Investing in Teacher Capacity

Results from the Impact Evaluation of the New Mexico Math-Science Partnership

Youth Development Evaluation Alliance

YDE/A

YDE/A

Youth Development Evaluation Alliance

Investing in Teacher Capacity

Results from the Impact Evaluation of the New Mexico Math-Science Partnership

July 2009

Prepared by

William P. Moore
Youth Development Evaluation Alliance

Holly A. McGregor
Youth Development Evaluation Alliance

Sharon L. Newbill
Folkstone: Evaluation Anthropology

This report was prepared for Folkstone: Evaluation Anthropology under contract number 26-003 by Youth Development Evaluation Alliance. The content of the publication does not necessarily reflect the views or policies of the New Mexico Public Education Department nor does the mention of trade names, commercial products, or organizations imply endorsement by the New Mexico Public Education Department.

Moore, W. P., McGregor, H. A.,& Newbill, S. L. (2009). Investing in teacher capacity: Results from the impact evaluation of the New Mexico Math-Science Partnership. Kansas City, MO: Youth Development Evaluation Alliance. Retrieved from: http://www.youtheval.org/publications.

This report is available on the Youth Development Evaluation Alliance website at http://www.youtheval.org/publications

Summary

Does the provision of content-specific mathematics professional development lead to changes in classroom practice that affect the mathematics achievement of students? At what levels of duration, intensity and focus must in-service professional development be in order to begin influencing classroom practice so that changes in achievement can be detected? A number of studies have documented the relationship between the quality of instruction and the outcomes of students. Teacher training and experience have been found to be associated with student achievement and In-service teacher professional development appears to positively influence the achievement outcomes of elementary school students. Researchers have noted a distinct curvilinear relationship between teacher experience and student achievement (achievement increases as teacher experience increases for 2 to 5 years, with in-service training, then levels off (Ferguson, 1991; Darling-Hammond, 2000; Rockoff, 2003).

A recent rigorous review of studies of teacher professional development (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007) has found that teachers who participate in more than 14 hours of professional development have a significant and positive impact on student achievement. Yoon, et al., after reviewing 9 rigorous studies of the affect of teacher professional development on student achievement, concluded that teachers who receive substantial professional development saw improvements of 21 percentile points in student achievement (p. 1).

This report describes the results of an external impact evaluation of the New Mexico Math-Science Partnerships (MSP) program after two years of implementation. The results of analyses of New Mexico state mathematics assessment scores, teacher professional development logs, content assessments, and student and teacher background and demographic information revealed statistically significant but practically non-significant improvements in latent growth trajectories of middle school students' mathematics achievement after two years of exposure to teachers participating in the MSP program.

The findings suggest that 1) the amount, type or focus of professional development support received by MPS teachers was not sufficient to bring about meaningful changes in instructional practice that would have a substantial and policy relevant impact on student mathematics achievement; 2) through multiple advanced statistical methods a coherent and consistent message emerged – student achievement did not change dramatically over the period of time studied by the evaluators – even with the provision of professional development targeted at mathematics content-specific knowledge and pedagogy. While several key hypotheses could not be tested because of missing data, the two most important evaluation questions and hypotheses did get answered: Does increased professional development make a difference in

the mathematics outcomes of students? And, does exposure to MSP teachers provide an incremental advantage in mathematics performance beyond that of non-MSP participating teachers? In both cases the results from this study provide no compelling evidence that MSP has dramatically changed the mathematics achievement trajectories of students. *The estimated mean rate of change was .09 units per year, or about a tenth of a proficiency category a year. While this rate of change was statistically significant, it does not appear to be practically significant*. Whether one considers the most policy-relevant outcome: moving students into the proficient and above category or looking at movement of percentages of students into the next highest proficiency category, MSP does not appear to have meaningfully improved the odds that students will progress above and beyond that of students who were not exposed to MSP.

The evaluation team notes that the complex analyses utilized to discern affects of professional development on student outcomes required a commitment by data owners to fulfill a highly structured and extensive data request --- one that spanned many years, many teachers, many schools, and many students. A relevant observation of the evaluation team experience in this project is that data quality and the fulfillment of data requests by the New Mexico Public Education Department lacked the capacity and resources necessary to provide the extensive data sets required. Data files received by the evaluation team contained duplicate student records, missing data elements, and coding errors contributing to significant delays in initiating analyses. Entire years of student data were not provided and requests for data were not fulfilled for months. The original scope of this evaluation included three MSP partnerships --- only one was responsive to the calls from the external evaluator to participate. The results obtained in this evaluation would have been more robust, compelling, and meaningful had all three partnerships contributed to the evaluation, had the NM PED been consistently responsive and able to fulfill data requests, and actively supporting and strongly urging the participation of MSP partnerships in this important evaluation of the impact of professional development on student mathematics achievement.

Table of Contents

Introduction

The Institute for Education Sciences (IES) has noted that "current levels of mathematics and science achievement at the elementary and secondary levels suggest that the United States is neither preparing the general population with levels of mathematics and science knowledge necessary for the 21st century workplace, nor producing an adequate pipeline to meet national needs for domestic scientists and mathematicians" (U.S. Department of Education, 2009, p. 13). Studies of NAEP administrations in the 2000 and 2005 school years indicated that 20 percent of Grade 4 students and 31 percent of Grade 8 students scored below the "basic" level. In the 2000 NAEP, 35 percent of grade 12 students scored below the "basic" level. Even with modest gains in mathematics performance over the last decade large achievement gaps remain among minority groups and students from low-income backgrounds.

Education reformers and researchers have identified the lack of well-qualified teachers as a factor in the poor performance of students. Fetler (1999), Klecker (2002), Clotfelter et al. (2007), found that student achievement in mathematics was significantly related to teacher experience and preparation although effect sizes were small. This finding is important for high-poverty, ethnically diverse schools, which have higher percentages of new teachers and teachers without subject area certification than those schools serving students in affluent communities. Moore and Esselman (1994) found that a teacher's personal efficacy and indirectly her teaching efficacy were related to the achievement performance of students. The intersection of teaching and student achievement is the classroom experience of both teachers and students – how they interact to bring about improved learning is the direct focus of much professional development offered to teachers today. Clotfelter, Ladd and Vigdor (2007), in a study of the influence of teacher experience and training on student outcomes conclude that "a teacher's experience, test scores and regular licensure all have positive effects on student achievement, with larger effects for math than for reading. Taken together the various teacher credentials exhibit quite large effects on math achievement" (p. 1).

Very few scientifically rigorous studies have been conducted to evaluate the effectiveness of mathematics curricula and instructional practices for improving student learning and math achievement. While some educational reformers (Connell and Broom, 2004) have undertaken efforts to improve the quality of instruction by focusing on key practices designed to improve student engagement and the rigor of teaching and learning, these variables have not made their way into many studies of the impact of mathematics interventions on student outcomes.

A recent review of the evidence on how teacher professional development affects student achievement (Yoon, Duncan, Lee, Scarloss & Shapley, 2007) examined the most rigorous studies available in the research literature. More than 1,300 studies were identified by the researchers and only 9 met the rigor standards of IES's What Works Clearinghouse. The overall conclusion of the reviewers was that teachers who participate in a substantial amount of professional development (49 hours or more) have students who perform significantly higher on measures of achievement – on average, 21 percentile points higher. Teachers who had more than 14 hours of professional development had a statistically positive effect on student achievement. Studies that tested the effects of less professional development (5-14 hours) found no statistically meaningful impact on achievement.

Yoon et al. cite a study by Birman et al. (2007) who demonstrated that few teachers "receive intensive, sustained, and content-focused professional development in mathematics (p. 2)." Birman found on average that teachers received 8.3 hours of professional development on how to teach mathematics and 5.2 hours on content-specific knowledge in the previous 12 months (p.2).

The nine studies included in the Yoon et al, meta-analysis only focused on elementary school teachers and students. No studies were included that focused on middle or high school mathematics professional development. Two of the studies included a focus on mathematics exclusively. So, while the field now has some compelling information about the impact of professional development on student outcomes, the extent of the research literature that is of sufficient rigor about mathematics professional development remains sparse and even more so for secondary school teachers and students.

Analyses associated with an evaluation of the impact of professional development on student outcomes should be focused on changes in student mathematics achievement in classrooms taught by teachers who have received professional development in mathematics content knowledge and mathematics pedagogy compared to those teachers who have not received professional development. Because students belong to classes taught by teachers, the extent to which students vary from one another in mathematics achievement is partially dependent upon the classroom (or more precisely, the teacher) to which the student belongs and has belonged in the past. Thus, in constructing analyses for evaluations of the impact on students of teacher professional development, the analysis must address questions of 1) what skills and knowledge teachers gain from professional development (content and pedagogy); 2) how much, what type, and in what mix of content and pedagogy is professional development to be maximally effective in producing changes in classroom instruction that will lead to improved student outcomes; and 3) what happens to student outcomes when professional development experiences varies?

The New Mexico Mathematics and Science Partnership (MSP) program specifically supports projects that will increase the subject matter knowledge and teaching skills of mathematics teachers and provide opportunities for career-long professional development that improves teachers' mathematics knowledge and pedagogy. This report summarizes the results of an independent evaluation of the impact of professional development activities provided by one MSP partnership on the student achievement outcomes of students of participating teachers.

Description of the Program

The Mathematics and Science Partnership (MSP) program administered by the New Mexico Public Education Department is a federal formula grant. The state prioritized funding of MSP projects supporting content-rich professional development for middle school mathematics teachers in high need schools in the state. According to the call for proposals issued by NMPED,

> MSP projects are designed and implemented by partnerships that include K-12 administrators, faculty, teachers, and guidance counselors in participating K-12 schools, STEM faculty, and administrators in higher education organizations. Other partners are encouraged and may include businesses, nonprofit organizations, and teacher training departments of an institution of higher education. These partners and other stakeholders engage in the effort at both the institutional and individual levels, and share goals, responsibilities and accountability for the project.

MSP funded projects must include 1) Content-based Professional Development: participating teachers will receive professional development on the mathematical content needed to effectively provide instruction for middle school students; 2) Needs Assessment: Projects are designed in response to a comprehensive needs assessment of teacher quality and mathematics professional development; 3) Scientifically Based Research: Planned activities carried out by partners must be informed by a review of the scientifically-based research and be justified based on scientifically based research; 4) Evaluation: partnerships must have an evaluation and accountability plan and be able to measure the impact of professional development activities on student outcomes.

MSP partnership grants were provided to colleges and universities to provide professional development, training and technical assistance to a number of partner school districts. Two partnerships began implementation of their projects in 2004. Two additional projects began implementation in 2005.

Background of the Independent Evaluation of the New Mexico MSP Projects

The New Mexico Public Education Department (NM PED) released a request for proposals for the independent, external evaluation of the Mathematics and Science Partnerships (MSP) projects in the May, 2005. Prior to that release, two MSP sites had completed one year of program implementation – New Mexico State University (NMSU) and the University of New Mexico (UNM) – and funding for additional MSP sites was available. A pre-proposal conference for new applicants was held in Santa Fe in January, 2005. A representative of *Folkstone* attended that meeting, and principle investigators from UNM and NMSU were presenters. During the pre-proposal conference, collaboration with the NM PED independent, external evaluation (hereafter referred to as the 'external evaluation' to distinguish it from the local, MSP site level evaluations) was presented as an MSP key feature in the scope of work for new applicants. Western New Mexico University (WNMU) and San Juan Community College (SJCC) were funded as MSP sites for the first time in 2005-2006 and funding for UNM and NMSU, the two original sites, was on-going. Therefore, the first year of the external evaluation (2005-2006), consisted of four MSP sites: WNMU and SJCC, which were new in 2005-2006, and NMSU and UNM were in the second year of program implementation in 2005-2006.

In 2006 Folkstone contracted with Youth Development Evaluation Alliance (YDEA) to conduct the impact evaluation of MSP. YDEA worked closely with Folkstone to design an impact evaluation that would allow the NM PED to draw conclusions about the efficacy of the program to bring about changes in instructional quality and on student mathematics outcomes. Collectively, Folkstone and YDEA had numerous conversations with NM PED and crafted a design and list of data elements that would be necessary to address the agreed upon evaluation questions. YDEA and Folkstone worked together to attempt to secure needed data from NM PED and participating partnerships in 2006-2008.

As an MSP key feature, the second release of MSP funding included stipulations for cooperation with the external evaluation team. Therefore, WNMU and SJCC anticipated collaboration with the external evaluation. Such stipulations may not have been included in the original release, and NMSU and UNM subsequently found the data requirements of the external evaluation to be an unexpected and untenable demand on their resources. Also, during much of first year of the external evaluation, NM PED did not encourage MSP sites to cooperate with the evaluation even though *Folkstone* discussed challenges the evaluation was facing in a telephone conversation as early as November, 2005 with the MSP program manager, Ms. Claudia Ahlstrom. Ahlstrom did intervene on behalf of the external evaluation in May, 2006.

Consequently, despite best intentions from all MSP sites and efforts of *Folkstone* and YDEA to support the local evaluation efforts in data collection for the external evaluation, critical data elements (e.g., the content knowledge survey matched pre and post) were not collected for all sites during the first year. Only WNMU delivered a complete set of data for external evaluation for Year One. Final data sets submitted to YDEA only included WNMU data and were missing key data elements necessary to conduct analyses of several hypotheses. Regardless, YDEA analysts worked with the data provided in order to draw substantive conclusions about the impact of the WNMU partnership on student outcomes. More information about the obtained data supporting the impact evaluation is detailed in subsequent sections of this report.

The next section of this report will describe the purpose of the impact evaluation and the guiding evaluation questions.

Purpose of the Evaluation

The purpose of the external impact evaluation of the New Mexico Mathematics-Science Partnership was to determine if MSP had any measurable impact on mathematics student achievement through teacher participation in MSP professional development activities.

The basic theory of change guiding the impact evaluation hypothesizes that student math achievement is impacted through the individual differences of students and the quality of instruction delivered by mathematics teachers. In practice, the theory is more complex and has been articulated by a number of professional development researchers (Cohen & Hill, 2000; Fishman, Marx, Best & Tal, 2003; Garet, Porter, Desimone, Birman & Yoon, 2001; Guskey & Sparks, 2004; Kennedy, 1998). In an effort to remain consistent with the work of others, we have adopted Yoon, et al's (2009) assumptions about how professional development impacts student outcomes: Professional development influences teacher knowledge and skills which influences classroom teaching practice which impacts student achievement. At all levels of this theory of change is the influence of standards, curricula, accountability and assessments (p. 4)

This study was designed to provide additional information about how and to what extent professional development provided by MSP partnerships influences or impacts the mathematics achievement of students who received instruction from teachers participating in the MSP program. The study is responsive to calls from the Institute of Education Sciences (IES) that scientifically rigorous studies of the impact of professional development focused on improving the quality of mathematics curricula and instructional practice are needed.

Guiding Evaluation Questions

Five questions guided the impact evaluation of the New Mexico Math-Science Partnership:

1. Is there a differential impact of professional development on student outcomes that is due to **variations in the amount** (i.e., dosage in terms of contact hours) of mathematics professional development received by teachers?

2. Is there a differential impact of professional development on student outcomes that is due to **differences in the types** (e.g., content knowledge, and pedagogy specific) of professional development opportunities teachers participate in?

3. Is there a maximally impactful **mix of content-specific and pedagogy-specific** professional development (and in what ratio) that will lead to higher-levels of student outcomes?

4. Is there a differential impact on mathematics outcomes for students whose teachers participated in professional development of **longer duration and intensity** than students who teachers participated in opportunities that were shorter in duration and intensity?

5. Is there a greater positive impact on mathematics achievement outcomes for students who have been taught by **more than one teacher** who has participated in MSP professional development sessions?

Evaluation Design

The evaluation was designed to answer questions that would help disentangle the effects associated with mathematics professional development for teachers and the resultant achievement outcomes of their students. The design specifically addresses questions of *how much professional development and how much exposure do students need* in order to bring about improvement in mathematics achievement. The strength of the analytic approach lies in its ability to partition sources of variance at different levels of the outcomes measurement model and to remove or hold constant variance components that increase measurement error.

Another strength of our approach is that we can model student outcomes at different levels of aggregation (individual student, individual teacher, and partnership) and not lose accuracy or power to detect meaningful effects. All analyses begin at the individual student level and aggregations up to the teacher or school level continue to rely on individual student data.

The design of the impact evaluation is dependent upon the already operational implementation of the intervention and the broader evaluation of the Math-Science Partnerships. Given the variations in focus, design and implementation allowable under the Partnerships grant program, the ability to standardize interventions, measures and dates of measurement are not possible. Furthermore, the external evaluation team was brought into the project after program implementation had already begun. The ability of the evaluation team to influence implementation to maximize the rigor of the evaluation was lost --- and subsequent attempts to secure data that was not part of the original grants funded by NM PED were difficult and in some cases unsuccessful. Furthermore, while data collection efforts were already underway in some phases of the implementation evaluation, the impact evaluation required a substantial data request of participating sites, as well as advanced statistical procedures to create cohesion among measures and times. Before delving into a description of our analytic approach and the variables/data we requested of the NM PED, we will describe the overall evaluation design for the impact evaluation. Given the goal of the impact evaluation, *to measure the extent of change in student mathematics achievement scores as a function of teacher participation in professional development focused on mathematics pedagogy and content*, and the desire to bring as much rigor to the scientific investigation as possible, multiple design choices were considered during the planning phase of the impact evaluation:

1. Experimental design in which teachers (and their students) were randomly selected and then randomly assigned to one of several groups receiving:
    a. no professional development in mathematics (comparison group);
    b. a fully designed sequence of specific professional development opportunities (experimental group)

2. Quasi-experimental design with random assignment of teachers (and their students) to experimental groups but no random selection to groups as described above. Evaluators would use statistical controls to rule out pre-existing differences in teachers and students on the most important predictor variables associated with mathematics achievement;

3. Non-equivalent groups design with comparative groupings of teachers and their students occurring post-hoc via individual and school-level decisions regarding participation in professional development. This design would be strengthened through statistical controls and an analytic design that considers the important predictor

variables associated with the outcome variable as well as covariates and variance decomposition strategies that isolate and explain potential contributors, mediators and moderators of the outcome measure.

Options 1 and 2 above were not possible given the variations in local decision-making and topical emphasis, as well as the goal of ensuring that all teachers have access to professional development opportunities, even those who do not teach mathematics. Consequently, the impact evaluation was constrained to adopt design #3 above and to use advanced statistical methods to examine student outcomes and the effects of teacher professional development. The optimal analytic design in this case is multilevel or hierarchical, in which individual student scores are nested within classrooms (or teachers).

Data collection included three consecutive school years (2004-05 to 2006-07) to allow for sufficient follow-on time following participation of teachers in professional development opportunities. The reader is reminded that students are included in the analysis as a result of their teacher's participation in the MSP program. Comparison students are those students who are included in the analysis as a comparison because their teacher *did not* participate in the MSP program. Consequently, the evaluation team requested state test scores for all middle school students in the 2004-05 through 2007-08 school years from schools whose teachers participated in the MSP program and for a matched comparison teacher in each participating teacher's building. Data from the 2007-08 school year was incomplete and could not be used in subsequent analyses.

Procedures

Student achievement data were received from the NM PED. Additional data files containing information about MSP participating teachers and comparisons were also received from NM PED. The data files were matched using an unique teacher ID and cleaned. The resultant merged file provided the sampling frame by which comparison teachers would be selected and matched to MSP teachers based on teacher and student demographic and background characteristics.

To create a statistically similar comparison group the evaluators matched MSP participating middle school teachers to non-MSP middle school teachers based on teacher ethnicity, gender and mathematics certification. A second level match was conducted based on the demographics of students that teacher's had in their classroom. Several variables were requested by the evaluation team to add precision to the matching process but were not provided by NM PED. These included years of teaching experience and years teaching mathematics (see **Appendix A** for a more detail explanation of the selection and matching process).

Participants

MSP Teachers

Seventy teachers (twenty-five male and forty-five female) participated in the Western New Mexico University's Math and Science Partnership program across three years. Of these, forty-three participated in one year, nineteen participated in two years, and 7 participated all three years. These seventy teachers represented eight school districts, and forty-four elementary, middle and high schools.

Student level data for the year following the teachers' last participation in MSP was provided by the New Mexico Public Education Department for 17 middle school teachers, all of whom were teaching middle school at the time that they participated in MSP. All but two of the 17 MSP teachers were certified/licensed in mathematics. One teacher was removed from the analytic sample because a suitable comparison teacher could not be found. Of the final sixteen teachers in the MSP sample, all but one were certified in mathematics, 62% were female, 56% were Caucasian, and 50% held a Masters degree.

MSP Students

Individual student level data for the year following teachers' most recent participation in MSP were provided for 1,145 students. Of these students, 50% were female, 43% were Hispanic, 46% were Native American, 11% were enrolled in special education, and 79% qualified for free or reduced lunch.

Comparison Teachers

The evaluation team received a data file of non-MSP participant teachers and their associated student achievement data by the New Mexico Public Education Department. Middle-school teachers with mathematics certification/endorsements were isolated to serve as potential matches to MSP teachers. The result of the matching generated 16 comparison teachers matched to 16 MSP participant teachers. Of these comparison teachers all were certified in mathematics, 65% were female, 35% were Hispanic, 65% were Caucasian and 29% held a Masters degree.

Comparison Students

The 16 teachers included as comparisons (those who had no MSP experience) taught 1,218 students across two of the three years that MSP occurred; no data were by provided by NM PED for comparison participants in 2007-2008. Of these 1,218 students 50% were female, 40% were Hispanic, 40% were Native American, 9% were enrolled in special education, and 77% qualified for free or reduced lunch.

A summary of the final sample used for analysis is included in the table below.

TABLE 1
SAMPLE CHARACTERISTICS FOR MSP EVALUATION

	MSP		Comparison	
	TEACHERS	**STUDENTS**	**TEACHERS**	**STUDENTS**
Gender: % Female	62.5%	50.4%	64.7%	49.8%
Ethnic:				
Native American	0%	45.7%	0%	40.0%
Hispanic	37.5%	42.5%	35.3%	40.3%
Caucasian	56.3%	10.9%	64.7%	18.7%
Asian/Pacific Islander	6.3%	.5%	0%	.6%
African-American	0%	.4%	0%	.4%
Special Education Students	----	10.5%	----	9.1%
Subsidized Lunch (F/R)	----	78.8%	----	77.4%
Grade:				
6	----	23.1%	----	17.5%
7	----	60.1%	----	59.1%
8	----	16.8%	----	23.4%

TABLE 1 (continued)
SAMPLE CHARACTERISTICS FOR MSP EVALUATION

	MSP		Comparison	
	TEACHERS	**STUDENTS**	**TEACHERS**	**STUDENTS**
Degrees Earned:				
Bachelor only	50%	----	70.6	----
BA + Masters	50%	----	29.4	----
Beyond Masters	0%	----	0%	----
Total N	16	1,145	17	1,218

Note: Teacher and student demographic data obtained from the New Mexico Public Education Department.

A test of the statistical equivalence of the MSP and comparison group was conducted and is reported in the results section of this report.

Analysis

Overview of Data Analytic Plan

Because the MSP program is designed to impact student achievement *through* the student's experience during instruction with a teacher trained via MSP partnership professional development activities the analytic model first considered three potential levels of variation: 1) individual students; 2) individual teachers; and 3) the school that students and teachers are nested within. As such the data analytic design is hierarchical in nature with three potential levels of nested data: Level – 1 consists of individual students; Level – 2 consists of classrooms or teachers; Level – 3 consists of schools. In this design, students belong to, or are nested within a classroom, and each teacher belongs to, or is nested within a school.

Hierarchical Linear Modeling (HLM, Version 6.06, Raudenbush, Bryk, & Congdon, 2008) was used to assess the impact of teacher professional development on individual student mathematics achievement scores. Fully nested designs have many advantages over previously used techniques to analyze such data, in that they are able to simultaneously model the effects of student, teacher and school level variables on individual student mathematics achievement scores. In what follows, we describe the rationale for the use of a fully nested design to analyze these data, the types and levels of data we used, the hypotheses that were tested, and present the data analytic plan.

Rationale for Using Hierarchical Linear Modeling

Previously, techniques for the analysis of multi-level or nested data have involved either aggregation of data to a higher level, or the disaggregation of data to a lower level. Since the evaluation of the MSP programs is interested in understanding the effects of teacher training on mathematics achievement scores, analytic techniques would have likely involved the aggregation of individual student outcome and demographic data to the class or teacher level. Such aggregation necessitates the use of the classroom mean as the outcome variable, rather than the individual student scores. Aggregation of student scores presents several problems. First, using a class mean as an outcome variable prevents an investigation of how students are differentially impacted by teacher's professional development. For instance, the use of the class mean as the outcome variable does not allow the use of prior achievement scores, or other individual student characteristics, as covariates. Because covariates may account for a substantial amount of variance in the outcome, and in turn increase the power to detect an effect, limiting use of them presents serious problems for the analysis (Porter & Raudenbush, 1987; Raudenbush & Bryk, 2002). Furthermore, the effects of such student level covariates likely indicate that there is heterogeneity of regression, thus violating the assumptions of standard ANOVA or ANCOVA analyses. If heterogeneity of regression *within* classrooms is present, the use of class means as outcomes offers no way to explain why some classes improve dramatically, while others may not improve at all. In other words, there may be characteristics of the students that are more important in some classrooms than others, and the use of class means allows no ability to explain these differences.

Second, aggregation of student mathematics achievement scores to the classroom level reduces the within-class variability to zero, causing the differences between class means to be inflated because there is no estimate of the within class variability. Such aggregation leads to smaller sample sizes (classrooms rather than students), and all issues associated with small sample sizes, including low power to detect effects, and less reliable estimates of effect. (Heck & Thomas, 2000; Raudenbush & Bryk, 2002; Raudenbush & Wilms, 1995).

Third, administrators, teachers and MSP partnership leaders may be most interested in the inferences that can be made about the impact of teacher professional development on subsets of *individual* students. For instance, what is the impact of teacher professional development on students that already have historical patterns of underachievement? Using aggregated class achievement scores to make inferences about student level data is known as the ecological fallacy (Robinson, 1950). This results from an aggregation bias that can result in wildly varying correlations between variables depending upon the level of aggregation of the data. Such a bias will lead to erroneous associations between variables in the model and quite possibly incorrect conclusions about the effect of teacher professional development on student achievement.

For these reasons we proposed to use data at all levels of the model, including student level achievement and demographic data, teacher data related to the MSP trainings, teacher demographics, and characteristics of the training (i.e., focus, type and duration) and school. As will be described later in the results section of this report, our analytic plans were altered due to under specified data from the MSP partnership as well as the inability of the NMPED to provide data on key variables.

Levels and Types of Data

The following variables were requested for a minimum of two years prior to MSP implementation and for at least three implementation years. Unfortunately, NM PED was only able to provide 2004-05 through 2006-07 data. While 2007-08 data was requested, no comparison teacher/student data was sent and only three MSP teacher data elements were sent --- effectively rendering any analysis that included 2007-08 data useless. The data elements and the level of the data requested were:

Student level data (provided by the NM PED):

1) individual student demographic characteristics including: ethnicity, gender, English language proficiency, socio-economic status in terms of subsidized lunch, educational plan (i.e., regular education, special education) for each year the child participated in the New Mexico state reading and mathematics assessments;

2) achievement test scores (require scale scores or transformed scores) from the New Mexico state assessment in reading and mathematics;

3) grade level in the years tested;

4) student attendance (i.e., days attended/days enrolled) during each test year;

5) name of school attended for each year tested;
6) first and last name of the child's teachers in reading (language arts) and mathematics;
7) unique identifier for each student that remains consistent across years.

Participating MSP teacher level data (provided by NM PED and WNMU:

1) the number of hours and type of professional development received for each year of participation,
2) whether the teacher attended annual MSP summer academy or not,
3) teaching endorsements and certifications,
4) teaching concentration,
5) primary content area taught,
6) courses taught,
7) self-assessment administered prior to participation in professional development,
8) test of content knowledge both prior to and following professional development,
9) unique identifier for each teacher that remains consistent across years,
10) first and last name of teacher,
11) name of school, and
12) demographic information for each teacher (includes ethnicity, gender, years of teaching experience, years of experience teaching mathematics, years at their current school, and advanced degrees earned).

Non-participating MSP teacher level data:

In order to establish a meaningful comparison to MSP participating teachers, the evaluation team requested that non-participating teacher and student data be secured and transmitted to the evaluators. For each participating MSP teacher we requested that a matched comparison be selected from the participating teacher's school. NM PED did not provide a specific matched comparison but instead provided data for all teachers in the districts that WMNU was providing

MSP professional development. For teachers not participating in MSP professional development the following variables were requested:

1) the number of hours and type of professional development received for each year,
2) teaching endorsements and certifications,
3) teaching concentration,
4) primary content area taught,
5) courses taught,
6) self-assessment administered at same time as participating teacher,
7) test of content knowledge administered at same times as participating teacher,
8) unique identifier that remains consistent across years,
9) first and last name of teacher,
10) name of school, and
11) demographic information for each teacher (includes ethnicity, gender, years of teaching experience, years of experience teaching mathematics, years at their current school, and advanced degrees earned).

In addition to these two levels of data, the evaluation team believed that school may be associated with a significant portion of the variance in student math achievement. Variables associated with this level included the socio-economic status of the school, the size of the school, the historical achievement profile of the school, frequency of principal changes, the adopted mathematics curriculum and the enacted curriculum, and the amount and type of school or district-sponsored mathematics professional development already in place.

Hypotheses

Our working hypotheses were based on understanding professional development in three ways: (1) as a dosage variable in which teachers will have more or less opportunities, (2) with a specific focus (e.g., content knowledge, and pedagogy specific) and (3) with a measurable intensity and duration (i.e., summer academy, one-day site visits). Several hypotheses were constructed by YDEA and Folkstone for testing:

Dosage of Professional Development

The first hypothesis addressed whether a larger dosage (amount) of professional development would lead to higher levels of subsequent student mathematics achievement.

Hypothesis 1: Students whose teachers participated in more (contact hours) professional development during the summer and school year will have significantly higher mathematics test scores on state assessments.

Focus of Professional Development

The second set of hypotheses dealt with the comparison of different types of professional development opportunities and the relative impact each type has on student outcomes. Three hypotheses were proposed:

Hypothesis 2: There is a differential and statistically significant impact of professional development on student test score performance that is due to the types (e.g., content knowledge, and pedagogy specific) of professional development opportunities teachers' experience.

Hypothesis 3: Participation in math-specific professional development for math teachers leads to higher levels of student mathematics test score performance.

Hypothesis 4: There is a maximally impactful mix of content-specific and pedagogy-specific professional development that will lead to higher-levels of mathematics test performance.

Intensity of Professional Development

The third set of hypotheses were formulated to examine whether the duration and intensity of teacher professional development opportunities have a differential impact on student outcomes.

Hypothesis 5: Students whose teachers participated in professional development of longer duration and intensity (i.e., summer academies) will have higher mathematics test score performance than students whose teachers only participated in brief/low intensity professional development (e.g., 1/2 hour to 2 hour PD visits).

Impact on Student Mathematics Achievement of Professional Development

The sixth hypothesis focuses on the impact of MSP professional development for teachers on the achievement outcomes of students. In particular, we hypothesize that students who have multiple years of mathematics instruction from MSP-trained teachers will have steeper achievement trajectories i.e., larger gains from year to year).

> **Hypothesis 6: Students who have received mathematics instruction from multiple teachers who have participated in MSP partner professional development experiences will have significantly higher mathematics test scores over time.**

Specific Analyses

The original intent of this evaluation was to test the impact of MSP teacher professional development delivered by three partner universities on student mathematics achievement scores on the NMSBA. As proposed, these data conformed to a three-level nested design in which student mathematics achievement (and other student level characteristics) was nested within teachers, who, in turn, were nested within University Partnership. Unfortunately, we were able to obtain individual student level data for only those teachers receiving professional development from one university partner, Western New Mexico University. As stated in the proposed design, in the event that the data showed no variability due to MSP partnership, school designation would be used as the third level variable. Thus, initial analyses were performed to determine the variability of each level of nesting. Where a level contributes negligible variance to the outcome, it was not retained in the analyses testing the hypotheses. The initial model included students nested within teachers, who are nested within school.

Hierarchical Linear Modeling (Raudenbush & Bryk, 2002) software was to be used to test the nested designs proposed for hypotheses 1 through 5. In this type of modeling, the effects of both student level and teacher (or classroom) level variables on individual student mathematics achievement scores can be appropriately determined. Specifically, student level variables may be tested for variability (or differences) within the teacher, and the variability in these student variables can then be considered when determining the impact of teacher professional development.

The original intent of the proposed analytic design was to use the New Mexico State Based Assessment (NMSBA) scale scores as the primary outcome variable for analyses. Unfortunately, NMSBA scores differ across grade levels in terms of the cut points used to determine which proficiency category students fall into. Because of this, the scale score cannot be used as the dependent variable. Because state assessment information most relevant to policy and accountability is whether students have reached proficiency, we determined that the appropriate outcome to be used is a dichotomy between those students scoring "nearing proficiency and below" and "proficient or above". Thus, the outcome for every student is a single binary outcome. Specifically, we are interested in the characteristics of teachers and of students that will predict the probability that students will score either "nearing proficiency or below" or "proficient or above on the NMSBA.

In multi-level models with dichotomous or binary outcomes, the structure of the model is typically non-linear with non-normally distributed errors. This diverges from the assumptions of normality underlying a "typical" multi-level model with continuous (or nearly continuous) outcomes. In this type of analysis, the approach is to base the expected value on the probability that a given student will score in the "proficient or above" (probability of success) category on the NMSBA. Because, in these analyses, we are considering only one score per student, the student has only one "chance" to score in the "proficient or above" category. Single "chances" for success on an outcome conform to the Bernoulli distribution.

In standard multi-level modeling, the level-1 model may be written as

$$Y_{ij}|\mu_{ij} \sim NID(\mu_{ij}, \sigma^2)$$

where Y_{ij} is a level-1 outcome with a given predicted value of μ_{ij} is normally and independently distributed with and expected value of μ_{ij} and a constant variance of σ^2. In the standard model the level-1 predicted value is transformed so that the predictions are constrained to lie within a given interval. In a multi-level model with binary outcomes the level-1 model may be written as

$$\eta_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{pj}X_{pij}$$

where η_{ij} is the transformation that constrains the predicted values, β_{pij} is the slope for a given predictor variable, and X_{pij} is the predictor variable. η_{ij} is a logit function, and is the log of the odds of a student scoring in the "proficient or above" category. Using the logit function allows us to predict a value for a given student that is a probability falling between the values of 0 and 1. This probability may be calculated using the following formula:

$$\varphi_{ij} = \frac{1}{1 + exp\{-\eta_{ij}\}}$$

where φ_{ij} is the probability that a student will score "proficient or above".

The level-2 and level-3 models in multi-level modeling using binary outcomes are the same as in standard multi-level models with continuous outcomes and will be specified for each analysis. Thus, using the above modeling procedure for dichotomous outcomes we are able to determine the impact that students, teachers, and schools have on the probability that students will score proficient or above on the NMSBA math assessment.

Specifically, the level-2 model may be written as:

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + \mu_{qj}$$

Where β_{pij} is the predicted value of a slope specified in the level-1 equation, γ_{q0} is the average intercept across level-2 variables, γ_{qs} is the average slope of the level-2 predictor, and μ_{qj} is the random effect or error of the level two predictor.

Analytic Block 1 – Dosage of Professional Development

This hypothesis is concerned with determining if increased levels of professional development lead to higher student mathematic achievement scores. Within MSP, teachers are given the opportunity to participate in professional development trainings and on-site one-on-one sessions with professional developers, both throughout the school year and via attendance in the MSP summer academy. Not all teachers attend all trainings nor participate in one-on-one site visits and technical assistance, and thus, teachers will vary in the amount of training hours that they receive. This difference in training hours is hypothesized to have an impact on student achievement. We hypothesized that, controlling for teacher characteristics, students being taught by teachers who have participated in a greater number of professional development hours will have higher mathematics achievement scores than those with fewer professional development hours.

Hypothesis 1: Students whose teachers participated in more (contact hours) professional development during summers and school years will have significantly higher mathematics test scores on state assessments.

To test this hypothesis, a three-level hierarchical model was used in which student level (level 1) achievement scores were predicted by the number of professional development hours received by each teacher (level 2). In addition, the effects of teachers being nested within schools (level 3) was tested but school was found to add less than .01% to the explanation of student mathematics achievement. Significant covariates in the model were retained.

Analytic Block 2 – Features of Professional Development

The second block of analyses dealt with the comparison of different types of professional development opportunities and the relative impact each type has on student outcomes. Three hierarchical linear models were planned to test the three hypotheses.

Hypothesis 2: There is a differential and statistically significant impact of professional development on student test score performance that is due to differences in the types (e.g., content knowledge, and pedagogy specific) of professional development opportunities teachers' experience.

Hypothesis 3: Participation in math-specific professional development for math teachers leads to higher levels of student mathematics test score performance.

Hypothesis 4: There is a maximally impactful mix of content-specific and pedagogy-specific professional development that will lead to higher-levels of mathematics test performance.

These analyses were to be exploratory since little is known about the most effective mix of content-specific and pedagogy-specific professional development in raising student test scores. Data received from professional development logs did not provide information at sufficient specificity to code the emphasis of the professional development experience so this block of hypothesis tests were discarded from the evaluation.

Analytic Block 3 – Duration and Intensity

The third block of analyses examined the impact of duration of teacher professional development opportunities on student outcomes.

Hypothesis 5: Students whose teachers participated in professional development of longer duration and intensity (i.e., summer academies) will have higher mathematics test score performance than students whose teachers only participated in brief/low intensity professional development (e.g., 1/2 hour to 2 hour professional development site visits).

A number of professional development sessions are made available to teachers.

Teachers may attend professional development sessions or site visits by professional development providers given during the school year (ranging from half an hour to several hours in a given day), and/or professional development sessions offered in summer academies (intensive multi-week trainings). Teachers may participate in both types of professional development opportunities, or may opt to participate in just one. As a result, teachers may vary in the duration and intensity of training they have received. This variation will likely have impact on student mathematics achievement.

The model tested is a two-level hierarchical model similar to those previously outlined. Student mathematics achievement scores (level 1) will be predicted by the duration and intensity of training received by teachers (level 2). We expect professional development trainings with longer duration will have a positive impact on student mathematics achievement scores.

Analytic Block Four---Impact on Student Achievement

The analyses for each of the above hypotheses were designed isolate information about how much and what type of professional development is needed to impact teacher classroom practice at levels sufficient to impact student outcomes. Thus, the analyses address the *features* of professional development and do not consider the *long term consequences* of professional development opportunities. These long term consequences are, arguably, the most important outcome of this evaluation.

Increased and ongoing exposure to teachers who have had mathematics-specific professional development may be observed by analyzing the rate at which students' math achievement scores increase (or decrease) with increased exposure to these teachers. We propose that this rate of growth will be positively impacted when a student is taught by more than one teacher who has participated in professional development focused on content knowledge and content pedagogy. Thus, each semester that a student is exposed to a teacher who has received MSP professional development, should have an increasingly positive impact on that student's mathematics achievement scores.

Hypothesis 6: Students who have received mathematics instruction from multiple teachers who have participated in MSP partner professional development experiences will have significantly higher mathematics test scores over time.

Students being taught by teachers receiving professional development in the school years 2005-06 and 2006-07 were used for these analyses. In contrast to the other analytic blocks, the appropriate analysis examining this hypothesis is a Latent-Growth Trajectory (LGT) Analysis. In this type of analysis, the repeated measurement of math achievement is used to estimate a latent growth trajectory for each student for whom data has been collected. In LGT a growth trajectory for each student may be calculated, and from this an average rate of growth computed. It is then possible to calculate how variable growth trajectories are from student to student, and whether the variability in growth trajectories may be predicted by the level of exposure to teachers who have received professional development. We expected that greater exposure to teachers, who have received professional development, will be related to greater increases in math achievement scores.

Results

In our analyses of the first five hypotheses we began with a three-level, fully unconditional model; no student level, teacher level, or school level predictor variables were specified. This analysis partitioned the total variability in post professional development student mathematics achievement into its three components: variability among students within classrooms, variability of among teachers within school, and variability among the schools. From this we were able to estimate the proportion of the total variance that exists within each of these components. If there was no significant variation at any one of the three levels, such an analysis allowed for the removal of that level from the model, and would revert the analysis to a two-level design.

Upon determination that there was sufficient variation at each level of the analysis, a series of exploratory analyses were conducted to determine whether student level (i.e., grade, age, gender, ethnicity), teacher level (i.e., years experience, gender, ethnicity), and school level (i.e., number of trainings offered, types of trainings offered, hours of professional development offered) covariates contribute to the explanation of variance in student achievement scores. Covariates that attain significance were retained in the analysis.

Changes to the Analytic Plan based on Data Obtained

The original intent of this evaluation, and specifically Analytic Blocks One, Two, and Three, was to test the impact of MSP teacher professional development delivered by three partner universities on student mathematics achievement scores on the NMSBA. As proposed, these data conformed to a three-level nested design in which student mathematics achievement (and other student level characteristics) was nested within teachers, who, in turn, were nested within University Partnership. Unfortunately, we were able to obtain individual student level data for only those teachers receiving professional development from one university partner, Western New Mexico University (WNMU). School-level variation in test scores (.01%) did not reach significance and indicated that school was not a significant source of variance in explaining student test score performance. As a result our hypotheses will be tested as a two-level, rather than a three-level, nested design. Specifically, student mathematics achievement will be nested within teachers receiving MSP professional development from Western New Mexico University.

Other important changes to the analytic plan occurred as a result of the data obtained and the initial analyses:

1. Mathematics achievement scores obtained through the NMSBA state assessments are not equatable across grades --- meaning that a score of 690 in the 6th grade (Proficiency Level 3) is not equivalent to a score of 690 in the 7th or 8th (Proficiency Level 2) grades. [1] As such, meaningful change in achievement scores from one year to the next based on the obtained scale scores of students cannot be directly calculated. Instead, proficiency levels were used as the primary mathematics outcome measure for each year a student was tested. A weakness of this approach is the precision of the measurement. Scales scores are interval level data and proficiency categories are ordinal level data. Small changes in scale scores may not translate into changes in proficiency levels. So, students who make small progress from one year to the next in terms of change in scale score but do not move from one proficiency level to another do not appear to have had any change in mathematics achievement. While our adopted approach may lack something in terms of sensitivity to change, it does capture the most meaningful changes that school districts and schools are held accountable for – that being the movement of more students into proficient and above categories.

2. Several key variables were not provided as requested to the NMPED. Our models of student mathematics achievement included total years of teaching and years of teaching mathematics for both MSP teachers and matched comparisons. The evaluation team did not receive the data for these two variables for the comparison group of teachers. Both of these variables were to be used in selecting a matched comparison group of teachers and then were planned for inclusion as covariates in models of mathematics achievement. Consequently, selection of the sample and subsequent analyses could not control for years of teaching experience.

3. Professional development logs (as completed by WMNU professional development providers) did not contain the necessary level of specificity in order to code professional

[1] For more information on NMSBA proficiency levels and cut scores see http://www.ped.state.nm.us/div/acc.assess/assess/NMSBA_Statistics_and_Data/downloads/Scale%20Scores%20_2_%20_2_.pdf

development activities in terms of type and focus. Consequently, hypotheses 2, 3, and 4 could not be tested and were excluded from the impact evaluation.

Initial Analyses

First, we examined the extent to which MSP participating teachers and comparison teachers differed from each other in terms of their demographic and background characteristics, self-assessments, and pre to post content knowledge change.

Teachers

Samples of teachers in the MSP program and in the comparison group were not statistically different in demographic characteristics (gender (χ^2 = .02; p ≤ .90), ethnicity (χ^2 = 1.71; p ≤ .56)) based on univariate tests. Groups did not differ in the number of degrees earned (χ^2 = 1.46; p ≤ .23). The evaluators requested years of teaching experience for both the MSP participants and their comparisons but comparison group teacher experience was not provided by NM PED for this evaluation and thus no direct comparisons on this variable could be made. MSP participating teachers' years of overall teaching experience was 10.9 years and their years of experience teaching mathematics was 5.4 years.

Students

Students, by virtue of their teacher being selected for inclusion in the comparison group or their teacher being a participant in MSP, did not differ significantly across groups on gender (χ^2 = .07; p ≤ .79), percent of special education students in teachers' classes (χ^2 = 1.365; p ≤ .24), percent of students receiving federally subsidized lunches (χ^2 = .62; p ≤ .43), or by grade level of student. The MSP and comparison group students did differ in terms of ethnicity (χ^2 = .28.48; p ≤ .00). Observed percentages indicate that both groups were predominately Native American and Hispanic (MSP =88.2%; Comparison = 80.3%) and the comparison group had significantly more Caucasian students (MSP = 11%; comparison = 19%).

Schools

Initial analyses were conducted to determine if MSP schools differed from each other in terms of the amount professional development participants engaged in through MSP. The results indicated that there were no significant differences in the number of contact hours of MSP professional development experienced by teachers in each school (F = 1.01; p= .502). The mean number of contact hours (including participation in MSP Summer Academies was 52.6 (sd = 19.7) over the course of two years (2005-06 and 2006-07). The minimum and maximum number of contact hours across schools was 30.5 to 82.5. Unfortunately, we were unable to determine the amount of professional development received by teachers in schools in the comparison group.

Inclusion of Covariates

Given that there was significant variation among students and teachers, the next initial analysis was performed to determine whether student and/or teacher characteristics accounted for a significant proportion of this variation. If a student or teacher characteristic accounted for significant amounts of variance, that characteristic was retained in the model for testing of hypotheses 1 through 4. Most student-level and teacher-level covariates were found to be non-significant sources of variation in student mathematics achievement and were excluded from analyses.

Table 2
Level-1 Covariates included in the initial model

Variable name	Coefficient	Odds Ratio (95% CI)	t-value	df	p
Attendance Rate	2.09	8.08 (0.43,152.89)	1.39	1073	.164
Lunch status	-0.12	0.89 (0.55,1.43)	-0.49	1073	.621
Ethnicity	0.08	1.09 (0.79,1.49)	0.51	1073	.611
Grade	0.22	1.25 (0.60,2.62)	0.59	1073	.533
Gender	-0.35	0.71 (0.46, 1.09)	-1.56	1073	.12
Special Ed Status	-0.08	0.93 (0.33,2.59)	-0.14	1073	.89
NMSBA Reading Achievement*	2.50	12.20 (6.93,21.46)	8.67	1083	<.001
Previous NMSBA Math Achievement*	2.86	17.48 (10.18,30.02)	10.37	1083	<.001

*These variables retained as Level-1 covariates, values are for final level-1 specification without other covariates in the equation

The only significant student-level (Level 1) covariates in the hypothesis testing models were: reading achievement in the year(s) tested; and a prior measure of mathematics achievement.

Covariates that were non-significant sources of variance in math achievement were: ethnicity, gender, grade level, attendance rate, subsidized lunch status, and special education status.

The only significant teacher-level (Level 2) covariate in the hypothesis testing models was teacher ethnicity. Non-significant covariates included: gender, degrees earned, change in content-specific knowledge measured through content-specific assessments administered pre and post professional development, classroom attendance rate, and mathematics certification. The evaluators had planned to include self-assessments of teaching practice as a potential covariate in the model but only 4 teachers submitted their self-assessments for evaluation purposes.

Table 3
Level-2 Covariates included in the initial model

Variable name	Coefficient	Odds Ratio (95% CI)	t-value	df	p
Teacher Gender	0.35	1.42 (0.55,3.69)	0.78	17	.445
Teacher Ethnicity*	-0.56	0.57 (0.35,.95)	-2.31	19	.03
Degrees Earned	0.47	1.60 (0.61,4.21)	1.032	17	..32

Significant covariates were retained and used in the following hypothesis tests.

The first step in our analysis was to determine if student mathematics proficiency was significantly different in the three years of test data for students who have teachers in MSP compared to students whose teachers were not participants in MSP.

Using the hierarchical model tested above and the inclusion of significant level 1 and 2 covariates we found that MSP and comparison group students did not differ in terms of their state math assessment proficiency in each year tested. In addition, as seen in the Figures on the next page, participation in MSP does not provide an advantage to students at any of the levels of the achievement distribution (t(18), -1.67, p = 11). What this means is that students whose teachers participated in MSP professional development do not score significantly different than students whose teachers did not receive MSP professional development. This finding does not speak to potential incremental change over time as exposure to professional development increases for MSP students. That analysis and the results are presented later for Hypothesis 6.

Figure 1
The effect of MSP on NMSBA Math proficiency scores
(Controlling for Reading Proficiency, Previous Math Proficiency and Teacher Ethnicity)

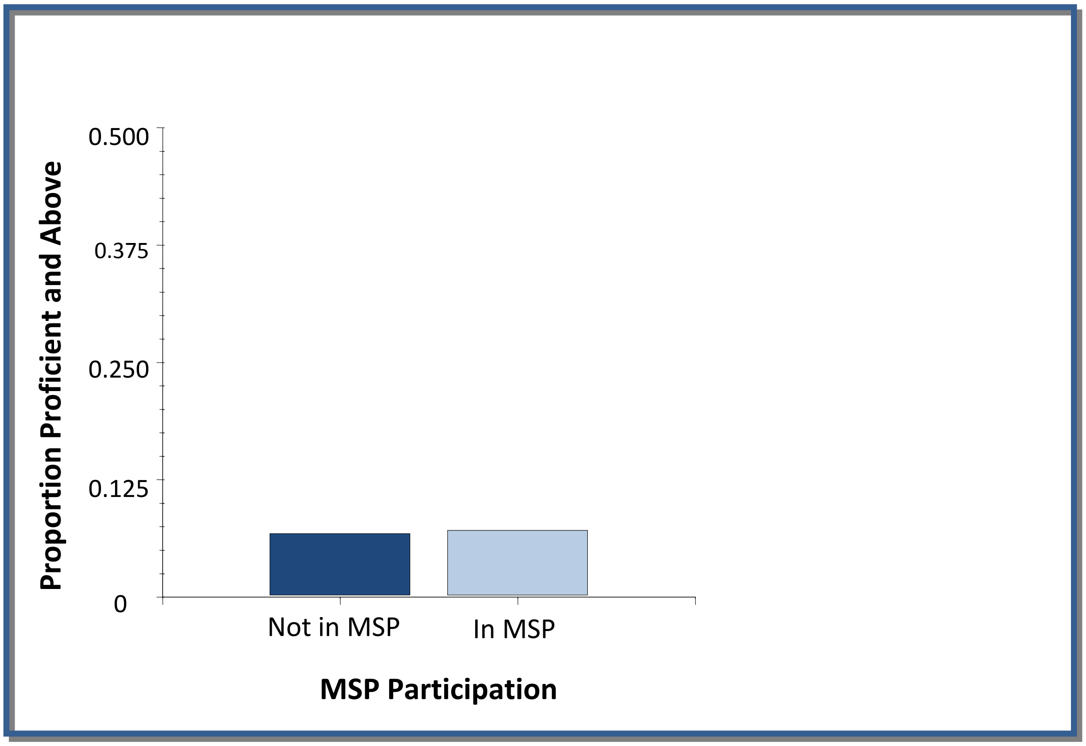


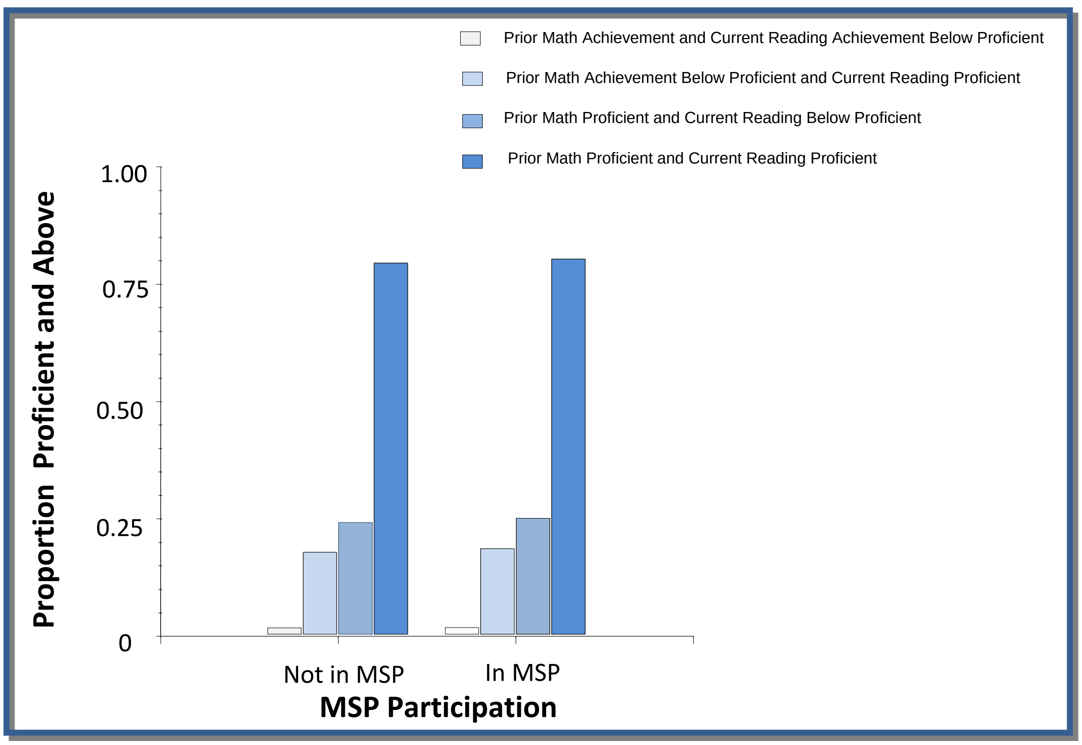
Table 4
Descriptive Statistics for Hypothesis 1: Effect of MSP on Student Outcomes

Level-1 Student Variables					
Variable Name	N	Mean	sd	Min	Max
Previous Math Proficiency	1,090	.11	.31	0	1
Current Year Reading Proficiency	1,090	.44	.5	0	1

Table 4 (continued)
Descriptive Statistics for Hypothesis 1: Effect of MSP on Student Outcomes

Level-2 Teacher Variables					
Variable Name	N	Mean	sd	Min	Max
Teacher Ethnicity	33	3.55	.67	1	4
In MSP	33	.49	.51	0	1

Figure 2
The Differential Effect of Prior Math Achievement and Reading Achievement on Current Math Achievement based on Participation in MSP



While these findings presented in Figures 1 and 2 are important, they do not address other more important questions regarding professional development. Subsequent analyses, presented in hypotheses 1 and 5 focused on just MSP teachers and their students to determine if the amount (e.g., dosage) and type (extended time in MSP Summer Academies versus brief site visits by professional developers) of MSP professional development that teachers received brought about changes in instruction sufficient to impact student outcomes.

Hypothesis Tests

Hypothesis 1: Students whose teachers participated in more (contact hours) professional development during the summer and school year will have significantly higher mathematics test scores on state assessments.

Two specific analyses were conducted:

1. A two-level model with teacher ethnicity as a teacher-level covariate and reading achievement as a student level covariate.
2. A two-level model with teacher ethnicity as a teacher-level covariate and reading achievement and prior math achievement as student-level covariates.

Table 5

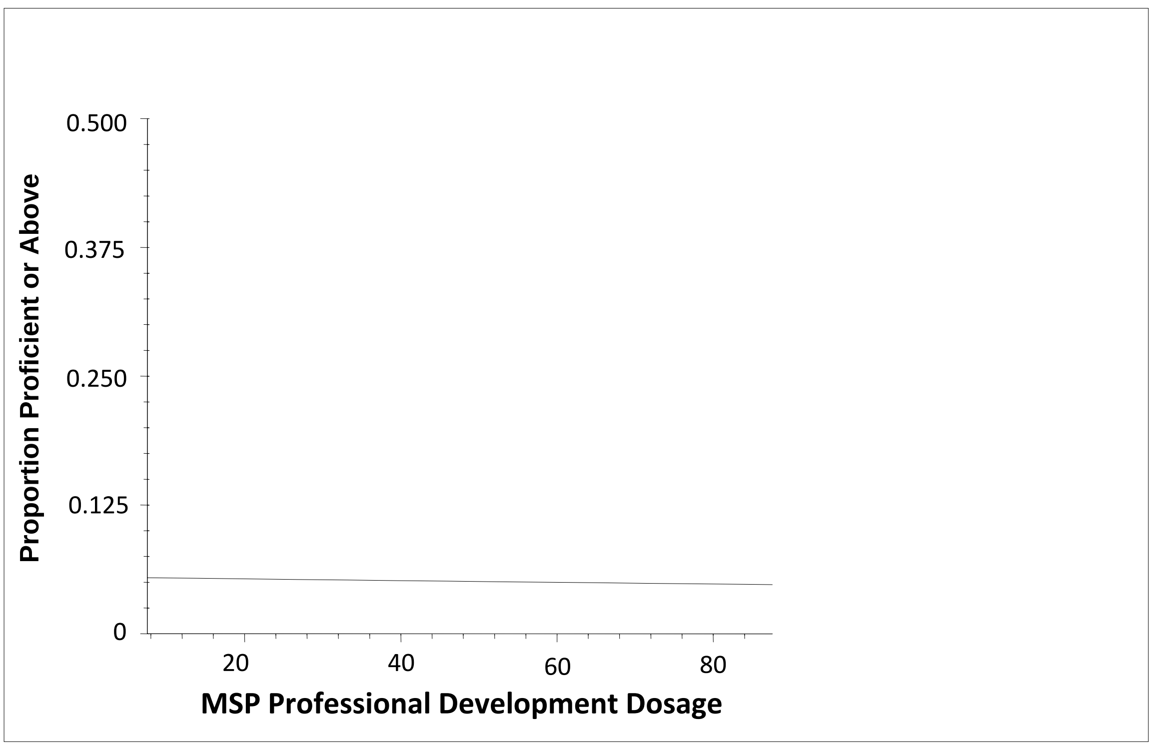
Descriptive statistics for Hypothesis 1: Effect of Professional Development Dosage on Subsequent Student Mathematics Achievement

Level-1 Student Variables					
Variable Name	**N**	**Mean**	**sd**	**Min**	**Max**
Previous NMSBA Math Proficiency	394	.08	.27	0	1
Current NMSBA Reading Proficiency	394	.46	.50	0	1
Level-2 Teacher Variables					
Variable Name	**N**	**Mean**	**sd**	**Min**	**Max**
Teacher Ethnicity	16	3.44	.81	1	4
Contact Hours Summer Academy	16	39.38	14.36	30	60
Contact Hours PD	16	13.89	7.73	.50	29.5
Total PD Contact Hours	16	52.56	19.70	30.50	82.50

Both analyses found that the dosage of MSP-provided professional development received by MSP teachers was a non-significant predictor of student mathematics achievement. With prior mathematics achievement out of the model dosage was non-significant ($t = .84$; $p \leq .41$). When prior math achievement was added into the model the dosage remained a non-significant predicator ($t = -.06$; $p \leq .952$). Prior math achievement was a significant covariate ($t = 5.34$; $p \leq .001$) as was reading achievement ($t = 5.99$; $p \leq .001$) and teacher ethnicity ($t = -2.67$; $p \leq .05$). These analyses were repeated with the ordinal, 4-level NMSBA math proficiency categories, and we found a similar pattern of non-significant findings ($t = 61$; $p \leq .56$)

Figure 3 demonstrates the lack of impact on proximal (most recent to the experienced of professional development) student mathematics achievement by total professional development contact hours (site visits, local trainings, and summer academies).

Figure 3
Effect of Number of Total Professional Development Contact Hours (Dosage) on Student Mathematics Proficiency Scores



This pattern of non-significant effect of professional development dosage was found across the student achievement distribution where historically low performers as well as high performers did not benefit differentially from increased or decreased levels of professional development.

The same was found when we excluded MSP Summer Academy contact hours. We hypothesized that since all MSP teachers in our sample participated in the Summer Academies that there may be a differential impact on student achievement associated with site visit, one-on-one, and local trainings. Consistent with the earlier findings regarding total professional development contact hours, student mathematics achievement did not vary significantly as a function of the number of contact hours teachers had with forms of professional development.

Hypothesis 2: There is a differential and statistically significant impact of professional development on student test score performance that is due to differences in the focus (e.g., content knowledge, and pedagogy specific) of professional development opportunities teachers' experience.

Professional development logs provided by WNMU professional development providers did not provide sufficient detail to code the supports provided to MSP teachers. Agenda for Summer Academies were not received by the evaluation team and could not be coded for professional development focus.

Hypothesis 3: Participation in math-specific professional development for math teachers leads to higher levels of student mathematics test score performance.

This hypothesis could not be tested because all teachers in the MSP project were math teachers and all professional development provided was math-specific. No non-mathematics teachers were included in the data files received.

Hypothesis 4: There is a maximally impactful mix of content-specific and pedagogy-specific professional development that will lead to higher-levels of mathematics test performance?

In order to test this hypothesis the professional development logs of MSP teachers would have had to include the percentage of time (hours, minutes) of exposure to professional development for content knowledge and math pedagogy). While this is what was requested by the evaluation team, the professional development logs only provided total minute(s) and/or hour(s) for any given site visit by professional developers and did not break out the time spent on each.

Examples from the professional development logs illustrate the lack of specificity of information provided to the evaluation team for analysis purposes.

Table 6
Examples from MSP Professional Development Logs

Date/Duration	School	Topic	Planned Activity	Actual Visit Activity
3/26 1 hour	Deming Middle	Teacher development	Curriculum, book adoption, and plan	Curriculum, book adoption, and rest of year plan.
4/20 1 hour	Zuni High	Resources discussion	Resources & Relationship development	Resources & Relationship development.
2/10 2 hours 50 min	Ramah High	Algebra activities	Barbie bungee	Assisted students with their A+ Assessment. This class is an enrichment class used for students to catch up in math.
4/19 2 hours	Quemado High	Graphing coordinates & linear equations	New car prices activity with Internet searches	New car prices activity with Internet searches.
1/25 2 hours	Animas High	Unknown/ no contact	Planned visitation on needs	Students were preparing for homecoming, so <TEACHER> and I discussed needs and next visit.
1/30 1 hour 50 min	La Plata Middle	Fractions	Folding fraction activity	Started folding activity, but changed to computer assessment.

These six examples are not unique within the professional development logs received from schools participating in MSP. Coding these experiences based on the emphasis on content knowledge and pedagogy was not possible.

Hypothesis 5: Students whose teachers participated in professional development of longer duration and intensity will have higher mathematics test score performance.

This hypothesis could not be tested due to a lack of variability in the data sets provided. All MSP teachers participated in the MSP Summer Academy (either once or twice) and all MSP teachers had site visit, one on one, local training professional development. Consequently, we could not isolate teachers who *only* participated in Summer Academies from teachers who *only* participated in MSP site visit professional development.

Hypothesis 6: Students who have received mathematics instruction from multiple teachers who have participated in MSP partner professional development experiences will have significantly higher mathematics test scores over time.

The purpose of this hypothesis was to examine the long-term impact for students whose teachers have participated in MSP professional development. In this analysis we used Latent Trajectory Modeling to determine if student exposure to MSP-trained teachers provided a measurable advantage in mathematics achievement over time.

Impact with Increasing Exposure

In this analysis, we examined a group of students for whom NMSBA math achievement scores and proficiency categories were available across a three year period. We could then examine the rate of growth or change in these scores over the three year time period, and determine what may be influencing the change. We are particularly interested in understanding the extent to which the change (or growth) in NMSBA math scores is due to the exposure of the student to teachers who have received MSP professional development. What we hypothesized is *greater positive growth over three years if a student has been exposed to more teachers who have received MSP professional development*.

The analysis will differ from those previously reported. The previous analyses were not concerned with the cumulative and differential impact of being exposed to teachers who have participated in MSP. The impact of increased and ongoing exposure to MSP trained teachers is, arguably, the most important outcome of this evaluation.

Increased and ongoing exposure to teachers who have had mathematics-specific professional development may be observed by analyzing the rate at which students' math achievement scores increase (or decrease) with increased exposure to MSP participating teachers. We propose that this rate of growth will be positively impacted when a student is taught by more than one teacher who has participated in professional development focused on content knowledge and content pedagogy. Thus, each semester that a student is exposed to a teacher who has received MSP professional development, should have an incremental impact on that student's mathematics achievement scores.

Definition of Sample Used

The data sample used in this analysis was constructed from the files sent by NM PED, and included students who took math related courses in the 2004-2005, 2005-2006, and 2006-2007 school years. As outlined in Appendix A, these files contained unique identifiers for students and teachers, student and teacher demographics, and NMSBA math and reading scale and proficiency scores. These files were examined and duplicate entries for unique teacher-student pairs were removed. Thus, for each school year for which data were available, each case represented a unique teacher-student pair. Importantly, students were associated with more than one teacher, and from this we were able to determine the number of teachers a student had in a given year that had participated in MSP professional development. Duplicate student records were then removed resulting in a file in which each case was a unique student, with a variable representing the number of MSP teachers they had in that year. The number of resulting cases in each year were: 2004-2005: N = 4,947, 2005-2006: N = 4,495, and 2006-2007: N = 3,501.

The three files were then combined and matched by student identification number. The resulting file contained, for each unique student id, the number of MSP teachers in each year, demographic characteristics, and NMSBA math and reading scores.

Measures

NMSBA Math Proficiency

The primary measure used in this analysis was student proficiency score on the NMSBA math assessment for each of three years in each cohort. Two NMSBA math proficiency indices were used: (1) the original NMSBA four category coding scheme which ranged from 1 'below basic' to 2 'nearing proficiency' to 3 'proficient' to 4 'advanced'; and (2) a dichotomous coding scheme which ranged from 1 'nearing proficiency and below' to 2 'proficient and above.' Descriptive statistics for both variables are given in Table 7.

NMSBA Reading Proficiency

Because reading proficiency accounted for so much of the variance in NMSBA math proficiency in the multi-level model analyses, it was also included in this analysis. Two NMSBA reading proficiency indices were used: (1) the original NMSBA four category coding scheme which ranged from 1 'below basic' to 2 'nearing proficiency' to 3 'proficient' to 4 'advanced'; and (2) a dichotomous coding scheme which ranged from 1 'nearing proficiency and below' to 2 ' proficient and above'. Descriptive statistics for both variables are provided in Table 7.

Cumulative Exposure to MSP Teachers

For each student a cumulative exposure index was created. Specifically, the total number of MSP teachers a student had in the 2005-2006 school year was added to the total number of MSP teachers a student had in the 2006-2007 school year. Descriptive statistics are given in Table 7.

Table 7
Descriptive Statistics for Hypothesis 6: Latent Trajectory Model of the Impact of MSP Professional Development on Student Mathematics Achievement

Variable Name	Year	N	Mean/sd	Min	Max
Four Category NMSBA Math Proficiency	2004-2005	490	1.60/.56	1	4
	2005-2006	490	1.59/.59	1	3
	2006-2007	490	1.77/.60	1	4
Dichotomous NMSBA Math Proficiency	2004-2005	490	.027/.16	0	1
	2005-2006	490	.053/.22	0	1
	2006-2007	490	.08/.27	0	1
Four Category NMSBA Reading Proficiency	2004-2005	490	2.03/.60	1	4
	2005-2006	490	2.19/.67	1	4
	2006-2007	490	2.31/.721	1	4
Dichotomous NMSBA Reading Proficiency	2004-2005	490	.19/.40	0	1
	2005-2006	490	.32/.46	0	1
	2006-2007	490	.43/.50	0	1
Two-Year MSP Exposure Index	---	490	.59/.50	0	2
Ethnicity	All	490	3.70/.90	2	5
Special Education Status	All	490	.15/.35	0	1
Free/Reduced Lunch Status	All	490	.90/.30	0	1
Gender			1.47/.5	1	2
Grade	2004-2005	490	6.02/.134	6	7
	2005-2006	490	7.00/.09	6	8
	2006-2007	490	7.97/.17	7	8

Student characteristics

For each student, ethnicity, special education status, free/reduced lunch status, grade and gender were included. The descriptive statistics and codes for these are given in Table 7.

Analytic Overview

A Latent Growth Trajectory Analysis (LGTA) will be used to examine the effect of teacher exposure on the change in NMSBA math proficiency over a three year period (baseline: 2004-05 to 2006-07). The hypothesis and data are particularly suited to the use of LGTA as we are interested understanding not only how student NMSBA math scores change over time in general, but the extent to which there are differences between students in that rate of change. In addition, we would like to be able to estimate the effect that having teachers who have participated in MSP Professional development has on the change in NMSBA scores over time.

Historically, the analysis of change over time has been limited to the examination of the between person differences in change over time. In this type of analysis we might discuss how people differ from each other in the average degree of change over time. What is not considered in these more rudimentary analyses is the impact that the between person differences have on the within person change over time. In other words, and specific to the current evaluation, we can think of each student as having their own growth trajectory on NMSBA math assessments taken across several different grades. When we combine this student with other students we can talk about how that student's growth trajectory is different from other students in terms of both where they started and the extent to which they had positive (or negative) change over time. If it is the case that students differ from each other both in where they start and the extent of change over time, we are likely to be interested in the reason for this difference, that is the between person factors. LGTA is particularly suited to modeling both the within person growth trajectory and the between person factors that influence those trajectories.

LGTA is a highly structured type of structural equation model (SEM). SEM (Jöreskog & Sörbom, 1993) is an analytic technique that allows for the simultaneous estimation of the relations between observed variables and underlying latent or unobserved constructs, as well as the relationships among the unobserved constructs themselves. In LGTA, the observed repeated measures are assumed to be indicators of an underlying, unobserved and latent, true growth trajectory (McArdle, 1986; Meredith & Tisak, 1990; Muthen, 1991). In the current analysis, we assume that the repeated measures of NMSBA math proficiency scores are indicators of a true (and latent) growth trajectory in NMSBA math proficiency scores. As with any growth trajectory, the latent NMSBA growth trajectory is assumed to have an intercept (or starting point) and a slope (or rate of change) that may be indicated by the observed repeated measurement of NMSBA proficiency scores.

All LGTA models to be estimated in this block of analyses will be analyzed using *AMOS 17.0* Structural Equation Modeling software (Arbuckle, 2008). The maximum likelihood method was used to estimate parameters, and goodness-of-fit was assessed by examining the following indices: the chi-square statistic, the incremental fit index (IFI), the comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA). A model is considered to have adequate fit when the IFI and CFI have values that exceed .90, and when the RMSEA is less than .08 (Bollen, 1989; Byrne, 2001; Kline, 1998).

Initial Analysis

The Unconditional Model

Figure 4 provides a graphic representation of the unconditional LGTA model for NMSBA proficiency scores. The rectangular boxes labeled *NMSBA Math Proficiency 1*, *NMSBA Math Proficiency 2*, and *NMSBA Math Proficiency 3* represent the observed measurement of NMSBA Math assessments across three years. The ovals labeled *NMSBA Math Proficiency Intercept* and *NMSBA math Proficiency Slope* represent underlying, latent growth factors that are defined by the observed variables. The circles labeled *e1, e2, and e3* are the errors in measurement of the NMSBA Proficiency Scores at times 1, 2, and 3 respectively. The three arrows going from the NMSBA Math Proficiency Intercept and the NMSBA Math Proficiency Slope to the three observed NMSBA Math Proficiency measures may be thought of as factor loadings that are associated with the regression of each of the NMSBA Math Proficiency measures on the latent constructs of intercept and slope. The factor loadings relating the three observed NMSBA Math Proficiency score to the latent intercept construct are all fixed at 1 to define the starting point of the NMSBA Math Proficiency Growth Trajectory. The factor loadings relating the three observed NMSBA Math Proficiency scores to the latent slope construct are fixed loadings that best represent the functional form of the growth trajectory over the three time points. Because there are only three time points of NMSBA Math Proficiency Scores available, only a linear growth trajectory may be modeled and the factor loadings are appropriately fixed to 0, 1, and 2 to represent a linear or straight line. The latent constructs of NMSBA intercept and NMSBA slope were allowed to correlate with each other.

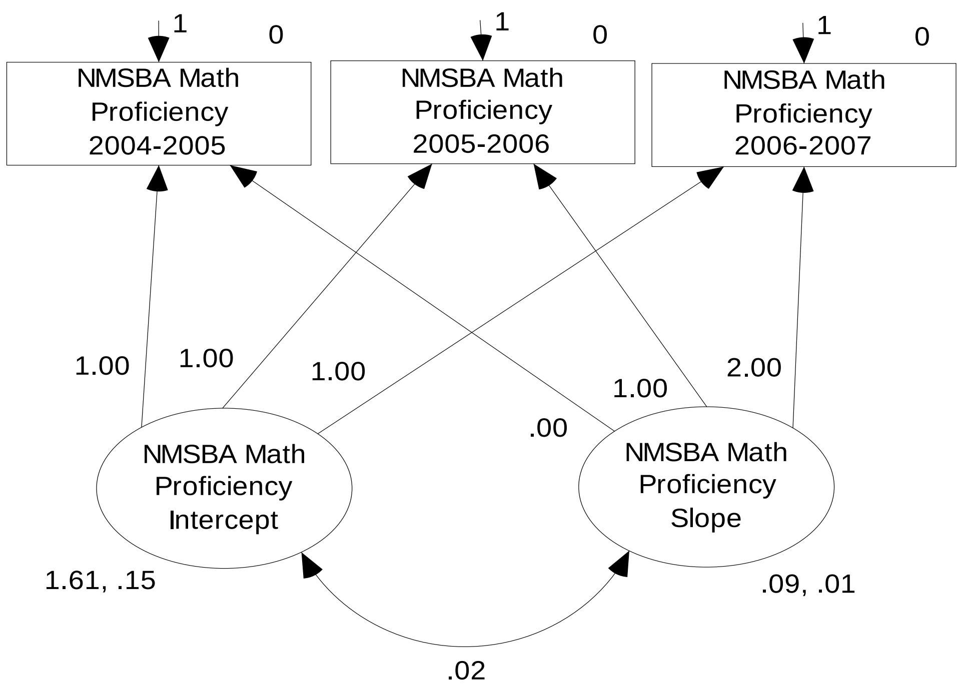
Table 8
Parameter Estimates for the Unconditional Model

Parameter	Estimate	Critical Ratio	*p*
Mean of NMSBA Intercept	1.61	78.10	<.001
Mean of NMSBA Slope	.09	8.22	<.001
Variance of NMSBA Intercept	.15	7.214	<.001
Variance of NMSBA Slope	.006	.55	.583
Covariance of NMSBA Intercept with NMSBA Slope	.021	1.82	.069

The initial model to be estimated (see Figure 4) is the unconditional model in which the means and variances of the two latent factors are allowed to freely vary. Specifically we are allowing the mean initial status, and the mean rate of change to vary freely. Likewise, we are allowing the variability among individuals in initial status and in the rate of change over time to vary freely as well. It is the variance in the two latent constructs that infers the presence or absence of individual differences in the growth trajectories of NMSBA over time. The unconditional model fit the data only marginally well ($\chi^2(1) = 11.63$, $p<.005$, IFI = .98, CFI = .98, RMSEA = .12). Estimated parameters, along with their significance tests are presented in Table 8. Inspection of the estimated parameters indicated that the means of both NMSBA Intercept and NMSBA Slope were significantly different from zero.

Figure 4

Latent Growth Trajectory Unconditional Model



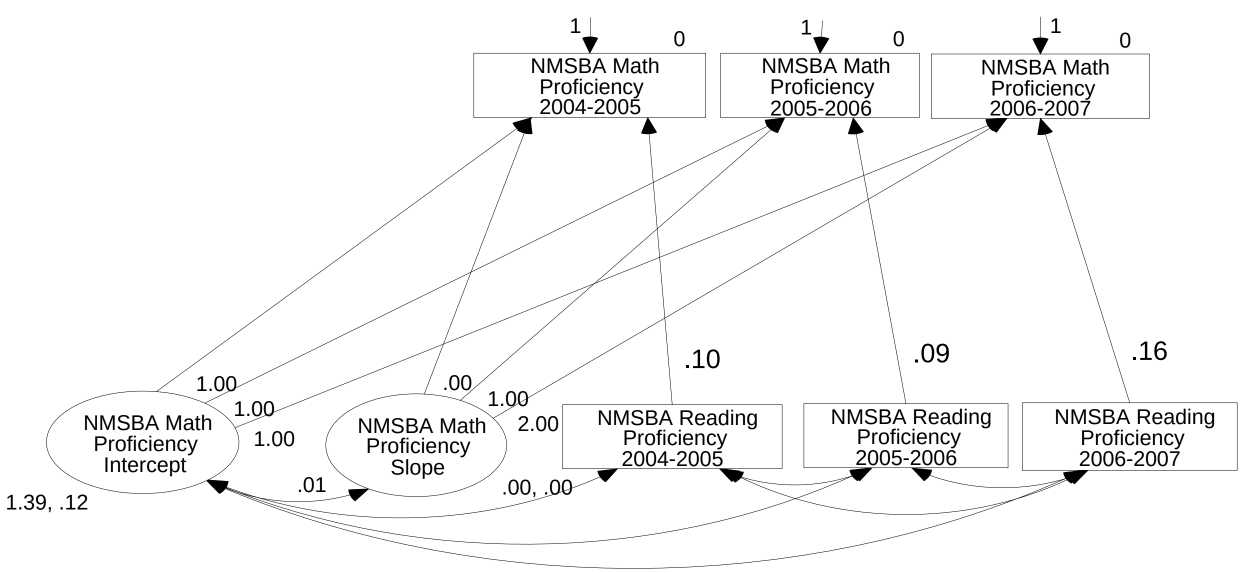
This indicated that the estimated starting point for NMSBA was 1.61 or between the Beginning Step and Nearing Proficiency categories. *The estimated mean rate of change was .09 units per year, or about a tenth of a proficiency category a year. While this rate of change was statistically significant, it does not appear to be practically significant*. Presumably states would prefer that students move from one proficiency category to another in under 10 years. Inspection of the variance components of the latent NMSBA Intercept and Slope indicate that there is significant variability around the mean in students initial status, but that there was not significant differences among students in the rate at which NMSBA scores changed over the three school years sampled. Such lack of variability argues for fixing the variance component on the NMSBA Slope factor to zero, and then observing the change in model fit.

In fact, when the variance on NMSBA was constrained to zero, the model did gain a non-significant increase in model fit ($\chi^2(1) = 11.91$, $p<.005$, IFI = .99, CFI = .98, RMSEA = .08) indicating that, in fact, there is little individual variability in the rate of change in NMSBA proficiency categories over the three school years sampled.

Time-Varying Covariate Model

Figure 5 provides a graphic representation of the LGTA model with time varying covariates for NMSBA proficiency scores. This model is the same as the unconditional model with the following exceptions: NMSBA Reading Proficiency scores for each of the three school years were added as covariates of NMSBA math proficiency. The arrowed lines going from the NMSBA Reading proficiency scores to NMSBA math proficiency serves to remove the variance in NMSBA math proficiency associated with NMSBA reading proficiency scores in the estimation of the underlying latent growth factors. Because no variability among students was found in the latent slope factor when examining the unconditional model, the variance of the latent slope was fixed to zero. The means on both the latent intercept and latent slope, as well as the variance on the latent intercept were allowed to freely vary.

Figure 5
Latent Growth Trajectory Model with Time-Varying Covariates



This LGT model examined the mean level and variability around that mean in the initial status of the growth as well as the mean of change over time in NMSBA math proficiency Scores, controlling for the variance in time-varying NMSBA Reading Proficiency Scores. This analysis was analogous to the cross-sectional multi-level analyses, such that for all estimations involving NMSBA Math Proficiency outcome, we controlled for performance on the NMSBA Reading assessment of the same year. It should be noted that inclusion of a prior year NMSBA Math Proficiency score as a control variable was not needed in these analyses because we modeled a growth trajectory that included that previous year's scores. The time-varying covariate model fit the data well ($\chi^2(1) = 6.72$, p=.24, IFI = .99, CFI = .99, RMSEA = .027). Estimated parameters, along with their significance tests are presented in Table 9.

Table 9
Parameter estimates for the time-varying covariate model

Parameter	Estimate	Critical Ratio	*p*
Mean of NMSBA Intercept	1.39	17.73	<.001
Mean of NMSBA Slope	-.005	-.11	.91
Variance of NMSBA Intercept	.12	6.77	<.001
Variance of NMSBA Slope	Fixed to 0	---	---
Covariance of NMSBA Intercept with NMSBA Slope	.01	1.98	.047
Regression of 2004-2005 NMSBA on 2004-2005 NMSBA Reading Proficiency Score	.10	2.62	.009
Regression of 2005-2006 NMSBA on 2005-2006 NMSBA Reading Proficiency Score	.09	3.21	.001
Regression of 2006-2007 NMSBA on 2006-2007 NMSBA Reading Proficiency Score	.16	5.34	<.001

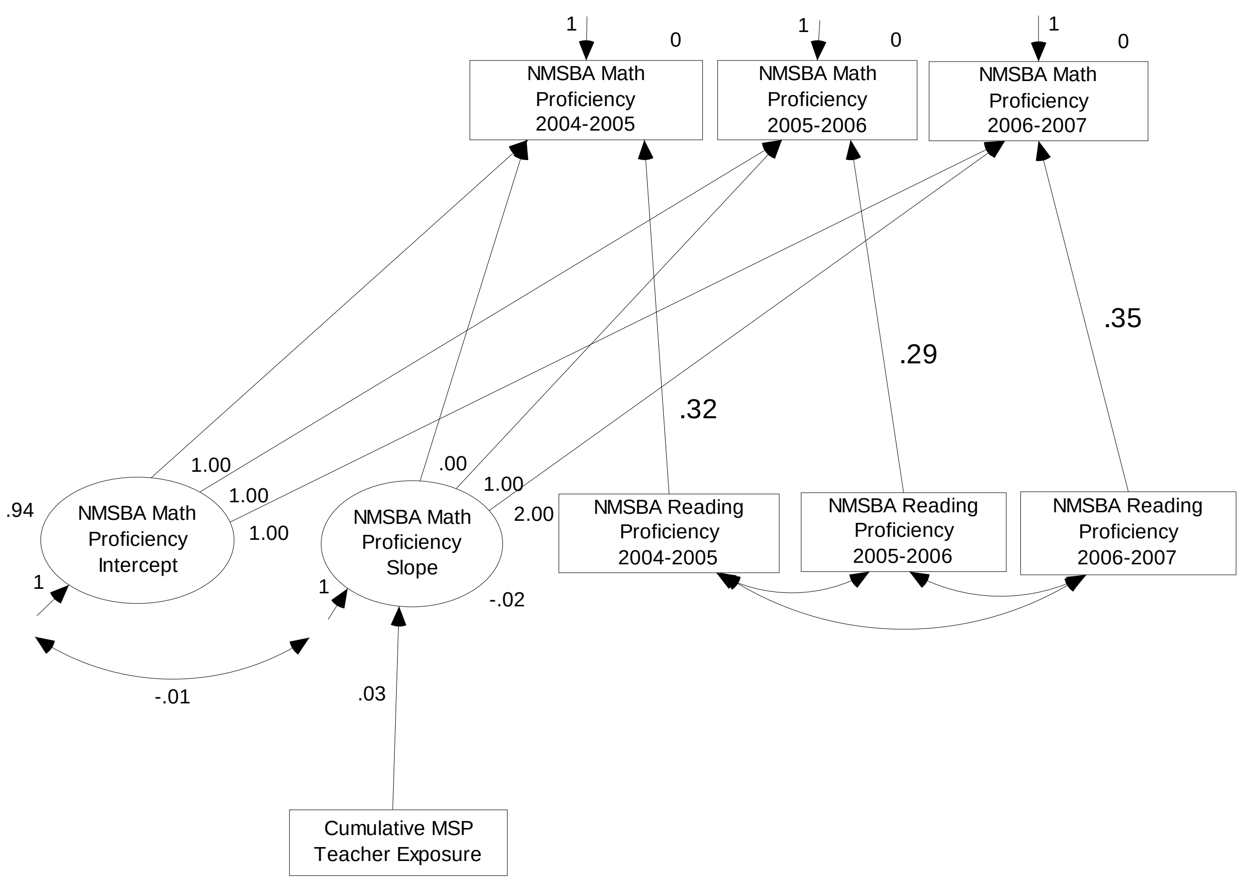
Inspection of the estimated parameters indicated that, unlike the unconditional model, only the mean of the NMSBA math proficiency Intercept was significantly different from zero, indicating that the estimated starting point for *NMSBA Math Proficiency*, controlling for *NMSBA Reading Proficiency* Scores was 1.39. The estimated mean rate of change in *NMSBA Math Proficiency Score* was not significant, indicating that when *NMSBA Reading Proficiency* is controlled for, change over time in the NMSBA is essentially flat. This finding supports our conclusion that the statistically significant mean rate of change found in examination of the unconditional model was functionally or practically insignificant, and that when *NMSBA Reading Proficiency Scores* were taken into account, there were no substantive increases (or decreases) over time in NMSBA math proficiency. Inspection of the variance component of the latent *NMSBA Intercept* indicates that there remains significant variability among students' initial NMSBA proficiency scores.

Effect of Cumulative MSP Teacher Exposure on NMSBA Growth Trajectories

Figure 6 provides a graphic representation of the final LGTA model in which the effect of Cumulative Exposure to MSP teachers was considered. The model is the same as the Time-Varying Covariate model with one exception: *the inclusion of the cumulative exposure index as a predictor of the latent slope factor (achievement trajectory)*. In the previous two models we were able to identify the existence of variation among students at baseline, and functionally non-significant change over time from baseline, of NMSBA math proficiency scores. The strength of the LGT approach, however, is that it allows us to examine the *between student* reasons for variability in growth trajectories. Specifically, we hypothesized that there is a greater positive impact on NMSBA math proficiency for students who have been taught by more than one teacher who participated in MSP professional development sessions. Given that no MSP training took place prior to the 2005-2006 school year, it is theoretically (and practically) unlikely that cumulative exposure to MSP teachers would account for explanatory variance in the mean and variability of student's baseline achievement status. As a result, cumulative exposure to MSP teachers is hypothesized to be predictive of *only* the latent slope factor. That is, to be predictive of the degree to which students change over time on NMSBA math proficiency. When predicting latent constructs in a SEM, the mean and variance of that construct are no longer estimateble parameters. Rather, the variance associated with the mean and variability of the latent factor is estimated by the factor loading of the predictor variable and its associated error. As a result, we will not provide estimates of the mean and variability in the latent growth factors for this analysis.

Several caveats must be made, however, prior to this analysis. First, we found no significant variation among students' rate of change over time in NMSBA math proficiency, even when controlling for concurrent NMSBA reading proficiency scores. The inclusion of a predictor variable to explain little or no variability is likely to add little to the parsimonious understanding of how and why student NMSBA math proficiency scores change over time. Second, when controlling for NMSBA reading proficiency, we found that the average rate of growth in NMSBA math proficiency scores did not differ from zero, indicating that on average student performance was not improving over the three time points included in our sample. Again, including a variable meant to predict average NMSBA growth does not, in this case, add to our understanding of how NMSBA scores change over time. Consequently, given the findings from both the previous SGT models and the cross-sectional multi-level models, we do not anticipate that the inclusion of cumulative exposure to MSP teachers to add useful information to the model.

Figure 6
Latent Growth Trajectory Model of Cumulative Exposure to MSP Teachers



The final model including cumulative exposure to MSP teachers as a predictor of the latent slope factor did not fit the data well ($\chi^2(12) = 124.70$, $p<.001$, IFI = .91, CFI = .91, RMSEA = .14), indicating that, as expected, the inclusion of this explanatory variable (exposure) did not add to a parsimonious understanding of *how* NMSBA math proficiency scores change over time. Estimated parameters, along with their significance tests are presented in Table 10.

Table 10
Parameter Estimates for the Cumulative Exposure Model

Parameter	Estimate	Critical Ratio	*p*
Regression of NMSBA Math Proficiency Slope on Cumulative MSP Teacher Exposure	.03	2.20	.03
Error Covariance of NMSBA Intercept with NMSBA Slope	-.01	-.49	.63
Regression of 2004-2005 NMSBA on 2004-2005 NMSBA Reading Proficiency Score	.32	10.47	<.001
Regression of 2005-2006 NMSBA on 2005-2006 NMSBA Reading Proficiency Score	.29	13.50	<.001
Regression of 2006-2007 NMSBA on 2006-2007 NMSBA Reading Proficiency Score	.35	13.10	<.001

Inspection of the estimated parameters, though, yielded a surprising result. Specifically, cumulative exposure to MSP teachers significantly predicted the NMSBA math proficiency latent slope factor ($B = .031$, $p=.03$), indicating that *for each additional MSP teacher a student is exposed to, the student's gain on the NMSBA math proficiency will increase by .03 units.* Interpreted differently, given the current impact of MSP teacher on NMSBA math proficiency (controlling for NMSBA Reading Proficiency), for a student to have a growth by just 1 proficiency category, they would need to be exposed to over 30 teachers who have participated in MSP professional development. The only other parameter estimate that changed from previous analyses was the covariance or relationship between the error terms on the latent intercept and latent slope. In the previous two models the relationship between the latent intercept (baseline) and slope (trajectory) was significant or nearly significant, indicating that students' initial status was related to their rate of change over time. This, is a rather obvious effect – when students start out performing well, they tend to have more positive growth rates over time that those who start out performing poorly. The current result indicates that regardless of how well a student performs initially; their rate of change over time will be essentially flat. That is that students who start high stay high, while those who start low, stay low. This single finding in combination with our early results suggests that the MSP program, as implemented in the WNMU partnership, did not significantly impact student achievement.

Conclusions

The results of the analyses of data from this impact evaluation of the New Mexico Math-Science Partnerships program lead the evaluation team to conclude that 1) the amount, type or focus of professional development support received by MPS teachers was not sufficient to bring about meaningful changes in instructional practice that would have a positive and practical impact on student mathematics achievement; 2) through multiple advanced statistical methods a consistent message has emerged – student achievement did not change dramatically over the period of time studied by the evaluators – even with the provision of professional development purportedly targeted at mathematics content-specific knowledge and pedagogy. While several key hypotheses could not be tested because of missing data, the two most important evaluation questions and hypotheses did get answered: Does increased professional development make a difference in the mathematics outcomes of students? And, does exposure to MSP teachers provide an incremental advantage in mathematics performance beyond that of non-MSP participating teachers? In both cases the results from this study provide no compelling evidence that MSP has changed the mathematics achievement trajectories of students. Whether one considers the most policy-relevant outcome: moving students into the proficient and above category or looking at movement into the next highest proficiency category, MSP does not appear to have dramatically improved the odds that students will progress above and beyond that of students who were not exposed to MSP.

A relevant observation of the evaluation team experience in this project is that data quality and the fulfillment of data requests by the New Mexico Public Education Department are in need of improvement. Data files received by the evaluation team contained duplicate student records, missing data elements, and coding errors contributing to significant delays in initiating analyses. Entire years of student data were not provided and requests for data were not fulfilled for months. The original scope of this evaluation included three MSP partnerships --- only one was responsive to the calls from the external evaluator to participate. The results obtained in this evaluation would have been more robust, compelling, and meaningful had all three partnerships contributed to the evaluation, had the NM PED been consistently responsive to data requests by providing the requested data in a timely manner and actively supporting and strongly urging the participation of MSP partnerships in this important evaluation of the impact of professional development on student mathematics achievement.

References

Arbuckle, J. L. (2008). *AMOS* (Version 6.06). Chicago, IL: SmallWaters Corp.

Birman, B. L., LeFloch, K. C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., Wayne, A., & Yoon, K. S. (2007). *State and local implementation of the No Child Left Behind Act, volume II-Teacher quality uner NCLB: Interim report*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Bollen, K. (1989). *Structural equations with latent variables.* New York: Wiley.

Byrne, B. M. (2001) *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum Associates.

Clotfelter, C, Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement in high school: Cross-subject analysis with student fixed effects, *NBER Working Paper No. 13617* (ED501993).

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*(2), 294-343.

Connell, J. P. and Broom, J. (2004). *The Toughest Nut to Crack: First Things First's (FTF) Approach to Improving Teaching and Learning*. Report prepared for the US Department of Education. Philadelphia: Institute for Research and Reform in Education.

Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A Review of State Policy Evidence. *Education Policy Analysis Archives, 8*(1). (Retrieved from http://epaa.asu.edu/epaa)

Felter, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7*(9). Retrieved from http://epaa.asu.edu/epaa/v7n9.html.

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*, 643-658.

Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915-945.

Guskey, T. S., & Sparks, D. (2004). Linking professional development to improvements in student learning. In E. M. Guyton & J. R. Dangel (Eds.), *Research linking teacher preparation and student performance: Teacher education yearbook XII* (pp. 233-247). Dubuque, IA: Kendall-Hunt.

Heck, R.H. & Thomas, S.L. (2000). *An Introduction to Multilevel Modeling Techniques.* Mahwah, NJ: Lawrence Erlbaum Associates.

Jöreskog K. G., Sörbom D. (1993). Structural Equation Modeling with the SIMPLIS Command Language. Chicago: Scientific Software Inc.

Kennedy, M. (1998). *Form and substance of inservice teacher education* (Research Monograph No. 13). Madison, WI: National Institute of Science Education, University of Wisconsin-Madison.

Klecker, B. M. (Nov, 2002). *The relationship between teachers's years-of-teaching experience and students' mathematics achievement*. Paper presented at the annual meeting of the Mid-South Education Research Association, Chattanooga, TN.

Kline, R. B. (1998). *Principles and practice of structural equations modeling.* New York: The Guilford Press.

McArdle, J.(1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, 16, 163-200

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrica*, 55, 107-22

Moore, W. P., & Esselman, M. E. (1994). *Exploring the context of teacher efficacy: The role of achievement and climate.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Muthen, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC: American Psychological Association.

Porter, A.C., & Raudenbush, S.W. (1987). Analysis of covariance: Its model and use in psychological research. *Journal of Counseling Psychology*, 34(4), 383-392.

Quint, J. C., Akey, T. M., Rappaport, S., & Willner, C. J. (2007). *Instructional Leadership, Teaching Quality, and Student Achievement: Suggestive Evidence from Three Urban School Districts*. New York: MDRC (Retrieved from www.mdrc.org).

Raudenbush, S. W., & Bryk, A.S. (1988). Methodological advances in studying effects of classrooms and schools on student learning. *Review of Research in Education,* 15, 423-476.

Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models (Second Edition*). Thousand Oaks: Sage Publications.

Raudenbush, S.W., & Willms, J.D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics,* 20*(*4), 307-335.

Robinson, W. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357

Rockoff, J. E. (2003). *The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data*. Harvard University, Kennedy School of Government: Cambridge, MA. (Retrieved from http://econwpa.wustl.edu:8089/eps/pe/papers/0304/0304002.pdf).

Tenenbaum, G. (1988). The relationship between the quality of instruction and intellectual achievement responsibility following positive and negative educational outcomes. *Journal of Experimental Education, 56*.

U.S. Department of Education (2009). *Institute of Education Sciences Request for Proposals: Education Research Grants, CFDA Number 84.305A*. p. 13 (Retrieved from www.ed.gov/about/offices/list/ies/2009-305A.pdf).

U.S. Department of Education. (2007). *Institute of Education Sciences Request for Proposals: Education Research Grants, CFDA Number 84.305*. p. 14 (Retrieved from www.ed.gov/about/offices/list/ies/2007-305.pdf).

Yoon, K.S., Duncan, T., Lee, S.W-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-033). Washington DC: US Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. (Retrieved from http://ies.ed.gov/ncee/edlabs).

Appendix A

MATCHING AND SELECTION OF COMPARISON TEACHERS AND STUDENTS FOR MULTI-LEVEL ANALYSES

Construction of the student level file

The data sample used in the multi-level models was constructed from files sent by the New Mexico Department of Education. These files included all middle school students who took math related courses in the 2004-2005, 2005-2006, 2006-2007 and 2007-2008 school years. Included in this file was the name and identification number of the teacher who taught the course, the name of the course, teacher characteristics (ethnicity, gender, degrees earned, certifications, and a code indicating participation in MSP), student characteristics (ethnicity, gender, free/reduced lunch status, special education status, and grade), as well as student scale scores and proficiency categories on the NMSBA math and reading tests.

For each year the files were examined to determine that there were not duplicate entries for each unique student/teacher combination. In the event that there were duplicates they were removed from the file. This resulted in a file for each year containing one record for each unique student – teacher pair.

The year specific files were then combined to form an overall base file of student records, with one exception. Specifically, teacher student records from the 2004-2005 school year were not included in this file because no MSP participation had taken place prior to, or during, that school year. The resulting file contained one record for each unique teacher-student combination for each year. In some cases students within a year were enrolled in more than one math class taught by different teachers. In this case, each of these records were retained in the base file. The resulting file consisted of 8,877 unique teacher-student pairs across three school years

The next selection step was designed to identify which school year was the most proximal to teacher participation in MSP. The years teachers participated in MSP were provided by the evaluation team and that information was included in the base file described above. A summary variable for each teacher was created that indicated the year of most recent participation. Unique teacher-student records from the year of last participation were then selected. In some cases, unique teacher-student records were from the year prior to a teacher's participation in MSP. These records were not included in the final file. In other cases, unique teacher-student records were from the year following participation in MSP (i.e., participation in 2007-2008, but data available for teacher-student pair in 2006-2007).

These unique teacher-student pairs were also not included in the final file. All non-MSP participants were excluded from the preceding selection criteria and were retained in the file. The resulting file consisted of 6,277 unique teacher-student pairs (1,278 MSP and 4,999 non-MSP) across the three school years. It should be noted that of the 881 unique teacher-student pairs obtained for the 2007-2008 school year, all teachers had participated in MSP at some point in during either the 2005-2006, 2006-2007, and/or 2007-2008 school years. As a consequence, these teacher-student records were either included as part of the MSP teacher-student sample or were excluded from possible inclusion in the comparison sample. Thus, we were not able to obtain any comparison teachers for the 2007-2008 school year.

Next, we selected the unique teacher-student pairs that would serve as comparisons in the analysis. As before the records are student level data that includes teacher information for that student's math classes. To determine the comparison sample of unique teacher – student pairs, teacher and student summary statistics were produced for those unique teacher-student pairs for which the teacher was a MSP participant. These summary statistics included teacher ethnicity, teacher gender, degrees earned by teachers, average student gender, average student ethnicity, average student special education status, and average student free/reduced lunch status. In addition, the number of students enrolled in MSP teachers' courses was also obtained. The same summary statistics were also obtained for the non-MSP teacher-student pairs. Matching of teacher and the characteristics of the student in their classes was conducted as follows:

1. Comparison teachers who were math certified or endorsed were selected.
2. Comparison teachers were matched to MSP teachers based on teacher ethnicity, gender, degrees earned, as well as grade taught. This matching was successful with two exceptions. First, one MSP participating teacher reported an ethnicity of Asian, but no non-MSP teachers were Asian. Thus, no match was found regarding ethnicity for this single teacher, and other match criteria took precedence for this single MSP teacher. Second, among MSP teachers' classes, 7th graders made up 60% of the students, while grade was relatively evenly distributed among the non-MSP teachers classes. In addition, one MSP teacher who taught 9th grade math was included in the original file, but no potential comparison teacher was provided by NM PED who taught 9th grade math, the MSP teacher was excluded from the analyses. As a result, while an attempt was made to match on grade, because the teacher characteristics took some priority, matching on the grade variable alone was not entirely successful.

3. The remaining possible comparison teachers were then matched to MSP teachers based on the ethnic make-up of the students in their classes. When differences existed and no suitable match could be found, matching on ethnicity and free/reduced lunch status took priority.
4. Finally, an attempt was made to match the number of students in MSP teachers' classes to those in the remaining possible comparison teachers' classes.

The resulting file had 2,363 unique teacher-student pairs, of whom 1,145 had MSP participating teachers and 1,218 had non-MSP participating teachers.

Construction of the teacher level file

The resulting student level file (containing unique teacher-student pairs) was used to construct the teacher level file. The process for constructing this file was relatively straightforward:

1. Unique teacher-student records were aggregated by unique teacher identification number. A teacher level summary variable of average student attendance rate was included in this aggregation. Attendance rate was calculated for each student in the student level file by taking the ratio of days present to days enrolled. Average attendance rate for the students in each teacher's class was simply the mean of these attendance ratios across students with whom that teacher was paired in the student level file.
2. The resulting file had 33 records each corresponding to a single teacher in a single year. Of these 16 were MSP participating teachers and 17 were non-MSP participating teachers. There was one less MSP participating teacher because we were not able to match the single 9th grade math teacher provided in the MSP sample.
3. To this file were added teacher level variables including content knowledge test pre- and post-test scores, self-assessment pre- and post-test scores, years teaching in total (only for MSP), years teaching math (only for MSP), and the contact hours of MSP professional development (MSP only).
4. The resulting file was used as the level two file for all multi-level analyses.

Construction of the school level file

The resulting teacher level file was used to construct the school level file. Because school was found to be a non-significant source of variation in student level outcomes based on an initial 3-

level unconditional model no school specific characteristics were included in this file. The resulting file consisted of 9 records, each corresponding to a different school.