed for the sample with 33 treatment teachers.

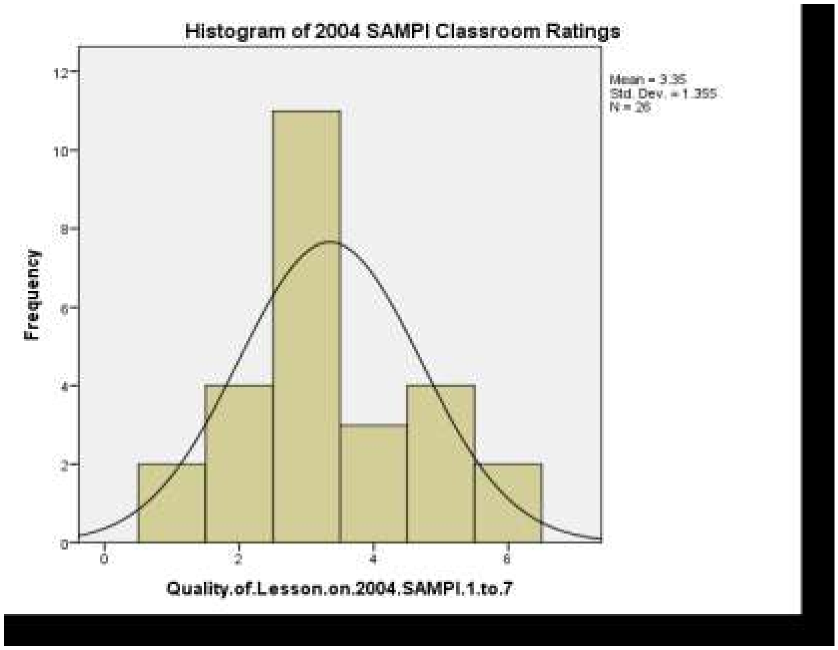
In the analysis of the SAMPI ratings, the independent variable was the testing time, 2004, 2006, or 2008, and the dependent variable was the rating on a scale from 1 through 7 on the quality of each item being measured.

|1 |2 |3 |4 |5 |6 |7

Since these data being analyzed are ratings of the quality of classroom instruction for the same treatment group teachers gathered at three different times, repeated measures ANOVA will be used. Before using this statistical technique, the data were inspected for violations of assumptions underlying the use of this test. As noted above, the ratings are scalar satisfying the level of measurement. Concerning the issue of random sampling, it has previously been mentioned that while these teachers were not randomly selected, they really were not selected. Once the two high-poverty, low-performing districts were selected, ALL teachers who taught mathematics to students in those districts were to attend. The PM³ project used a whole district approach as researchers have suggested for challenged districts (Miles et al., 2005; Loucks-Horsley & Matsumoto, 1999)

Each lesson rating was scored by a trained, certified observer and the teachers were not shown the SAMPI rubric or the items on the protocol on which they were being assessed. So while each observation was independent, participants in PM³ might have learned from the instructional modeling at the mathematics institutes and from the discussions about instructional decisions to gain insight into the effective instruction the SAMPI was seeking. Figure 10 shows the histogram of 2004 SAMPI ratings with the normal curve superimposed. While not perfectly normal, the data are reasonably close to normal; if a couple of the 3 ratings had been 4 ratings instead, it would be quite close to a normal curve.

Figure 10. Histogram of 2004 SAMPI ratings.



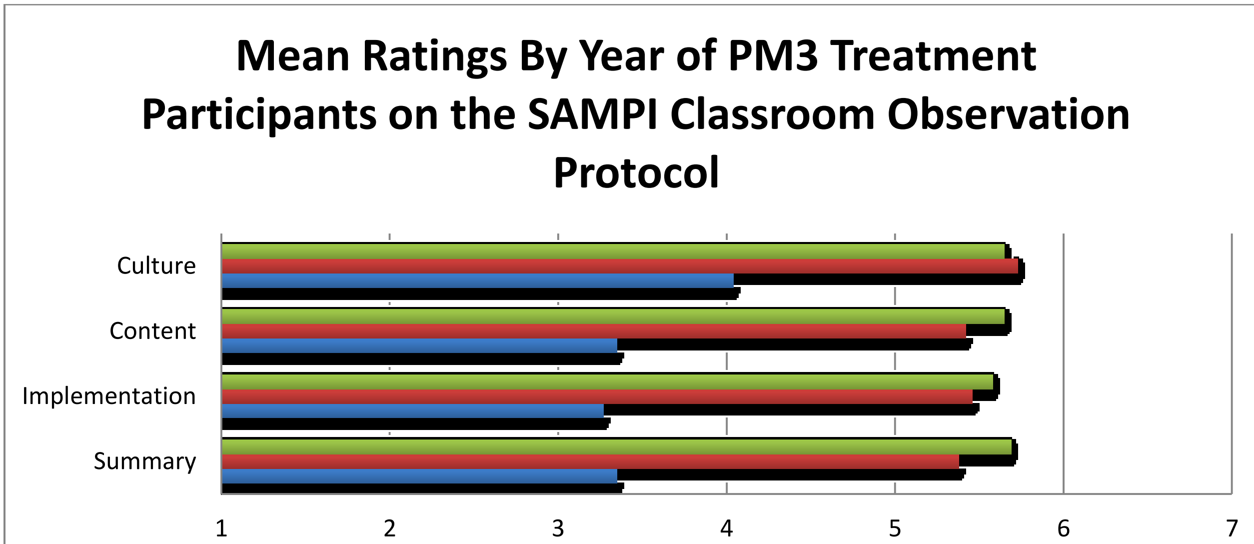
With assumptions for its use reasonably met, the repeated measures ANOVA was performed on two groups of treatment teachers, the subset of 26 and the larger set with 33 teachers with 9 of 99 scores imputed. Table 15 begins the reports from the repeated measures ANOVA performed on the SAMPI classroom quality ratings by reporting means and standard deviations for the summary or overall lesson ratings with the statistics for the group with 26 participants on the right and those for the group with 33 participants on the left.

Table 15. Comparisons of SAMPI descriptive statistics for N=26, N=33: Summary

N	SD	Mean		Mean	SD	N
33	1.372	3.48	SAMPI 2004	3.35	1.355	26
33	.936	5.42	SAMPI 2006	5.38	.898	26
33	.663	5.76	SAMPI 2008	5.69	.679	26

When looking at the means for both groups in Table 15, the overall lesson quality increases approximately 2 points—a substantial gain on an instrument with a 7 point scale. The mean summary ratings from the 2006 observations to the 2008 observations showed only a small increase of about three tenths of a point. These data are very similar and Figure 11 will use and show the data from the group with 26 participants for visual inspections of the ratings. In addition to the means shown in Table 15, these bar graphs will also show the means for each of the SAMPI sub-scales. These sub-scales are implementation of the lesson, mathematical content of the lesson, and the classroom culture during the lesson.

Figure 11. SAMPI ratings for treatment teachers on lesson quality.



	Summary	Implementation	Content	Culture
2008	5.69	5.58	5.65	5.65
2006	5.38	5.46	5.42	5.73
2004	3.35	3.27	3.35	4.04

In Figure 11, the three bars at the bottom of the figure labeled "Summary" show the means for 2004, 2006, and 2008 for the sample of 26 participants. The large increase in length from the bottom blue bar to the middle red bar and the small increase from the red to the top green bar are visually presented. Shown in the bars above the overall or holistic summary rating of the observed lesson are the three sub-scale ratings for implementation of the lesson, mathematics content in the lesson, and classroom culture during the lesson. The patterns seen in all four sets of bars graphs is nearly the same. Each had a rating of about 3.3 in 2004, experienced a large increase to about 5.4 in 2006, and grew slightly to 5.6 in 2008. The only slight variation in this pattern occurred with classroom culture ratings where the 2006 rating was the highest of any of the means and its mean decreased slightly in 2008. Despite the fact that these two, high-poverty districts were experiencing multiple changes in district and building leadership, these teachers' scores indicate they were sensitive to student learning needs in their classrooms.

As reported in Table 15, these SAMPI data are quite similar whether the analysis includes the original group of 26 or the 33 when seven scores out of 99 are imputed or whether it is an analysis of any of the three sub-scales. For further comparative analyses of the SAMPI summary ratings of these PM³ teachers, additional output from the repeated measures ANOVA follows in Table 15. The testing cycle is the independent variable and the SAMPI summary rating is the dependent variable performed using the Bonferroni correction for multiple comparisons.

Table 16. Comparisons of SAMPI repeated measures ANOVA Results: Summary

Sample	Wilks' Lambda	F	Sig.	Partial Eta Squared	Observed Power	Time as Source of Within Contrasts	F	Sig.
SAMPI 26	.240	38.03	.000	.760	1.0	Linear	59.746	.000
SAMPI 33	.262	43.62	.000	.738	1.0	Linear	77.320	.000

Once again the output from the repeated measures ANOVA testing shown in Table 15 report similar results for the two samples for SAMPI analysis, with the first row showing analysis of the subset of 26 participants and the second row giving output for all 33 participants. The Wilks' Lambda value of .24 with $F(2, 24) = 38.03$, leads to the conclusion of significant differences ($p < .0001$) in SAMPI ratings through the years of participation in PM³. For the group of 33, the Wilks' Lambda value of .26 with $F(2, 31)$ is 43.62, indicating the changes in SAMPI ratings are significantly different at $p < .0001$. The numerical values and findings are similar for the two groups. The effect sizes are also very large for both samples with about 75 percent of the variance explained by the time in the program (2004, 2006, or 2008). When Cohen's effect size was computed using the 2004 and 2008 means, the effect size was again VERY large with a value of 2.116. This large an effect size poses some questions about the sensitivity of the SAMPI tool. Observed power is high at 1.000. The ANOVA results performed with either 26 or 33 participants indicate the same conclusion; the SAMPI classroom quality ratings are significantly different over time. This finding suggests a difference, but does not tell where the ratings differ. Pairwise comparisons among all possible pairings of SAMPI ratings using the Bonferroni adjustment follow in Table 17 to clarify significant differences.

Table17. Pair-wise comparisons of SAMPI means, n=26: Summary

(I)testing cycle	(J) testing cycle	Mean Difference (J - I)	Std. Error	Sig.
2004	2006	2.038	.282	.000
	2008	2.346	.282	.000
2006	2004	-2.038	.282	.000
	2008	.308	.282	.355
2008	2004	-2.346	.282	.000
	2006	-.308	.282	.355

These pair-wise comparisons confirm what may have already been predicted from consideration of Figure 8; the differences between participants' 2004 and 2006 SAMPI ratings are significantly different ($p < .0001$) as well as between their 2004 and 2008 means ($p < .0001$). However, the increase between the 2006 and 2008 SAMPI scores did not meet the test for significance at $p = .355$ when using the Bonferroni correction. When the pairwise comparisons were performed on the three sub-scales of SAMPI, the implementation of the lesson, the mathematical content of the lesson, and the classroom culture during the lesson, this same pattern of significance emerged. Next, these same pairwise comparisons using the Bonferroni correction for multiple comparisons will be computed for the sample with 33 participants.

Table 18. Pair-wise comparisons of SAMPI means, n=33: Summary

(I)testing cycle	(J) testing cycle	Mean Difference (J - I)	Std. Error	Sig.
2004	2006	1.939	.213	.000
	2008	2.273	.258	.000
2006	2004	-1.939	.213	.000
	2008	.333	.161	.139
2008	2004	-2.273	.258	.000
	2006	-.333	.161	.139

When comparing results for the samples of 26 or 33 treatment participants as reported in Table 17 and Table 18, the means differ slightly, but the statistical results are the same. For both groups of treatment teachers (26 or 33), the increases in lesson quality are significant between the original SAMPI assessment in the fall of 2004 and the next SAMPI observation two years later in the spring of 2006 and between fall 2004 and spring 2008 at the end of the study.

However, the increases for both samples between the 2006 and 2008 ratings are not significant. It appears that the imputation of the seven missing data points does not have a significant impact on the means comparisons. Going forward in analyses of these data, either sample could be used for the analysis, but the sample of 26 with fewer missing data points will be used.

Given the appearance of the data on Figure 11, with the large ratings increase of at least two points on a seven point scale when comparing means in 2004 with those in either 2006 or 2008, these findings of significant growth are not particular surprising. However, the lack of significant growth between 2006 and 2008 does raise some interesting questions. Why did the participants' classroom practices not continue the growth seen between the first two rating cycles? Are four years of an intensive, professional development project for teaching mathematics unnecessary? Did the PM^3 professional development project stop providing the support needed for continued growth? Was the SAMPI observation tool implemented correctly? Was SAMPI sensitive enough to distinguish the amount of change occurring in the quality of classroom instruction? Did the scale of 1 through 7 result in participants experiencing a "ceiling effect" for their instructional ratings?

These questions led back to an examination of the raw data collected during the SAMPI classroom observations of instructional practices in participants' classrooms. Although the mean of the 2004 summary SAMPI ratings was below the midpoint of a 1 to 7 rating scale at 3.35, six

|1 |2 |3 x |4 |5 |6 |7

of the twenty-six PM^3 participants included in this analysis were rated at a summary level of 5 (n=4) or level 6 (n=2). The mathematics lessons implemented by these six teachers during their first observation already demonstrated high levels of lesson implementation, mathematics content, and classroom culture and these teachers could only improve their rating by one or two points or regress to the mean. At the outset of the PM^3 project, another six of the twenty-six teachers were rated at level 1 (n=2) or level 2 (n=4) and these teachers had an opportunity to improve their instructional quality by five or six levels. The initial quality of instruction demonstrated by the PM^3 teachers varied widely. This conclusion is supported both by the examination of the histogram of 2004 raw scores themselves (Figure 10) and by their large standard deviation of 1.355 points on a seven point rating instrument. The level of instruction in sixty percent of these PM^3 teachers' classrooms was rated low between 2.0 and 4.7—not a uniformly high level of instruction that might enable low-performing students in these challenged districts to accelerate their learning!

Within two years, the mean level of instructional quality had risen to 5.38 and sixty

|1 |2 |3 x |4 |5 x x |6 |7

percent of the classrooms were rated between level 4.5 and 6.2. The mean rating for overall lesson quality level in the final observation at the end of the 2007- 2008 school year increased slightly to 5.69 and the standard deviation in ratings had shrunk to .679, one-half of the 2004 statistic. In 2008 with the observed lesson quality in sixty percent of these classrooms rated between 5.0 and 6.4 and with raw scores showing all twenty-six participants rated at quality level 5, 6, or 7, the students in these classrooms in these challenged districts received considerably more uniform, higher quality instruction in 2006 and 2008 than in 2004!! So while the first two years of the PM^3 professional development program may have been sufficient for many teachers in the project to demonstrate higher quality instruction, some needed additional time.

Returning to the questions concerning the lack of significant growth in instructional quality between 2006 and 2008, there may be other factors to consider concerning the smaller gains. With a 2006 mean rating of 5.38 and a highest possible rating of 7.00, one problem for the same rate of growth to occur between 2006 and 2008 is the impossibility of it numerically. Beginning with a 2006 mean of 5.38 and adding 2.00 growth points as occurred between 2004 and 2006 would result in a mean of 7.38 which is larger than the top quality rating of 7. Participants' 2006 ratings were beginning to "top out" or experience a "ceiling effect" on the SAMPI tool. Participants rated at level 7 in 2006 could not demonstrate additional growth on this instrument in 2008 and any change in rating would have to be a lower one.

Examining the raw ratings of participants, it appears quite difficult to earn a 7 rating. The total number of summary ratings of 7 on the SAMPI was quite small with zero in 2004, two in 2006, and only three in 2008. An examination of the descriptors for a level 7 rating in any of the sub-scales provides further insight into the difficulty of demonstrating level 7 instructional quality. For example, the level 7 descriptors on the implementation subscale items are: teacher shows great confidence in facilitating lesson; classroom shows strong student/teacher interaction emphasizing higher order thinking, probing questioning, and exposing prior student knowledge; classroom management is very effective in engaging students; the lesson is well paced for the developmental level of the students; the student/student interaction is very productive for discussing and gaining a deep understanding of the content; the lesson was organized to allow considerable time for teacher and students for reflection on the lesson and content; and finally considerable time was devoted for closure of the lesson.

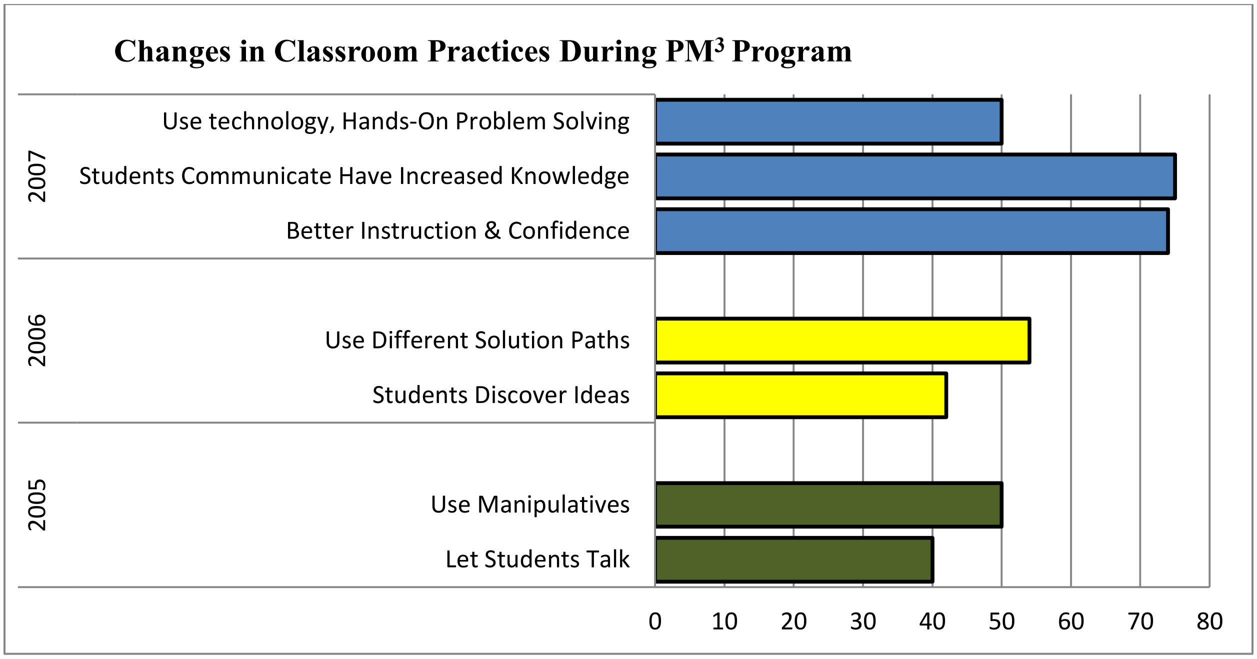
These represent less than one-third of the items upon which instructional quality of the teacher's lesson is rated during a SAMPI classroom observation. With this new insight into the demand of a 7 rating, the 2008 mean rating level of nearly 6 (5.69) would seem to represent a very high level of instructional quality and would help in supporting the hypothesis of topping out playing a role in the lack of additional significant growth between 2006 and 2008. The summary SAMPI rating mean of 5.69 or nearly 6 in 2008 as compared to 5.38 in 2006, indicates maintenance of the much higher level of instruction PM³ teachers demonstrated after two years of participation in this mathematics professional development project.

The null hypothesis that there will be no differences in the quality of mathematics instructional practices during teachers' participation in PM³ is rejected. While a few participants in the study began with higher quality lessons, the mean for the group was quite low at 3.35 and by 2008, this mean measure of quality of lesson had increased to 5.69—with ALL teachers rated at level 5,6, or 7—with the mode of the data being 6.

At the end of the section which analyzed teachers' increases in their pedagogical, content knowledge for teaching mathematics, the qualitative analyses of participants' responses to focus group questions about perceived changes in their content knowledge showed the teachers made a strong connection between their increases in mathematical knowledge to improved instruction. These teachers' perceptions appear to have been borne out in the quantitative analyses so far which have shown significant increases in treatment teachers' content knowledge for teaching mathematics and in the quality of their classroom instruction.

For additional and deeper insights into participants' thoughts about changes in classroom instructional practices, analysis will be performed of the focus group question asking, "How have you changed your teaching/classroom practice as a result of the Institute?" Figure 10 shows the themes that emerged through the years in response to this question and the percentages of teachers who made a similar statement about the theme.

Figure 12. Focus group teachers' responses about their instruction.



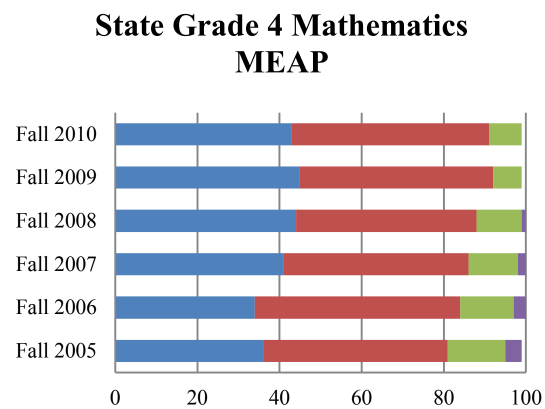
On Figure 12, the years in the institute increase from the bottom to the top. The themes most often mentioned in responses during the first spring focus groups shown in green on the bottom of the figure were teachers' use of manipulatives and letting students talk. While both of these may be quite positive, just having students use manipulatives or students talking in class does not necessarily mean better instruction is occurring; it is unclear. However, the responses in 2006 about students discovering ideas and over 50 percent of the teachers mentioning classroom use of different solution paths for problems is more indicative of more focused meaningful practices in classrooms. Students appear to be engaged in their learning of mathematics including finding and sharing the different ways they solved a problem. By the third set of spring focus groups, over 70 percent of the teachers note having more confidence when teaching mathematics and providing their students with better instruction. Two other themes had multiple responses; the first was a mention of the power of using technology and hands-on problem solving in their instructional practice. The last major theme of teacher responses again provides a transition into our final research question about increases in student mathematics achievement during their teachers' participation in PM³. When asked about changes in their instructional practices, over 70 percent of the treatment teachers observed that as a result of students communicating their mathematical ideas with one another, teachers perceived students have increased mathematical knowledge. In the next section, student mathematical performance will be examined.

Analyses of Student Mathematics Achievement

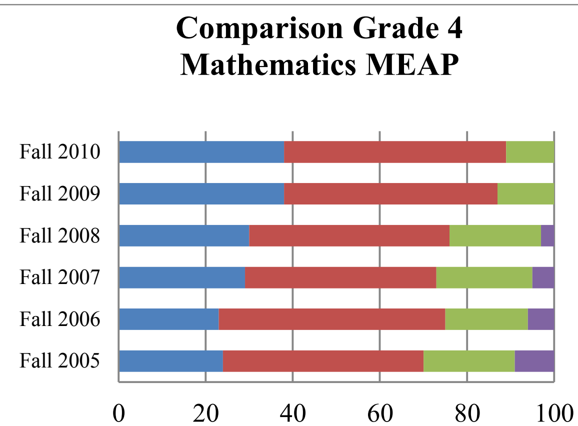
The examination of student mathematical performance during and after the time their teachers participated in PM³ will include descriptive examination of treatment and comparison MEAP scores, statistical analyses of student scale scores from treatment and comparison districts using One-Way ANOVA, lines of best fit, and analyses of qualitative data given by treatment teachers about changes in their students. These analyses will begin with bar graphs by grade of the mathematics MEAP results from fall 2005 through fall 2010 including the state-wide average, the comparison district average, and the two treatment district averages.

While the fall 2005 student MEAP testing occurred a full year after the beginning of teacher participation in the PM³ program, there really is no great pre-test for measuring student growth in mathematics. For analysis of student changes on the state-wide test during the life of the project, it would have been wonderful if the October testing of grades 3 through 8 on the new test would have begun in 2004. Certainly a test given in October of the year the project was beginning would not yet reflect increases due to the project. However, around that time the state changed the testing standards and when the test occurred. Table 3 and Figures 2 and 3 (p. 49-52) show the comparisons for the grades 4 and 8 for 2002 through 2004, before the beginning of the PM³ project. It was a different test and had measures for only two grades. Therefore, for this study, the first student tests that will be used are after a year of teacher treatment in fall 2005 and may not reflect the achievement of students prior to their teachers' participation in PM³.

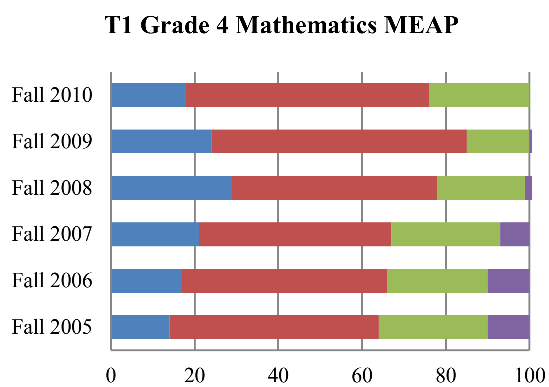
Figure 13. Grade 4 MEAP comparisons 2005-2010.



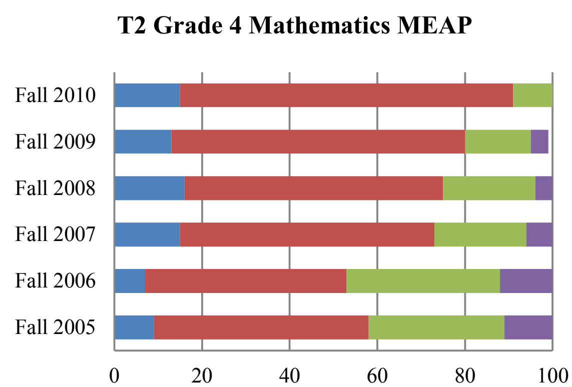
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	36	34	41	44	45	43
Level 2	45	50	45	44	47	48
Level 3	14	13	12	11	7	8
Level 4	4	3	2	1	0	0



	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	24	23	29	30	38	38
Level 2	46	52	44	46	49	51
Level 3	21	19	22	21	13	11
Level 4	9	6	5	3	0	0



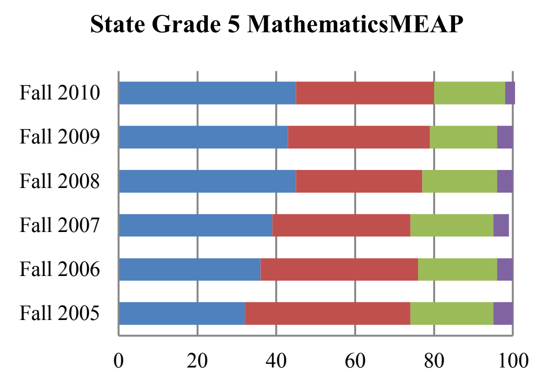
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	14	17	21	29	24	18
Level 2	50	49	46	49	61	58
Level 3	26	24	26	21	15	24
Level 4	10	10	7	2	1	0



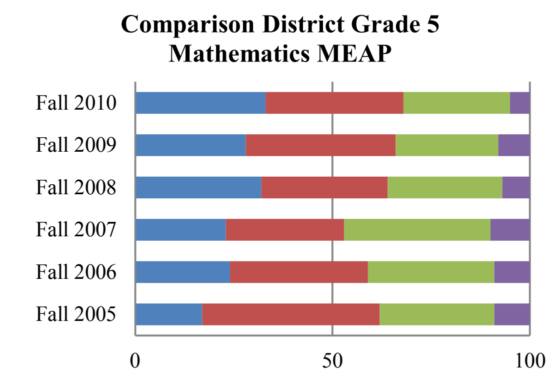
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	9	7	15	16	13	15
Level 2	49	46	58	59	67	76
Level 3	31	35	21	21	15	9
Level 4	11	12	6	4	4	0

While the fourth grade teachers participated in the institutes, the MEAP test was administered in the beginning of October. The impact of one month of instruction by a PM^3 participant is questionable. However, it is noteworthy that T2 achieved the state-wide average of 91 percent of its students rated as proficient in 2010. This is far from this district's three year average between 2002 and 2004 of 24 percent MEAP proficiency. While the state's percentage of students rated "advanced proficient" and "proficient" between 2005 and 2010 improved from 36 to 43 and from 45 to 48 respectively, in T2 the increase in proficiency in these categories was from 9 to 15 and 49 to 76 respectively. For T2 nearly all of the change is seen in decreases in the percentage of students in Levels 3 and 4 and in a large increase in Level 2, the "proficient" category. The percentage of T2 students in the "advanced proficient" category is still quite low.

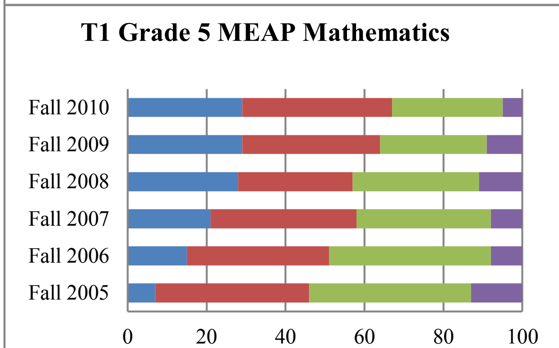
Figure 14. Grade 5 MEAP comparisons 2005-2010.



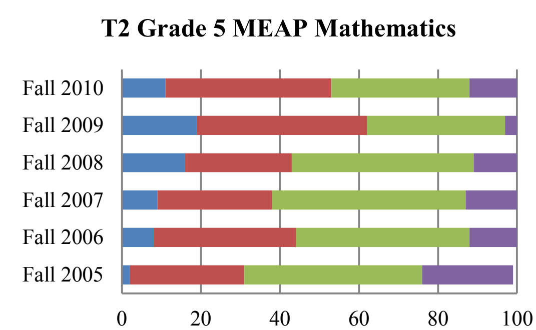
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	32	36	39	45	43	45
Level 2	42	40	35	32	36	35
Level 3	21	20	21	19	17	18
Level 4	5	4	4	4	4	3



	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	17	24	23	32	28	33
Level 2	45	35	30	32	38	35
Level 3	29	32	37	29	26	27
Level 4	9	9	10	7	8	5



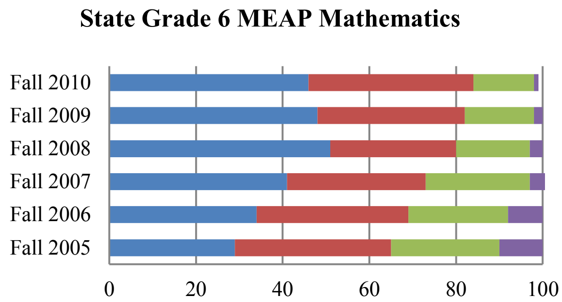
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	7	15	21	28	29	29
Level 2	39	36	37	29	35	38
Level 3	41	41	34	32	27	28
Level 4	13	8	8	11	9	5



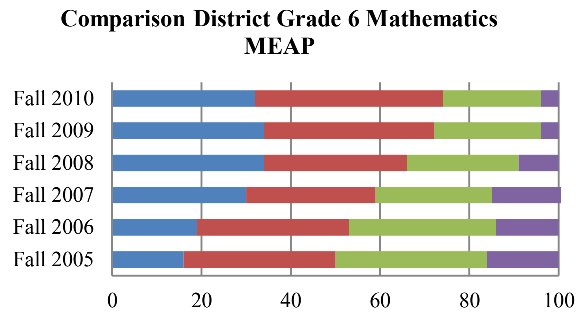
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	2	8	9	16	19	11
Level 2	29	36	29	27	43	42
Level 3	45	44	49	46	35	35
Level 4	23	12	13	11	3	12

For grade 5, both the state and comparison district scores remain fairly stable, while both treatment districts show a slow, but steady increase. District T1 shows a slow steady growth going from 46 to 67 percent proficiency while the comparison group went from 62 to 68 percent.

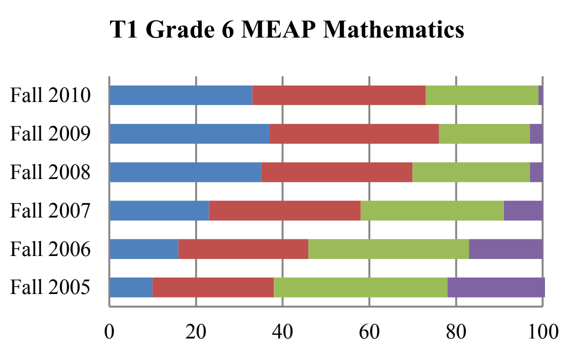
Figure 15. Grade 6 MEAP comparisons 2005-2010.



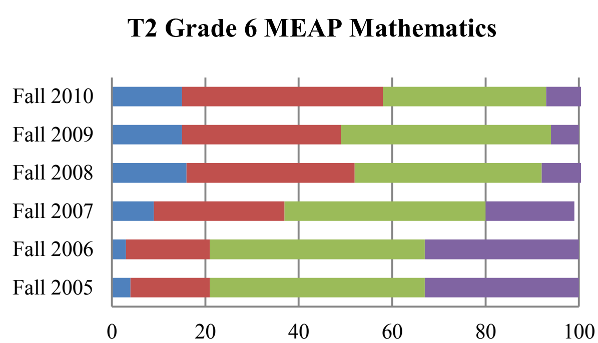
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	29	34	41	51	48	46
Level 2	36	35	32	29	34	38
Level 3	25	23	24	17	16	14
Level 4	10	8	4	3	2	1



	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	16	19	30	34	34	32
Level 2	34	34	29	32	38	42
Level 3	34	33	26	25	24	22
Level 4	16	14	16	9	4	4



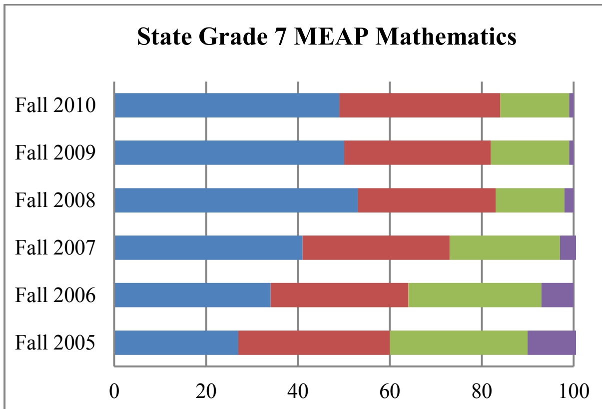
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	10	16	23	35	37	33
Level 2	28	30	35	35	39	40
Level 3	40	37	33	27	21	26
Level 4	23	17	9	3	3	1



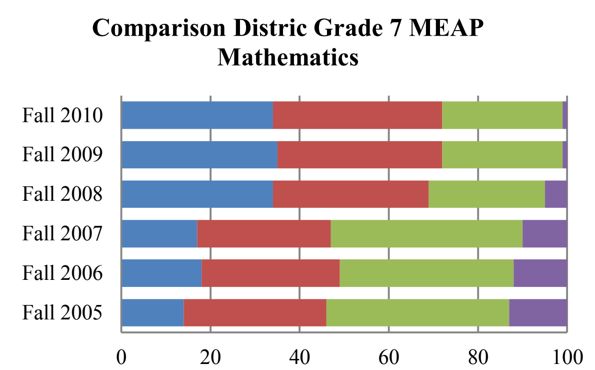
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	4	3	9	16	15	15
Level 2	17	18	28	36	34	43
Level 3	46	46	43	40	45	35
Level 4	33	33	19	9	6	8

The increases in scores for the state and the comparison group are similar with the state's proficiency rates always about 10 points higher than the comparison district's mean. The changes in both treatment districts are a bit more variable and the raw gain in means is about 35 points while the state and comparison group increase by about 25 points.

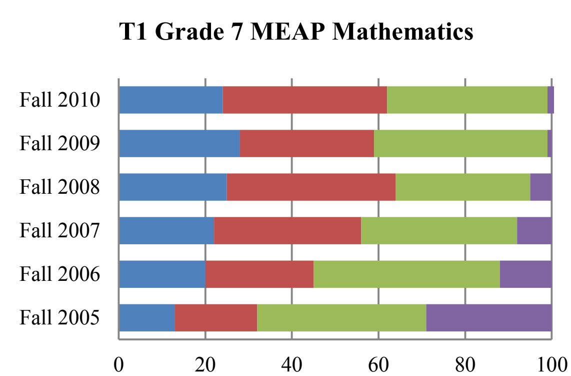
Figure 16. Grade 7 MEAP comparisons 2005-2010.



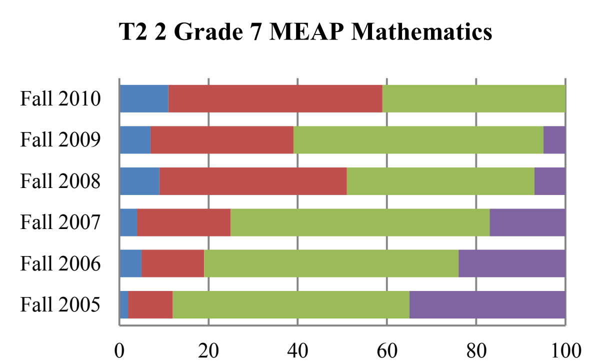
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	27	34	41	53	50	49
Level 2	33	30	32	30	32	35
Level 3	30	29	24	15	17	15
Level 4	11	7	4	2	1	1



	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	14	18	17	34	35	34
Level 2	32	31	30	35	37	38
Level 3	41	39	43	26	27	27
Level 4	13	12	10	5	1	1



	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	13	20	22	25	28	24
Level 2	19	25	34	39	31	38
Level 3	39	43	36	31	40	37
Level 4	29	12	8	5	1	2

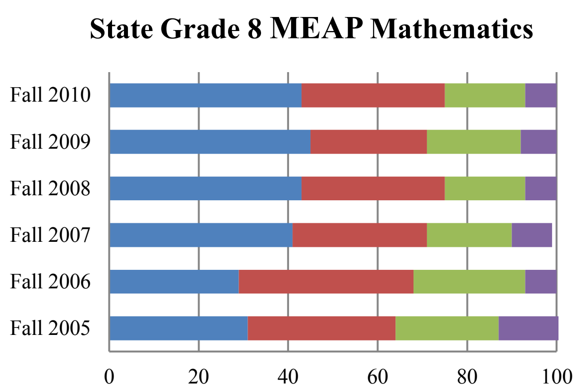


	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	2	5	4	9	7	11
Level 2	10	14	21	42	32	48
Level 3	53	57	58	42	56	41
Level 4	35	24	17	7	5	0

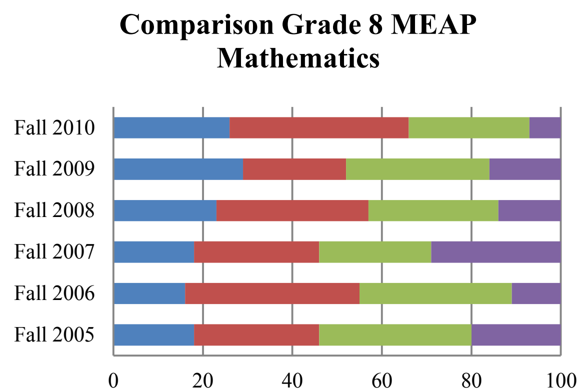
Although all four sets of bar graphs for grade 7 show increases, one noticeable difference for the treatment districts is their elimination, or near elimination of “Not Proficient” students. The increase in proficiency rate for the comparison district between fall 2007 and fall 2008, from

47 to 69 percent, is very large while for the other years the rate stays fairly level. For both treatment districts, there is an uneven but steady growth.

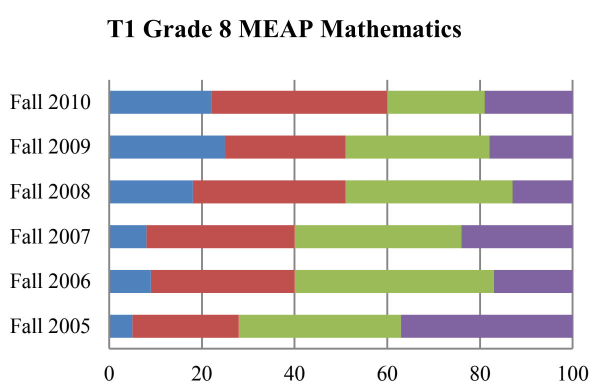
Figure 18. Grade 8 MEAP comparisons 2005-2010.



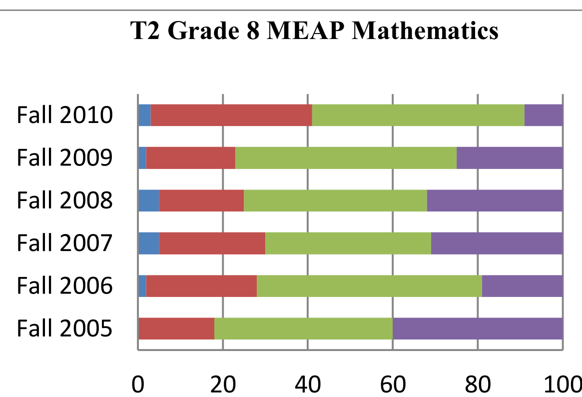
	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	31	29	41	43	45	43
Level 2	33	39	30	32	26	32
Level 3	23	25	19	18	21	18
Level 4	14	7	9	7	8	7



	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	18	16	18	23	29	26
Level 2	28	39	28	34	23	40
Level 3	34	34	25	29	32	27
Level 4	20	11	29	14	16	7



	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	5	9	8	18	25	22
Level 2	23	31	32	33	26	38
Level 3	35	43	36	36	31	21
Level 4	37	17	24	13	18	19



	Fall 2005	Fall 2006	Fall 2007	Fall 2008	Fall 2009	Fall 2010
Level 1	0	2	5	5	2	3
Level 2	18	26	25	20	21	38
Level 3	42	53	39	43	52	50
Level 4	40	19	31	32	25	9

All three of these low-performing districts are struggling in Grade 8 mathematics. The comparison district made a good gain between 2009 and 2010. District T1 shows a nice steady increase that took their students from 28 percent proficiency in 2005 to their most current proficiency rating of 60 percent. District T2 has reduced the percentage of their level 4 students from 40 percent to 9 percent.

After the previous visual examination of the student MEAP data that showed the changes in the percentage of students in each of four categories, the statistical analyses will use the student scale scores. MEAP student scale scores are determined when range finders and statisticians come together after the MEAP test has been given. This group determines the items a student in each of the grades should be able to answer correctly and they determine the cut scores for the each of the four levels in each grade. Once this process has been completed, a student who has correctly answered only the questions the group identifies as a minimal level of knowledge for a proficient fourth grade student is assigned a scale score of 400—the start of the range for grade 4 students in level 2 or the "proficient" category. For example, in 2010, the grade level scale score ranges were as follows:

Table 19. 2010 MEAP cut scores by grade

Grade	Level 1 Advanced	Level 2 Proficient	Level 3 Partially Proficient	Level 4 Not Proficient
3	327-427	300-326	279-299	205-278
4	432-554	400-431	378-399	285-377
5	527-671	500-526	477-499	358-476
6	622-761	600-621	580-599	472-579
7	722-873	700-721	676-699	581-675
8	820-961	800-819	784-799	676-783

*Source is Michigan Department of Education's MEAP Website

Describing the whole procedure of range finding and cut scores is beyond the scope of this study.

While the treatment districts are both challenged, they are also unique in their student bodies and communities. The analysis of student scores will therefore be performed while keeping each of the three districts' scores separate. As has been seen in a variety of tables and figures, T2 district students had the lowest scores going into the study with T1 scores somewhat higher, and the comparison district students scoring even higher. The research question guiding these analyses is whether there are significant differences in grades 4 through 8 MEAP mathematics scores between students in treatment districts and those in the comparison district.

A two-way ANOVA was performed on the approximately 4000 student scale scores for each grade over six testing years for the three districts. Each student scale score was coded for district and year. Prior to the ANOVA testing, techniques to detect anomalies were performed on the data and those data identified as outliers inappropriately affecting the outcomes were removed. The dependent variable for these analyses was the scale score and the independent variables were the district from which the score came and the year, 2005 through 2010, it occurred. Five, two-way ANOVAs were conducted; one for each of the grade levels 4 through 8.

Table 20. Results from 2-way ANOVA on student scale scores, grades 4-8

Grade	Sig. of interaction Group * Year	Partial Eta Squared	Power	Pairwise Comparisons Sig. Level T1 & C	T2 & C	T1 & T2
4	.048*	.004	.857	.000*	.000*	.009*
5	.039*	.004	.872	.000*	.000*	.001*
6	.068	.004	.829	.443	.000*	.000*
7	.000*	.007	.993	.000*	.000*	.000*
8	.002*	.006	.977	.000*	.000*	.000*

*significant at p < .05

The interaction between the group and year of testing was significant for grades 4, 5, 7, and 8 and the significance level for grade was .068. The effect sizes are quite small, but perhaps due to the large sample size, the observed power is above 80 percent for the significant interactions. The null hypothesis of no differences in student scales scores across groups through the years is rejected for all pairwise comparisons of T1, T2, and C, with the exception of that in sixth grade between T1 and C. With the knowledge that significant interactions are occurring through the years relating to the whether the students are taught in district T1, T2, or C, the locations of these differences must be determined.

To search for the indicated differences in the mathematical growth of students as indicated by their MEAP scale score, a series of means comparisons were made. One-way ANOVAs with Tukey's post hoc tests of multiple comparisons were performed for each year at each grade level to see where the means were or were not significantly different. For example, starting with 2005, each grade level of coded sets of student scale scores from that year were compared using a one-way ANOVA with the independent variable being the district, T1, T2, or C, and the dependent variable was the student scale scores. Therefore, five one-way ANOVAs were conducted for the year, one for each of the grades 4 through 8. This process of comparisons of student scale scores using one-way ANOVAs was repeated for each year, 2005 through 2010. In Table 21, the top section with the three comparisons for 2005, the student sets of scores from the treatment districts are shown to be statistically significantly different than the comparison district at all grade levels. The treatment districts' scores in grades 4 and 8 do not show a statistical difference ($p < .05$). The significance level from the Tukey's post-hoc multiple comparisons are shown in the cells in Table 21. While the analysis was conducted for sixth grade, the main interaction effect at that grade was .068 and failed to reach significance at $p<.05$.

Table 21. Significance level of students' MEAP scale scores comparisons

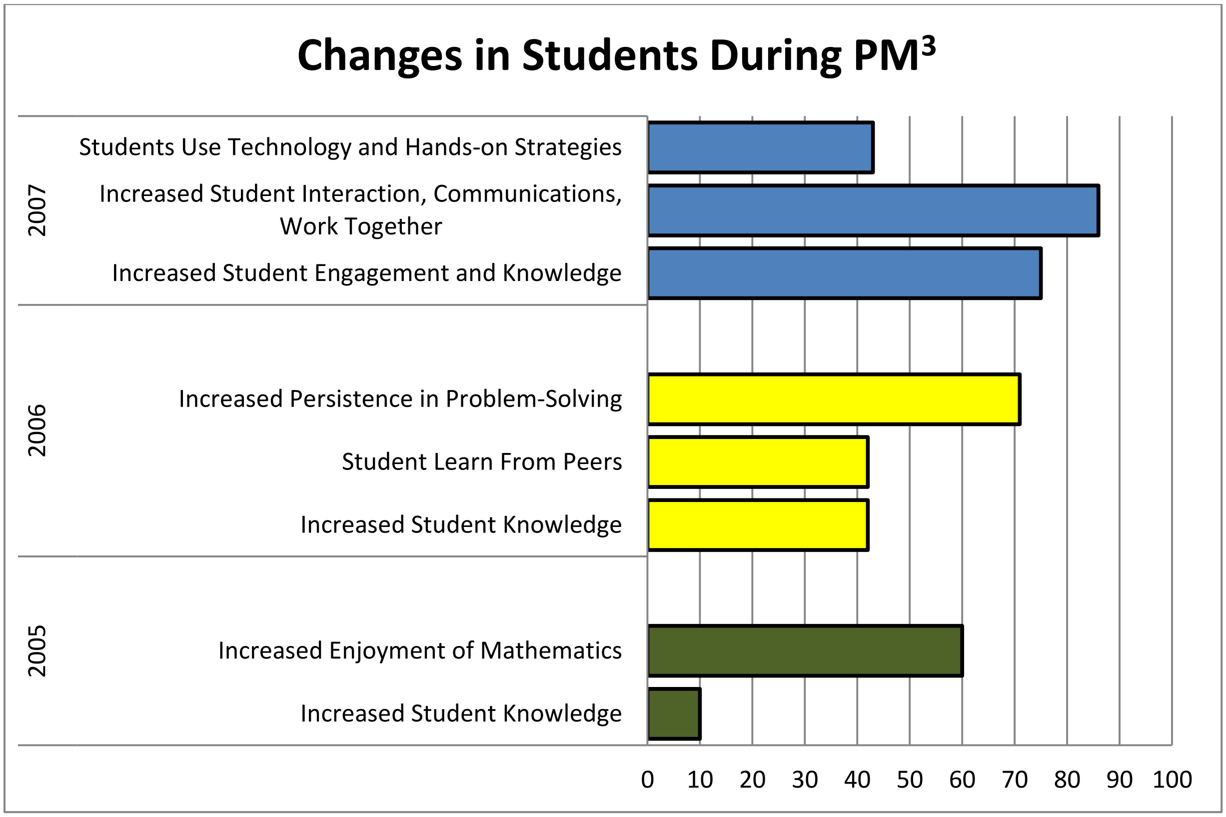
Year	Districts of comparison	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
2005	Treatment 1 & Treatment 2	.167	.006	.001	.000	.097
	Treatment 2 & Comparison	.000	.000	.000	.000	.000
	Treatment 1 & Comparison	.006	.000	.000	.000	.000
2006	Treatment 1 & Treatment 2	.001	.271	.000	.000	.011
	Treatment 2 & Comparison	.000	.000	.000	.000	.000
	Treatment 1 & Comparison	.008	.002	.144	.990	.000
2007	Treatment 1 & Treatment 2	1.000	.008	.000	.000	.999
	Treatment 2 & Comparison	.029	.000	.000	.000	.269
	Treatment 1 & Comparison	.033	.571	.970	.183	.172
2008	Treatment 1 & Treatment 2	.212	.105	.000	.002	.000
	Treatment 2 & Comparison	.026	.000	.001	.000	.000
	Treatment 1 & Comparison	.602	.015	.372	.024	.128
2009	Treatment 1 & Treatment 2	.286	.900	.000	.003	.009
	Treatment 2 & Comparison	.000	.478	.000	.000	.000
	Treatment 1 & Comparison	.000	.588	.689	.030	.003
2010	Treatment 1 & Treatment 2	.778	.009	.000	.654	.132
	Treatment 2 & Comparison	.029	.000	.000	.001	.000
	Treatment 1 & Comparison	.000	.387	.873	.000	.006

The cells in Table 21 that are shaded have values greater than .05, our criterion level for declaring sets to be significantly different. Therefore, for the comparison values in shaded boxes, these sets of student scale scores on the MEAP have not been found to be significantly different. Some of the cells are shaded in gray while others are shaded in green; the gray boxes are pairings between the two treatment districts and the green shading denotes the involvement of the comparison district with one of the treatment districts. Thus, the green shaded cells show where treatment districts' student scale scores on the state-wide MEAP test had accelerated and were no longer significantly different than those of the students in the comparison district!!

Looking back at Table 21 in the 2005 section, the two rows of zeroes jump out where the means of the MEAP student scale scores of each treatment district were analyzed next to the comparison group; at each grade level, the scores of treatment and comparison students were all found significantly different. As can be seen in figures 13 through 17, the difference was in favor of the comparison district with the highest scores, followed by treatment district 1, and the lowest scores generally from treatment district 2. In the years 2006 through 2010, all three of the districts made gains, but the treatment district students made greater gains than the comparison students so that some of their scores were no longer significantly different. These improvements in mathematics performance were seen in the number of cells shaded green in the following years: 2006, 1 cell; 2007, 5cells; 2008, 3 cells; 2009, 3 cells; and 2010, 2 cells.

In addition to these quantitative findings suggesting the students in the treatment districts are significantly improving their mathematics achievement in many instances relative to the students in the comparison district, the treatment teachers also perceive their students' increased mathematics achievement. When a focus group question asked teacher participants about what impact or changes they have seen in their students as a result of their participation in the institute, the responses changed in tone through the years. Figure 18 represents the major themes voiced by focus group participants through the years during the spring annual focus group session.

Figure 18. Teachers' focus group responses about changes in students.



After six months participation in PM³, 60 percent of the teachers reported their students were showing an increased enjoyment of mathematics and 10 percent believed student knowledge was increasing. While it is nice to have students enjoying mathematics, and the enjoyment may be a precursor to increased learning, it is a slightly shallow response as compared to those given by teachers the following year. In 2006, one of the more exciting themes expressed by over forty percent of the teachers was that students were learning from their peers. Not only does this mean students are taking more responsibility for learning and are willing to work together, but it also implies deeper learning. How many times have you heard, "I didn't really understand it until I taught it?" Another promising theme is a response by over 70 percent of the teachers about their students having increased persistence in problem solving. Also, a greater proportion of teachers reported increases in student knowledge. However, the following year over 70 percent of teachers noted increased student engagement and knowledge and over 80 percent referred to student to student interaction and communication as making a positive impact in their classrooms. It is difficult to know the amount of intersection between teachers' responses about student communication and interaction and the teachers' reports about students using technology and hands-on strategies. Each treatment participant received a document camera and projector for their classroom which enabled any student or group of students to come forward and show their work to the class. This equipment facilitates student discussions about their work and allows students to see a variety of ways to solve a problem.

As the analyses of improvements in mathematics achievement of students from schools where their teachers were involved in the PM^3 professional development treatment program comes to a close, this study has presented both quantitative and qualitative data suggesting student mathematics achievement improved during and after their teachers' participation in PM^3. While this researcher might have hoped for more and greater increases as compared to the comparison district, the power of really high or low scale scores on the MEAP test has been learned. While T2 scored 91 percent proficiency on the 2010 grade 4 MEAP that tied the state average and was higher than either T1 or the comparison district, the T2 scale score mean was still significantly lower than either of the other districts—probably because of its lower proportion of level 1, advanced students. For low-scoring schools to rise above the achievement gap, it is not necessarily enough for students to achieve proficiency, it is also about having higher scoring students in your district. It is easier to understand why the process is so slow and difficult—especially for students in challenged districts.

Chapter 5

Conclusions

The Importance of This Study

This study was based on an issue of crucial importance for our nation and of personal significance for students attending low-performing schools in high-poverty communities—equitable opportunities for high-quality, mathematics instruction and for improving mathematics proficiency. Increasingly, new technologies and innovations increase the mathematical, technological skills required of our citizen work force and of America's entrepreneurs. Other than in the relatively low-paying service sector industries, such as fast food, many of the repetitive work jobs that used to support middle class families are going to robots or to workers in developing countries. Therefore, if America intends to be a country of equal opportunity for all, and especially of equal educational opportunity for all, educational policies must shift from suggestions that racing and competing are solutions for the low performance of schools in high poverty areas, toward considerations of research findings about interventions which have been proven successful in high poverty settings and can be applied to achieve equity. Certainly university preparation of pre-service teachers is one place for working on improving mathematics instruction, but unless our nation is willing to simply give up on all of the students who are or will be being taught by our current in-service teachers, effective professional development is another critical juncture for improving the teaching and learning of mathematics.

Returning to Liping Ma's (Ma, 1999) explanation of factors which result in Chinese teachers having a more profound understanding of fundamental mathematics (PUFM) than American teachers, one component is future teachers' mathematics experience during their own schooling. Ma cites the fact that Chinese children live, go to school, and learn mathematics in an approach in which mathematical reasoning and understanding is stressed. Today's Chinese teachers are teaching mathematics the way they learned mathematics: as a cohesive system of pattern and order which can be used to model real-world situations. Additionally, Chinese students are taught by teachers who collaborate with their peers about the big ideas of the mathematics they are teaching and the appropriate sequencing of tasks for students to build their own mathematical reasoning skills. As new teachers come to a school, they become part of these collaborative, peer conversations.

American schooling is much different than Chinese schooling and has its own advantages; this study is certainly not advocating wholesale US adoption of a schooling system that works well in another culture. However, one US belief enables traditional mathematics instruction to remain unquestioned by many: in our culture there is an acceptance that mathematics is something that some individuals cannot do. The blame for mathematics failure is attributed to the person's problem— not to the instructional system. This same belief is not held about reading in the US. If someone cannot read, the system has failed that person and the wrong can be righted! Change is difficult and such widely held beliefs about what mathematics is and how it should be taught makes reform of mathematics instruction problematic— even though adults resisting the change may be the same ones saying they had problems understanding math.

Reform in mathematics instruction is an educational policy that has been strongly debated by two major factions for many years and has been called "the math wars." One group defended the traditional, theoretical, mathematics texts as enabling top students who could make mathematical connections and understand mathematics become the educational elite who would keep America competitive; the other group believed America and Americans would be stronger if more students were taught mathematics as a tool they could understand to help make sense of the world. This second group believed the traditional texts and approach to teaching mathematics left too large a percentage of Americans who could not contribute to their fullest potential due to a critical lack of mathematical understanding and reasoning. They promoted reformed—rather than traditional—texts that carefully designed and sequenced mathematical tasks intended to help students construct and communicate their own understanding of mathematics. This approach is more like Chinese instruction in which mathematical skills are connected, build upon each other, and result in the beginnings of PUFM.

These debates are probably moot as a result of the recent release of the Common Core State Standards (CCSS) intended to create national standards for reading, writing, literacy, and mathematics. In mathematics, the Standards for Mathematical Practice represent the means through which the Standards for Mathematical Content at each grade level are to be taught and learned. The CCSS attempts to sequence mathematics instruction so as to allow students to understand and see the connections between past, present, and future mathematical study. The CCSS Standards for Mathematical Practice for students stress higher order thinking such as perseverance in real-world problem solving, making connections in mathematical reasoning, constructing and communicating a valid mathematical argument in written as well as verbal form, and critiquing the reasoning of others. This is a very different curriculum and instructional approach—one more closely aligned with reform pedagogy and instruction. Traditional mathematics instruction usually involved teachers lecturing to students to convey a section in the mathematics textbook which was followed by students practicing numerous problems involving the solution method in that section. The real-world problems or "story problems" might just be assigned to the more advanced students or classes. With most traditional mathematics texts contracting with different authors to write chapters about their assigned topic, teachers and students using these texts are less likely to note connections between mathematical topics. Since the CCSS have been adopted by the vast majority of U.S. states and territories, the critical mass for reforming mathematics instruction has arrived.

Even as the CCSS continues the process of being adopted and implemented across the country, we might be well-advised to look again at the Chinese mathematics' experience to consider how long it might take until the American teaching force acquires the necessary PUFM to help students understand the mathematics they are studying. Most current teachers learned mathematics with traditional textbooks from teachers who focused more on rote memorization, procedures, and formulas than on deep conceptual understanding of the mathematics they were doing and why it worked. As a teacher in a low-performing, high-poverty community, how do you help students gain a deeper understanding of important mathematical concepts when you have not had an opportunity to develop your own understanding? In high-poverty communities, it is likely that fewer parents and community members have the time and/or mathematical skill set to help with students' mathematical understanding than in wealthier communities. The mathematical hope for students in low-performing, high-poverty districts is that their teachers will be afforded the opportunity to participate in an effective, mathematics, professional development program such as, "Project: Making Mathematics Matter."

The Findings of This Study

The PM³ project was designed as a quasi-experimental, matched-comparison group research project to evaluate the effectiveness of an intensive, mathematics, professional development model in very low-scoring, high-poverty districts. The PM³ principal investigators chose to work with districts most in need, despite the challenges the study would face as a result of the choice. Could a professional development project survive and succeed in such challenged environments?

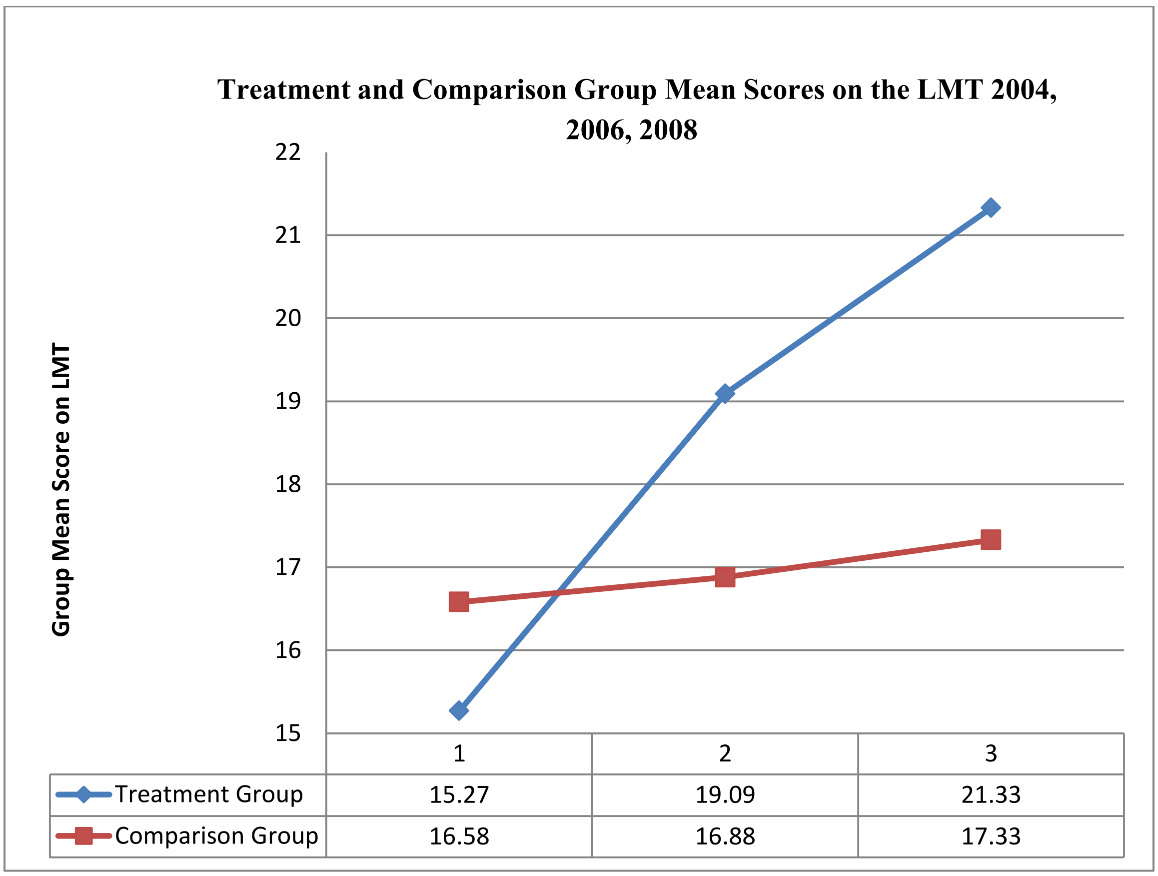
This study and analysis of the data collected during the 2004 through 2008 duration of PM³ was planned to answer these three research questions:

1. Are there significant differences in mathematics content knowledge for teaching between PM³ treatment group teachers and comparison group teachers as measured by pre- and post-LMT assessments over the period of the PM³ project?
2. Are there significant changes in classroom instruction and practice of treatment teachers as measured by the SAMPI classroom observation protocol during their participation in the PM³ professional development project?
3. Are there significant differences in grades 4 through 8 MEAP mathematics performances between students in treatment districts and students in the comparison district following participation in the PM³ project?

For the first research question, data from the LMT assessment of pedagogical, mathematical, content knowledge for teaching were compared for treatment teachers and their matched comparison teachers and the gains for the treatment group teachers were significantly higher than those of comparison teachers ($p < .001$). A mixed within- and between-subjects repeated measures ANOVA was used to assess differences. Since the teachers were matched and the comparison teachers began the study with higher—but not significantly higher—scores on the 2004 administration of "Learning Mathematics for Teaching," the line plots of the scores crossed and the effects of the interaction between the group to which a teacher belonged (treatment or comparison) and the year the test was given were significant ($p < .001$) and had a large effect size with this interaction accounting for 30.6 percent of the variance among scores.

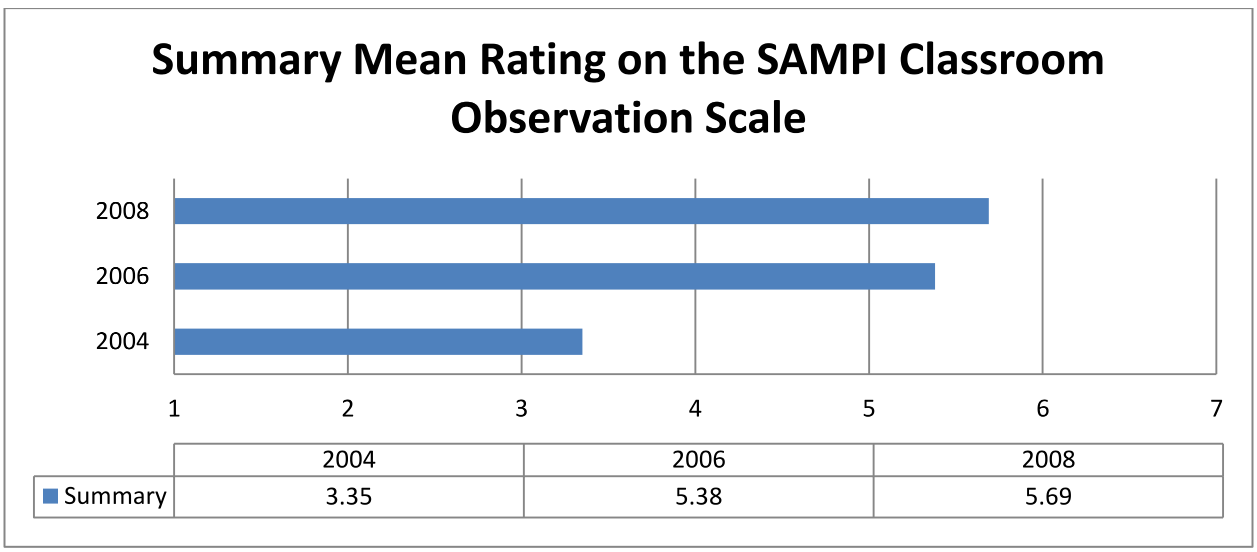
Additional qualitative analyses were performed on participants' responses to focus group questions and the depth and complexity of their responses about their mathematical knowledge supports the quantitative finding of treatment teachers demonstrating significantly higher increases in their pedagogical, mathematics, content knowledge for teaching as measured on the LMT during the time of their participation in PM³. Figure 9 shows this comparison visually with the blue plot representing the treatment teachers' means and the red line plot following the comparison group's means.

Figure 19. Means of "Learning Mathematics for Teaching" scores.



The second research question sought to discover whether the quality of mathematics instruction increased in treatment teachers' classrooms while they were engaged in the PM³ professional development program. The research study findings suggested a significant increase in the quality of instruction. A one-way repeated measures ANOVA was used to compare each teacher's ratings of classroom quality of mathematics instruction on the SAMPI Classroom Observation Protocol at the beginning, middle, and end of the project in 2008. Once the ANOVA indicated a significant difference, pairwise comparisons of means were performed to find where the means were significantly different. The pairwise statistical testing showed the quality of treatment teachers' mathematics instructional practices increased significantly ($p < .001$) between the start of PM³ (2004) and the middle of the project (2006) by about two points on a seven point scale. The pairwise comparisons between the beginning and the end were also significant ($p < .001$), with the 2008 ratings somewhat higher than those of the midpoint of the project. However, the increase between the 2006 ratings and 2008 ratings was non-significant; it appears teachers' ratings were topping out on the seven-point scale of the SAMPI instrument.

Figure 20. Means of instructional quality on SAMPI classroom observation protocol.



However, the qualitative analyses of teachers' responses in focus groups indicated their knowledge and sophistication about mathematics and mathematics instruction continued growing as their time of participation in PM³ increased. They progressively connected their increases in mathematical content knowledge with improved instruction in their classrooms.

Now that the statistical testing has suggested that the answers to the first two research questions are affirmative, with treatment teachers experiencing significant increases in their pedagogical, content knowledge for teaching mathematics and in the quality of their classroom instruction during the time of participation in PM³, the remaining question was whether their students' performances on the state-wide mathematics were significantly different than those of students in the comparison teachers' district.

The analyses of this third research question began with findings that all three of the districts had significant differences in student MEAP achievement in 2005, approximately one year after the start of the PM³ project. The comparison district's student MEAP scores were low, but not as low as the students in the treatment districts. For mathematics achievement demonstrated on the state-wide MEAP test among these three districts, the comparison district's students' scores were highest, followed by treatment district T1, and then by treatment district T2. In fifth grade, the state mean for proficiency was 74 percent, the comparison group showed 62 percent proficiency, with treatment district 1 having 46 percent of its students rated proficient, and treatment district 2 reporting 31 percent proficiency. A two-way ANOVA was conducted with the approximately 4000 student scale scores for each grade; the two independent variables were the year (1-6) and the district (1-3) and the dependent variable was the student scale score. The interaction effect of these two independent variables and the dependent variable in grades 4, 5, 7, and 8 was found to have less than a 5 percent probability of happening by chance—thus meeting the necessary $p < .05$ for significance. Sixth grade did not quite meet this criterion level with $p = .068$. Once a significant interaction effect was found between year after the start of PM^3 and district on students' MEAP scores (with the exception of grade 6), one-way ANOVAs were performed on the means of the three districts for every year and grade. For each ANOVA only one year of student scale scores were used, and therefore, the independent variable for this testing was the student's district and the dependent variable was the student's MEAP scale score.

The results of these one-way ANOVAs showed a number of comparisons in which one or the other of the treatment districts' sets of student scale scores was no longer significantly different than the comparison set of student scale scores. While all 3 districts showed improved means, the treatment districts demonstrated an accelerated rate of gain in some instances that resulted in their scores not being statistically different from the comparison district. Figure 21 shows a graph of the fifth grade student scale scores means for all three districts. In 2005, all three means are significantly different and both treatment means are below 500— the minimum rating for proficiency. All three sets of scores generally increase with some occasional setbacks from which districts generally recover. Then in 2009, there are no significant differences between the student MEAP achievement scores between any of the districts. T1 and C scores remain close to each other, but district T2 experiences a significant decline.

Figure 21. Fifth grade MEAP student scale score means comparisons, T1, T2, C

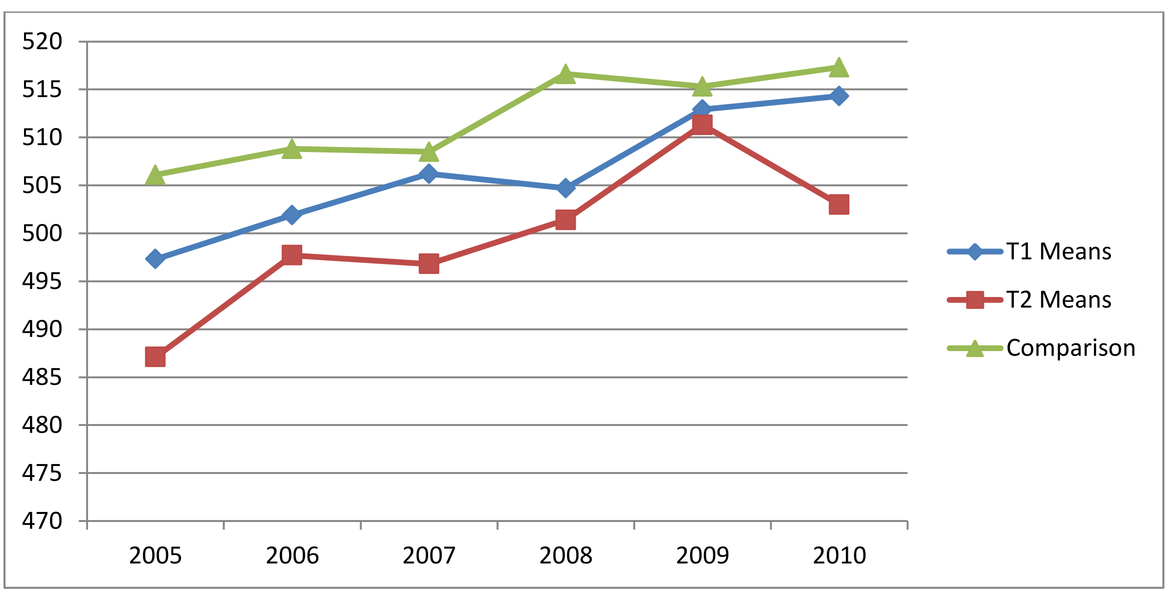


Figure 21 combines the hopes and fears about the future mathematics success of students in these treatment districts and others in similar settings. With support similar to that of PM³, teachers can increase their pedagogical, content knowledge for teaching mathematics, improve the quality of their classroom instruction, and students can accelerate their mathematics achievement to begin to have mathematical knowledge more similar to that of students in less challenged environments. What happens when the support leaves? Is the down-turn in scores of students in T2 a harbinger of the district challenges overwhelming the PM³ support from 2004 to 2008….or….is it just one of the temporary setbacks these districts have experienced and the 2011 scores will rise again?

Implications of this Study

These research findings add to the paucity of literature about successful mathematics, professional development in very low-performing, high poverty communities. Through the analyses of data collected during PM³, this research study found an intensive, sustained mathematics professional development program in very low-performing, high-poverty communities can show positive, significant impact on teacher content knowledge, mathematics instructional quality, and student mathematics achievement. However, programs such as "Project: Making Mathematics Matter" are not inexpensive and not a quick fix. Currently, the money being spent to help low-performing, high-poverty schools and districts seems to suggest competition and choice are the magic wand for increasing student achievement. How does that begin to help in the important work of making sure every student in every classroom has a good teacher? (Darling-Hammond, 1996; Sanders & Horn, 1998; Darling-Hammond & Barnett, 2006).

It's about the teacher, stupid! Just as President Clinton had to keep reminding himself that it's about the economy, our legislators and policy makers need to focus on the most important interactions that occur in schools where students achieve—those between the students and their teachers. If teachers do not have a deep understanding of mathematics, it is not a matter of them trying harder or teaching in another school. Just as the PM³ PD project found a large percentage of teachers in treatment schools at the beginning of the study without good preparation in mathematics, other research studies have similarly found less prepared teachers in challenged, low-performing, high-poverty districts. The current push toward charters creates an ever changing faculty status in schools. Even as the teachers, instruction, and student scores steadily improved in these treatment districts, they found themselves losing students to charter schools where students would likely have less prepared mathematics teachers. So many excellent, caring mathematics teachers were laid off as district enrollments shrank. Are the policy makers really trying to marginalize students in such challenged environments even more?

Suggestions for Future Study

Further study might include continuing analyses in 2011 and beyond of the student MEAP scores for these treatment districts to determine if their MEAP mathematics scores continue a general pattern of increase or not. It would also be interesting to try to find further information about who taught the mathematics each year to the students whose scores were studied. When the new state test is administered in the last 12 weeks of the year, studies might be tempted to try connecting the learning of a student with their teacher for that year. However, this study suggests there are many intervening variables in challenged environments with great mobility of students and teachers. Other studies might include following teachers who were trained during PM³ or similar intensive mathematics PD programs in challenged environments to find the current situation of these teachers who received so much training only to be laid off due to shrinking enrollments in the urban school chaos. This might also lead to a study about what information parents have and how they get their data when deciding where to send their child to school. Is this different in challenged environments? As difficult as it is to perform research in such challenged environments, it is among the most important places to study if we hope to provide equitable educational opportunities to children who find themselves at the mercy of the communities in which they live. This researcher would like to see more long-term commitments by policy makers to provide funding to replicate intensive, mathematics professional development programs, such as PM³, to deepen teachers' pedagogical, content knowledge for teaching mathematics and to improve the quality of classroom mathematics instruction to see if students in these communities have experience better mathematical outcomes.

This study raises questions for future research in three major areas: 1. How do intermediate school districts (ISDs) operationalize professional development in low-performing, high-poverty districts? Is a whole-district, content-specific model necessary in settings with high mobility of staff and students? 2. How do we measure implementation and what students are learning? Is there a good assessment that will include both the CCSS Practice and Content Standards that can be used for further research studies in low-performing, high-poverty settings? 3. What about the ways we analyze the collected data and the amount of variance that can be attributed to each source in such challenged settings? Is hierarchical linear modeling with nesting possible in settings with such high mobility rates? Must future studies use the district as the unit of measurement requiring large projects to have sufficient sample sizes or are there alternative units of analysis that might be used?

This study shows results of a sustained, mathematics, content-focused professional development program conducted in highly challenged environments and the strongest suggestion for future study is to conduct more well-designed studies in these settings to further inform educational policy makers about successful interventions for students in these communities.

BIBLIOGRAPHY

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, 69 (2), 334-345.

Addonizio, M. F. (2004). Toward a new adequacy in public school finance: Analytical and political issues. *Educational Considerations*, 32(1), 55-61.

Armor, D. J. (2006). Lessons learned from school desegregation. In P. E. Peterson (Ed.) *Generational change: Closing the test score gap*, (pp. 115-142). NY.: Rowman & Littlefield Publishers.

Ball, D. L. (1991). Teaching mathematics for understanding: What do teachers need to know about the subject matter? In M. Kennedy (Ed.), *Teaching academic subjects to diverse learners*, (pp. 63-83). NY.: Teachers College Press.

Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83-104). Westport, CT.: Ablex.

Ball, D. L., & Bass, H. (2003). Toward a practice-based theory of mathematical knowledge for teaching. In B. Davis & E. Simmt (Eds.), Proceedings of the 2002 Annual Meeting of the Canadian Mathematics Education Study Group, (pp. 3-14). Edmonton, AB.: CMESG/GCEDM.

Banilower, E. R., Heck, D. J., &Weiss, I. R. (2007). Can professional development make the vision of the Standards a reality? The impact of the National Science Foundation's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching*, 44(3), 375-395.

Blank, R., de las Alas, N., & Smith, C. (2008). Does teacher professional development have effects on teaching and learning? Evaluation findings from programs for mathematics and science teachers in 14 states. Council of Chief State School Officers, Washington, DC.: Publications The Council of Chief State School Officers.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3-15.

Borko, H., Eisenhart, M., Brown, C., Underhill, R., Jones, D., Agard, P. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, 23, 194-222.

Borman, G. D., & Rachuba, L. T. (1999). Qualifications and professional growth opportunities of teachers in high- and low-poverty elementary schools. *The Journal of Negro Education*, 68(3) 366-381.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL.: Rand McNally.

Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. *Phi Delta Kappan*, 78(3), 193-200.

Darling-Hammond, L. (1998). Unequal opportunity: Race and education. *The Brookings Review*, 16(2), 28-32.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1). Retrieved September 22, 2009 from the Educational Policy Analysis Archives website: http://epaa.asu.edu/epaa/v8n1/

Darling-Hammond, L., (2007). Race, inequality and educational accountability: The irony of- "No Child Left Behind". *Race, Ethnicity and Education*, 10(3), 245-260.

Darling-Hammond, L. & Barnett, B. (2006). Highly qualified teachers for all. *Educational Leadership*, 64(3), 14-20.

Darling-Hammond, L., Williamson, J.A., & Hyler, M. H. (2007). Securing the right to learn: The quest for an empowering curriculum for African American citizens. *The Journal of Negro Education*, 68(3), 366-381.

Dorfler, W. (2003). Mathematics and mathematics education: Content and people, relation and difference. *Educational Studies in Mathematics*, 54, 147-170.

DuFour, R. & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Alexandria, VA.: Association for Supervision and Curriculum Development.

DuFour, R. (2004). What is a "Professional Learning Community"? *Educational Leadership*, 61(8), 6-11.

Eisenhart, M., Borko, H., Underhill, R., Brown, C., Jones, D. & Agard, P. (1993). Conceptual knowledge falls through the cracks: Complexities of learning to teach mathematics for understanding. *Journal for Research in Mathematics Education*. 24, 8-40.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, 7(9).

Fleischman, H. L., Hopstock P. J., Pelczar, M. P. & Shelley B. E. (2010). *Highlights From PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context.* (NCES 201004). U.S. Department of Education, National Center for Education Statistics. Washington, DC.: U.S. Government Printing Office.

Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap? *The High School Journal*, 91(1), 29-42.

Fryer, R. G. & Levitt, S. D. (2006). Losing ground at school. In P. E. Peterson (Ed.) *Generational change: Closing the test score gap*, (pp. 88-114). NY.: Rowman & Littlefield Publishers.

Gallimore, R., Ermeling, B., Saunders, W., & Goldenberg, C., (2009). Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams. The Elementary School Journal, 109(5), 537-553.

Garrett, J. L. (2008). STEM: The 21st century Sputnik. *Kappan Delta Pi Record*, 44(4), 152-153.

Gonzales, P., Guzman, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). *Highlights from the trends in International Mathematics and Science Study (TIMSS), 2003*. (NCES 2005-005). U.S. Department of Education, National Center for Education Statistics. Washington, DC.: U.S. Government Printing Office.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). *Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context*. (NCES 2009-001). Washington, DC.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Griffith, J. E. & Medrich, E. A. (1992). *International mathematics and science assessments: What have we learned?* National Center for Research and Education Statistics. U. S. Department of Education Research and Development Report NCES 92-011 455-G-09 ED 1.328/6: M 42.

Gusky, T. R. (2003). Analyzing lists of the characteristics of effective professional development to promote visionary leadership. *NASSP Bulletin*, 87(637), 4-20.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L (2006). *Multivariate data analysis*, 6th edition. Upper Saddle River, N.J.: Pearson/ Prentice Hall.

Hanushek, E. A. (2006). Alternative school policies and the benefits of general cognitive. *Economics of Education Review*. 5, 447-462.

Harris, D. & Ray, L. (2003). *No School Left Behind? The Distribution of Teacher Quality in Michigan's Public Schools*, East Lansing: Education Policy Center at Michigan State University.

Haskins, R. (2006). Putting education into preschools. In P. E. Peterson (Ed.), *Generational change: Closing the test score gap* (pp. 47-87). New York: Rowman & Littlefield Publishers, Inc.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study* (NCES 2003-013 Revised). Washington, DC.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, 35(5), 330-351.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, 105(1), 11.

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *Excellence in the Classroom*, 17(1), 129-155.

Kilday, C., & Kinzie, M.. (2009). An Analysis of Instruments that Measure the Quality of Mathematics Teaching in Early Childhood. *Early Childhood Education Journal*, 36(4), 365-372.

Laczko-Kerr, I & Berliner, D. C. (2002). The effectiveness of "Teach for America" and other under-certified teachers on student achievement: A case of harmful public policy. *Education Policy Analysis Archives*, 10(37).

Lankford, H., Loeb, S. & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Education Policy Analysis*, 24(1), 37-62.

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics,* 2nd edition. Thousand Oaks, CA.:Corwin Press.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics*, 99(5), 258-271.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teacher's understanding of fundamental mathematics in China and the United States*. Mahwah, NJ.:Lawrence Erlbaum Associates.

Michigan Department of Education (2007). *Frequently asked question: Michigan merit curriculum high school graduation requirement*. Lansing, MI.: Michigan Department of Education. Available http://www.michigan.gov/documents/mde

Miles, K. H., Odden, A., Fermanich, M., & Archibald, S. (2004). *Excerpts from inside the black box-School district spending on professional development in education: Lessons from five urban districts*. Document Number ED485651: The Finance Report.

Miles, K. H., Odden, A., Fermanich, M., & Archibald, S. (2005). Inside the black box of school district spending on professional development: Lessons from five urban districts. *Journal of Education Finance,* 30, 1-26.

Mittag, K. C., Shoho, A., & Lien, V. (2004). Describing and evaluating an exemplary mathematics elementary staff development project. *School Science and Mathematics*, 104(4), 145-154.

National Center for Education Statistics (1992). International mathematics and science assessment: What have we learned? Washington, DC.: U.S. Department of Education. Available http://nces.ed.gov/pubs92/92011.pdf

National Center for Education Statistics (1999). *Highlights from TIMSS: The third international mathematics and science study.* Washington, DC.: U.S. Department of Education. Available http://nces.ed.gov/timss

Neal, D. (2006). How families and schools shape the achievement gap In P. E. Peterson (Ed.), *Generational change: Closing the test score gap* (pp. 26-46). New York: Rowman & Littlefield Publishers, Inc.

No Child Left Behind Act of 2001, H.R. 1, 110th Cong. (2002)

Odden, A. (2003). Equity and adequacy in school finance today. *Phi Delta Kappan*, 85(2), 120-125.

Odden, A., Archibald, S., Fermanich, M., & Gallagher, H. A. (2002). A cost framework for professional development. *Journal of Education Finance*, 28, 51-74.

Payne, K. J., & Biddle, B. J. (1999). Poor school funding, child poverty, and mathematics achievement. *Educational Researcher*, 28(6), 4-13.

Peterson, P. E. (2006). Toward the elimination of race differences in educational achievement. In P.E. Peterson (Ed.), *Generational change: Closing the test score gap*, (pp. 1-25). NY.: Rowman & Littlefield Publishers, Inc.

Saunders, W., Goldenberg, C., & Gallimore, R. (2009). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, quasi-experimental study of Title 1 schools. American Educational Research Journal, 46(4), 1006-1033.

Schilling, S. G. & Hill, H. C. (2007). Assessing measures of mathematical knowledge for teaching: A validity argument approach. *Measurement: Interdisciplinary Research and Perspectives*, 5(2-3), 70-80.

Shulman, L. S. (1986). Those who understand: Knowledge and growth in teaching. *Educational Researcher*, 15(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of a new reform. *Harvard Educational Reviewer*, 57(1), 1-22.

Slavin R. E., Lake C. & Groff C. (2009). Effective programs in middle and high school mathematics: A best evidence synthesis. *Review of Educational Research*, 79 (2), 839 –911.

Stein, M., Smith M., & Silver, E. (1999). The development of professional developers: Learning to assist teachers in new settings in new ways. *Harvard Educational Review*, 69(3), 237-269.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Dallas, TX.: National Staff Development Council.

Wei, R. C., Darling-Hammond, L., and Adamson, F.(2010). *Professional development in the United States: Trends and challenges*. Dallas, TX.: National Staff Development Council.

West, L. & Curcio, F. (2004). Collaboration sites: Teacher-centered professional development in mathematics. Teaching Children Mathematics, ??

West, L., & Staub, F. C. (2003). *Content-focused coaching: Transforming mathematics lessons*. Portsmouth, NJ.: Heinemann.

Wiliam, D. (2007). Changing classroom practice. *Educational Leadership*, 65(4), 36-42.

Wilson, S. M., Shulman, L. E., & Richert, A. E. (1987). 150 different ways of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers' thinking,* (pp. 104-124). Sussex: Holt, Rinehart and Winston.

Yoon, K. S., Duncan, T., Lee, S. W., Scarloss, B. & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement. Issues and answers. REL 2007-, No. 033. Retrieved February 4, 2009 from the Institute of Education Sciences, U. S. Department of Education website: http://ies.ed.gov.ncee/edlabs/regions/southwest

ABSTRACT

MAKING MATHEMATICS MATTER: PROFESSIONAL DEVELOPMENT IMPROVING OUTCOMES IN HIGH-POVERTY ENVIRONMENTS

by

CAROLYN A. SIEBERS

May 2012

Advisor: Dr. Michael Addonizio

Major: Educational Leadership and Policy Studies

Degree: Doctor of Philosophy

This research study is a mixed-methods quantitative and qualitative study that analyzed data from an intensive, long-term professional development project named, "Project : Making Mathematics Matter. (PM^3)" carried out in two low-performing, high-poverty districts. Two of the districts served as treatment districts and ALL teachers (general education, special education, ELL) who taught mathematics to students in grades 4-8 participated in a four-year PM^3 mathematical intervention.

Three elements identified in the research as effective components in mathematics professional development served as foundational for this project. They were: teachers attended monthly institutes focused on deepening the pedagogical, content knowledge for teaching mathematics; teachers received on-site coaching from an experienced mathematics educator to support implementation of new instructional strategies; and teachers participated in monthly, collegial, grade band meetings facilitated by their coach.

This study's analyses focused around three areas: 1. increased pedagogical knowledge for teaching mathematics as measured by the "Learning Mathematics for Teaching" (LMT)

assessment (Hill, Schilling, & Ball, 2004); 2. improved mathematics instructional quality as measured by the SAMPI Classroom Observational Protocol; 3. increased mathematics performance of students as measured on the state-wide MEAP test.

Research findings were: 1. LMT data collected at the beginning, middle, and end of PM^3 showed treatment teachers significantly increased their pedagogical, content knowledge for teaching mathematics over their matched-comparison teachers; 2. Analyses of SAMPI data demonstrated teachers significantly improved the quality of their classroom instruction; 3. Analyzing sets of MEAP student scale scores by grade and year, there was a significant interaction between district of the student and the year, during and following treatment teachers' participation in PM^3. Where in 2005, all sets of treatment scale scores were significantly different and lower than comparison scores, as the years of treatment teacher participation in PM^3 increased many sets of compared sets of MEAP scores were no longer significantly different.

This study suggests a connection between the increased content knowledge of teachers, the improved quality of classroom instruction, and the increased mathematics achievement of their students in high-poverty, low-performing districts.

AUTOBIOGRAPHICAL STATEMENT

Thirty years ago, I graduated from the University of Michigan and began my career as an educator. I was strongly aware of my passion for making a difference for students through my work in schools, but certainly did not anticipate all of the paths I have since explored in pursuit of my goal. From a young age, I had always pictured myself as a teacher and during my student teaching, I knew I had chosen the right career for me – I loved teaching!

My first position was mathematics teacher and coach in Byron, a small rural Michigan community. After four years in Byron, I moved to Ann Arbor and taught mathematics and computers at Community High School, an alternative school of choice in the Ann Arbor Public Schools. Community high was created in the early 1970's during student unrest around the Vietnam War to attract activist students and remove them from other high school locations in the district. A major component of the new school was a structure called forum, where twenty students met regularly with a teacher throughout their four-year high school career for school business including course planning and for social activities. I learned a great deal from these two very different school cultures and became convinced that an integral piece of high student performance is a strong connection with an adult at school.

Although I loved teaching, my desire to learn more about successful schools and how to create them led me to Wayne State University where I earned an Educational Specialist degree in Educational Leadership. Although I continued teaching mathematics for five more years after completing my degree, I began considering leaving my teaching to make a larger scale impact as an administrator. My first administrative post was as an assistant principal at Pioneer, a large high school in the Ann Arbor district. After four years in this position, I was named principal of Thurston High School in South Redford. In each position, I worked with other staff members to implement policies to increase student success – especially for at-risk students.

Four years ago, I moved to Wayne Regional Educational Services Agency (RESA) as a mathematics consultant to attempt to further increase my sphere of influence in my mission to increase student success. The project that has been most rewarding in this respect is a federally-funded Math/Science partnership grant RESA was awarded to work with mathematics teachers in two, low performing, high-poverty districts. After two years of professional learning work with these districts, our preliminary results look extremely promising. I believe this project may further the body of knowledge about successful mathematics interventions for teachers and students in challenged communities.

I hope to be able to complete by doctoral work at Wayne State University including a dissertation on this project's outcome so others might see the research and work to replicate it. This would truly be beyond the scope of my youthful dreams about making a difference as an educator.

The statement above is the personal statement I wrote in my application for admittance into the Wayne State University doctoral program in Leadership and Policy Studies. It is now five years later, this statement still traces my evolution in educational leadership, and I am approaching the pivotal point I sought. I am now submitting my dissertation which I hope will add to the literature and research in challenged environments by linking effective mathematics professional development for teachers with improved mathematical outcomes for students.