

Comparison of MSP and Non-MSP Schools in Six States

Kenneth K. Wong, Megan Boben, Chisoo Kim and Ted Socha
Brown University

This pilot study proposes a set of analytical steps for comparing schools that participate in the National Science Foundation's Math and Science Partnership (MSP) Program and their intrastate non-participating peers. This pilot is part of a larger effort to evaluate the MSP Program's role in student achievement, along with two companion analyses. While our pilot study uses a comparative approach, the paper by Dimiter Dimitrov (this volume) follows a within-group design. The third analysis by Robert K. Yin (this volume) covers the varied designs used by the MSPs themselves in their own evaluations. As this pilot study has progressed, there has emerged three distinct phases of analysis. Phase I focused on the participating schools within four Math and Science Partnerships (MSPs) located in three states. Phase II expanded Phase I to focus on three more MSPs in three additional states. The Phase III study was conducted using six cohort I MSPs in four states, and will eventually include the nine cohort I MSPs in all six states from Phase I and Phase II. For each phase, the MSP participating schools were carefully matched with the non-participating schools on eight demographic variables to form a comparison group. This paper offers detailed documentation on how we operationalize matching methods for comparative purposes. We conclude that carefully executed matching methods are promising for large-scale comparative analysis on the effects of the MSP Program across all involved states. The study draws on publicly accessible school-level standardized test data from six states and from data available at the National Center for Education Statistics' Common Core of Data (NCES CCD). In addition, the study uses documents available via MSPnet, and Web site information reported by the individual MSPs in the MSP Program accessible through the school year 2005-06.

The purpose of this pilot study is to propose a set of analytical steps for comparing schools that participate in the National Science Foundation's Math and Science Partnership (MSP) Program and their non-participating peers in their respective states.

This pilot is part of a larger effort, along with two companion analyses, to evaluate the MSP Program's role in student achievement. While our pilot study uses a comparative approach, the paper by Dimiter Dimitrov follows a within-group design. The third analysis by Robert K. Yin covers the varied designs used by the MSPs themselves in their own evaluations. The overall objective of this larger effort is to examine whether the Math and Science Partnership (MSP) Program is associated with student academic performance.

As there is increasing national concern about student performance, especially in the areas of mathematics and science, it becomes more important for policymakers and certification agencies to pay attention to programs that achieve results. One area of focus of educational reform has been on ways to improve teacher quality and teacher leadership (Sherrill, 1999). While there has been much research about teacher quality and its impact on student achievement, identifying the components of teacher quality that have the most influence on student achievement has been difficult to establish and measure. Teacher leadership has also been difficult to both define and measure (York-Barr & Duke, 2004). However, there have been several studies that strongly link teacher quality with high student achievement. Rockoff, for example, found that teaching experience is significantly correlated with an increase in test scores (Rockoff, 2004).

Professional development for teachers and in-service training have also been well-researched issues. Professional development has been a main focus of reform initiatives and many states require ongoing professional development for teachers (Garet et al., 2001). While some studies have found little or no impact on student achievement by in-staff development (Jacob & Lefgren, 2004), others have found promising results such as Angrist and Lavy's study (2001) on schools in Jerusalem.

In moving toward a comprehensive analysis of the outcomes associated with the MSP Program, we use state standardized test scores as a measure of student performance due to their public accessibility and prominence as accountability indicators. Ultimately, any conclusions drawn about the relationship between the MSP Program and student achievement will be based on the convergence of all three analyses (ours, Dimitrov's, and Yin's).

The purpose of the analysis is to examine whether MSP-participating schools compared to non-MSP schools are associated with different achievement trends. Because MSP activities primarily involve teacher training and professional development in multiple grade levels, we examine school-level achievement. We address the question: When schools in a state participate in the MSP Program, do their students perform better than they would have if they had not participated in the MSP Program?

In placing the MSP-participating schools in a comparative context, this pilot study uses social science methodologies that account for many confounding conditions, thereby making as fair a comparison as possible. Throughout the analysis, student achievement has been measured in terms of performance on state-administered

assessments in mathematics and science for specific grades in the sampled schools. Generally, we assume the individual states' classification systems of "proficiency" and greater.

Central to addressing the issue of school performance is accounting for a school's previous level of achievement. A simple comparison of MSP-participating to non-MSP schools in a given year does not tell us about the potential association with the MSP engagement because it does not account for how those MSP participating and non-MSP schools were performing before the program began. Therefore, our statistical model account for the prior level of achievement. We also consider factors such as student poverty levels and student family background, due to their documented association with student achievement outcomes.

Three distinct phases of analysis emerged during the pilot study. Phase I focused on the participating schools within four MSPs located in three states. Phase II expanded Phase I to focus on three more MSPs in three additional states. The Phase III study was conducted using six cohort I MSPs in four states, and will eventually include the nine cohort I MSPs in all six states from Phase I and Phase II. For each phase, the MSP-participating schools were carefully matched with the non-participating schools on eight demographic variables to form a comparison group.

Even though we have made a strong attempt to match demographically and academically similar MSP and non-MSP schools, the nature of any MSP-like activities in the non-MSP entities is still unknown. Because the MSP Program was not organized to follow a "treatment" and "no treatment" design, many of the non-MSP schools may very well be undertaking MSP-like activities, using other sources of funding. In fact, the MSP entities in our study are limited to those funded by NSF, and the present analysis has not yet had an opportunity to remove from the non-MSP group those districts and schools that might have received funding from the U.S. Department of Education as part of a counterpart MSP Program supported by that agency.

Future analyses will attempt to further define the non-MSP group more precisely. To the extent that data are available, our next step will differentiate within the non-MSP group those districts and schools known to have some MSP-like activities. Nevertheless, even though such sorting has not yet been possible because of a lack of needed data, the present pilot analysis provides an opportunity for testing the pertinent statistical methods on an otherwise appropriate array of information.

It should also be noted that this study has not made an attempt to research the alignment between any of the MSPs' programs and the resident states' standardized tests. For the purpose of this study we are assuming that the content of each MSP's program is aligned with the standardized test. Additional caveats surrounding the analysis are stated throughout the rest of this paper.

Analytic Design in School Matching

The MSP Program can be seen as an investment toward building the capacity of partnering schools and districts to improve teaching and learning in mathematics and science. The MSP Program has provided for the opportunity to expand the capacity of schools and districts by bringing resources and commitment from institutions of higher education (IHEs) to support mathematics and/or science curricula, teacher professional development, and increases in the supply of highly qualified teachers. Equally important, the Program is designed to build sustainable relationships between these K-12 school systems and other key institutions, including: business and industry, professional organizations, state education agencies, and others with a stake in educational improvement (National Science Foundation, 2005). This pilot study focuses on the relationship between the MSP Program and one set of outcome measures, namely standardized test scores.

Research Model Structure

As emphasized by King, Keohane, and Verba (1994), the goal of social science research is inference. In the present study, we wish to make inferences about the relationship between the MSP Program and concurrent student achievement trends in mathematics and science. Theoretically, we want to look at the performance of an MSP-participating school, and compare it to the counterfactual: ‘How would the school have performed without MSP participation?’ We cannot observe the counterfactual directly, but we use statistical methods designed to estimate the differences associated with MSP participation.

Our analysis includes MSP-participating schools and matched non-MSP schools within the states in which the individual MSPs are located. We match on student background and socioeconomic status variables. Overlooking variables such as these (known as *omitted variable bias*) can lead to incomplete conclusions about the marginal differences associated with the MSP Program. Our study includes a measure of previous school achievement. Including baseline measures of achievement is critical for understanding the incremental difference associated with the MSP Program. It is not sufficient to know how an MSP-participating school is doing this year; we want to know how it is doing this year relative to previous baseline and programmatic years. Finally, our methods attempt to specify the uncertainty surrounding our estimates. Determining statistical significance is important for understanding how strong any MSP and non-MSP differences might be.

Applying Mahalanobis Distance Matching

To control for a number of demographic variables, we employ the Mahalanobis distance matching to define an appropriate comparison school group before analysis

(Gu and Rosenbaum, 1993). We first characterize each MSP school using a set of eight variables and then use the Mahalanobis distance function to locate a “matching” school within the particular state.

The estimated statistical distance between the two N dimensional points is scaled by the statistical variation in each component of the point. For example, if \vec{x} and \vec{y} are two points from the same distribution that has covariance matrix C , then the Mahalanobis distance is given by: $((\vec{x} - \vec{y})' C^{-1} (\vec{x} - \vec{y}))^{\frac{1}{2}}$ (Takeshita, Nozawa, & Kimura, 1993). The resulting group of statistically “close” non-MSP schools is used as our comparison group for regression analysis. Though Mahalanobis distance matching is widely used in computer and spectrometry science, it is only beginning to be used in education policy studies (Good, Burross, & McCaslin, 2005).

Ordinary Least Squares (OLS) Regression

Once the distance matching score is computed, the pilot study employs Ordinary Least Squares (OLS) regression. OLS regressions have, for many years, been the standard statistical technique for evaluation in the field of educational policy (Hanushek, 1979). Along the lines of Hanushek (1986), we assume an education production function. In this model, the outputs of school mathematics and science achievement are seen as the function of a series of inputs. One of the inputs that some schools have is MSP participation, while others do not. Our goal is to see if MSP participation is related to the outputs of mathematics and science student achievement. The general form of the relationship is specified as:

$$[1] \quad \mathbf{O}_{it} = f(\mathbf{M}_{it})$$

where outcomes (\mathbf{O}_{it}) in school i in year t are understood to be a function of the vector of MSP Program activity (\mathbf{M}_{it}). We assume a linear form of the production function (Hanushek, Rivkin, & Taylor, 1996). Our linear estimation initially takes the following form:

$$[2] \quad \mathbf{O}_{it} = \beta_o + \beta \mathbf{M}_{it} + \epsilon_{it}$$

Considering the MSP Participation's Scope and Intensity

Thus far, we have only referred generally to a school's participation in an MSP. In our quantitative analysis, it is necessary to construct measures of MSP Program participation. We considered both the scope and intensity of the MSP participation. First, we used data from the MSP-MIS (Management Information System) to identify the scope of each MSP. By scope, we mean both the subject (mathematics and/or science), as well as the grade levels targeted. In many cases, an MSP's scope does

not align precisely with the grades and subjects tested by state assessments. This can be best seen at the beginning of the MSP Program where many states were in the process of implementing No Child Left Behind (NCLB). At the beginning of the millennium, right after the NCLB Act had been passed, many states only administered their standardized test to a few grades, but added more grades as the years went on, according to NCLB requirements. Also, while many states had a standardized test for mathematics, most states did not test for science since it was not required under NCLB until 2006. Therefore, using the elementary grade span in state E as an example, in school years 2002-03, 2003-04, and 2004-05 we were only able to collect test scores for mathematics in the 4th grade whereas in school years 2005-06 and 2006-07 we could analyze test scores for both mathematics and science in grades 3-5. Similar situations occurred in many of the states we used for our pilot study. This may have an effect on our regression results in that we may not be able to fully capture the effect of the MSP Program in the early years. However, given our method in Phase III to combine several MSPs from different states, along with the fact that some of the states in our analysis do have complete scores for all the years we are looking at, we believe that our results will only be minimally affected.

In addition to scope, we also looked at the intensity of the MSP participation. It should be noted that this pilot study focused on a small set of MSP-MIS data to define the notion of intensity.

These data identified whether or not schools had met one of the following three conditions during school years 2002-03, 2003-04, 2004-05 or 2005-06:

- MSP-MIS item 'q5Bald': Whether 30 percent or more of targeted teachers participated in 30 or more hours of MSP-sponsored activities.
- MSP-MIS item 'q5Bbld': Whether 30 percent or more of targeted students engaged in a challenging mathematics or science curriculum that was initiated or revised with MSP support.
- MSP-MIS item 'q5Bdld': Whether 30 percent or more of targeted students participated in an MSP-supported academic enrichment activity.

For the purposes of this pilot study, if at least one of the three conditions was met for any given year, the school was classified as a 'Participating' MSP school. All other schools that are part of the MSP Program were categorized as 'Partnership' MSP schools. Based on these definitions, the same schools were classified as Participating or Partnership each year, which makes it possible to observe trends in our analysis.¹

The criteria for being classified as a Participating school may seem a little light. The 30 percent targets were chosen as a cut-off point because that was the only participation threshold that was identified in the MSP-MIS data. Currently, we do not have a way of establishing a higher criterion for participation. It may also be argued that we have further increased the looseness our criteria by only requiring Participating schools to meet the 30 percent mark for one category in one year. We classified Participating schools the way we did in order to increase the sample size so we could do a regression

analysis. Increasing the years that a school has to meet the 30 percent mark or the number of categories per year is something we will consider in the future. We focused only on Participating schools for Phases I and II and included Partnership schools in Phase III.

Variables for Statistical Matching

In addition to MSP participation, other school-level conditions are likely to be associated with student achievement. To address alternative explanations, we use relevant and available school-level control variables, as provided by state departments of education and the U.S. Department of Education's National Center for Education Statistics' Common Core of Data (NCES CCD) for the most recent school year available.

Our first control variable is the size of the school, measured as the total enrollment found in the NCES CCD data for the 2005-06 school year. Larger schools operate under different conditions than smaller schools, and in turn, potentially influence student achievement outcomes. Use of "size" as a control variable reduces, if not eliminates, any contaminating effect.

The makeup of the school's student body is likely connected to student performance, which the following five variables seek to address. Schools/campuses serving larger percentages of Black and/or Latino students may experience lower overall achievement as they address the racial disparity that pervades American public education (Jencks & Phillips, 1998), therefore we included a variable for the percentage of population of the student body that is Black and another variable for the percentage of population that is Latino. Another important control is for the percentage of students with disabilities in the school. Larger percentages of students with disabilities may reduce the overall level of school achievement, as those students may face additional educational challenges, so a control variable was included to account for the percentage of the student body with disabilities.

We also included two measures for the percentage of students in the school/campus who are eligible for free and reduced-price lunch, and Title I eligibility. Since the Coleman Report in 1966, a consistent finding in the social science literature on education is a strong relationship between family background and student success. The percentage of free and reduced-price lunch eligible students serves as a proxy for the students' family background, as does the percentage of those eligible for Title I.²

The seventh control variable takes into account the number of pupils per teacher in a given school. Education research suggests that class size reduction can benefit certain populations of students (Rivkin, Hanushek, & Kain, 2005).

Finally, we match on the locale of the school. We believe that the size and classification of the municipality where a school is located has a strong effect on how a school system operates and is structured. This can be directly linked to student achievement. Also, we posit that the type of neighborhood setting where students live

has an influence on student achievement. The NCES CCD's categorical Locale Code variable was used to match non-MSP schools to the average MSP-participating school. The NCES CCD Glossary defines Locale as: "...the school is situated in a particular location relative to populous areas, based on the school's address." The possible categories are "Large City," "Mid-Size City," "Urban Fringe of Large City," "Urban Fringe of Mid-Size City," "Large Town," "Small Town," "Rural, outside [Core Based Statistical Area]," and "Rural, inside [Core Based Statistical Area]."³

This pilot study decided not to include student mobility rates as a control variable due to missing data from schools in the Phase I states. In fact, most states do not publish student mobility data. However, it is probable that this variable is highly correlated to the other eight variables already being used and therefore the characteristic does not need to be accounted for separately.

Measuring Achievement Gains

This study assumes that a connection exists between increasing teaching capacity in mathematics and science and improvement in student performance. We measured improvement in student performance by looking at state-specific standardized tests scores at the school level. These test scores are publicly accessible and allow us to collect multiple points of data over time to monitor trends regarding different schools. Both the direction and the magnitude of student achievement in specific subject areas and by grade levels can be informed by our preliminary analysis. It should be noted that some states will publish preliminary test scores on their Web site before the final scores are ready. In order to ensure the accuracy of our analysis, we only extracted final test scores from each state's Web site and data that were downloaded from the U.S. Department of Education's National Center for Education Statistics' Common Core of Data (NCES CCD) for the purposes of matching MSP schools to non-MSP schools, and all data were accessed at the same time for each phase. That being said, data accuracy is still dependent on each state department of education and the NCES CCD.

In conducting our analysis, we measure the achievement gains from (or value-added by) MSP participation. In the literature that examines the effects of school funding on achievement, this is typically accomplished by modeling, either by generating a dependent variable that measures "change in performance from year $t-1$ to year t " or by using performance in year $t-1$ as a statistical control variable on the right-hand side of the equation (Burtless, 1996). We adopt the second approach, including lagged achievement as a statistical control variable. This lagged achievement variable captures the MSP schools' performance in the previous year, relative to the matched non-MSP schools'. One of the reasons we do not calculate a direct change-in-performance variable is that the test instrument in states may have changed over the time period of interest.

Introducing this notion of value-added through the use of a lagged achievement control variable enables us to better estimate the trends associated with the MSP Program, distinct from influences such as unobserved family background influences (e.g., parental commitment to education). For instance, the assumption holds that if parental involvement is roughly the same year-to-year (e.g., active parents in year t-1 are still active in year t and vice versa), then those parental involvement factors will be captured by the lagged achievement variable. However, if parental involvement also changes from year-to-year (but such data are not available) and systematically with achievement, adjusting for such a contamination would be outside of our model's capability. Overall, given the limitations of the available data, we believe this is the most complete model we can develop.

School Level OLS Model

To perform our ordinary least squares (OLS) analysis, we employ STATA's reg command which regresses a variable on a single predictor (Hamilton, 2006). Our school-level statistical OLS regression model takes the basic form:

$$[3] \quad \mathit{ACHIEVE}_{jt} = \beta_0 + \beta_1 \mathit{ACHIEVE}_{j,t-1} + \beta_2 \mathit{MSP}_{jt} + \epsilon_{jt}$$

where $\mathit{ACHIEVE}_{jt}$ is the mathematics or science student achievement score for school j in year t ; $\mathit{ACHIEVE}_{j,t-1}$ is the school's previous achievement level; MSP_{jt} is a dichotomous (dummy) variable indicating whether or not this is an MSP participating school (after accounting for MSP participation scope and intensity [discussed earlier]); and ϵ_{jt} is the error term. We use this base model for our school-level analysis and can apply this to all grade levels where we have such data.

When performing the regression, the schools are entered as a group (MSP schools or non-MSP schools) and not as paired matches. We did this in order to have a large sample size to perform a meaningful regression analysis.

Combining MSPs Across States

In order to perform a meaningful OLS analysis, there needs to be a large enough sample size. There are not enough schools in each state at each grade-span taking part in the MSP Program to make this possible, so we had to combine MSPs across state lines. Since test scores vary from state to state, this meant we had to find a way to make test scores across many states comparable to each other. We accomplished this by standardizing the test results using Z-scores, which is a common practice. Z-scores are useful for standardizing across states because they measure the number of standard deviations each school is above or below their state mean. The formula for finding a Z-score for each school is as follows:

$$[4] \quad Z_{score} = (X - \mu/\sigma)$$

where X is the percent of students scoring at or above proficiency on the state test for the school; μ is the mean percent of students scoring at or above proficiency on the state test for the state; and σ is the state standard deviation within the population of all schools within the state.

Brief Explanation of Three Phases of Analysis

Phase I

For the first part of this pilot study, which occurred during 2007, we focused on four MSPs within three randomly chosen states (states A-C). The original analysis matched a group of ten non-MSP schools to an “average” MSP school. The thought behind this was that if we found significant effects we could double-check to ensure that other external conditions were not causing any portion of the observed effect. Also, aggregating up to the “average” MSP school would blunt the role of any possible conditions outside of the MSP school observations. Unfortunately, this method reduced the sample size, and although we did run several regression analyses, we did not feel that the results were as meaningful as they could be with a larger sample size. Therefore the method of matching ten schools to an average MSP school was discontinued in favor of one-on-one matching for each MSP school in Phases II and III.

Phase II

The second phase of this study started at the beginning of 2008 and added three MSPs in three states, which were also randomly chosen (states D-F). Phase II matched each MSP school individually to a non-MSP school and so provided us with an increased sample size. Unfortunately, in many cases, the sample size was still not large enough to do a regression. For example, state D has 91 participating schools, but 85 of them are elementary schools, while only 4 are middle schools and 2 are high schools. This discrepancy in sample size led us to perform two different types of analyses: a regression analysis for the groups that had a large sample size, like the elementary schools in state D, and a “trend analysis” for the groups of schools that had a small sample size, like the middle and high schools in state D. The trend analysis simply looked at each individual MSP school compared to its matching non-MSP school to determine if any trend was visible. We did not think it was appropriate to combine grade spans since each grade span is very distinct.

Due to the fact that we had to use different types of analyses depending on sample size, we decided to use only states D-F for this phase in order to determine whether it was worth expanding our analyses to states A-C. Because both our regression and

trend analysis results were inconclusive, we moved on to Phase III without analyzing states A-C in the same manner.

Phase III

In Phase III, we decided to combine the standardized test results from multiple states to increase our sample size sufficiently to be able to perform a meaningful regression analysis for all three grade spans. We combined test results across states C through F by standardizing all the results using Z-scores. We still kept each grade span separate and combined MSP scores within each grade span. In some cases, the same grade could be included in different grade spans. For example, depending on the district, grade 6 can either be included in elementary school or in middle school. Since the way 6th grade is taught can be vastly different if it's part of an elementary school as opposed to a middle school, we decided just to keep each school in the grade span the state defined it to be in. This means that sometimes grade 6 is included in the elementary school results and sometimes it's included in the middle school results. This also occurred for grade 5, and for grade 9, which is sometimes taught in middle school.

Matching of Comparison Schools with MSP-Participating Schools

The Phase III study was first conducted using six cohort I MSPs in four states, and will eventually include the nine cohort I MSPs in all six states. The states were chosen at random and analyzed in various combinations depending on the phase of the analysis. Table 1 shows which states were included in each phase.

Table 1
States Included in Each Phase

	Phase I	Phase II	Phase III
State A	X		
State B	X		
State C	X		X
State D		X	X
State E		X	X
State F		X	X

In sorting out the MSPs for Phase I, we looked at each MSP in each state to see if the MSPs were part of Cohort I (awarded in 2002-03) and contained Participating schools (state C was also analyzed in Phase III). State A has multiple operating MSPs (three in its case). One of the three MSPs located within the state is a Cohort II grantee (awarded in 2003-04), so it was dropped from the analysis because it is desirable to have at least three years of student achievement data to analyze: one year as a baseline

and two subsequent project years. State B has one operating MSP that focuses on mathematics instruction improvement at all levels of compulsory schooling. State C has three operating Cohort I MSPs with one focusing on mathematics instruction at all levels of compulsory schooling, one focusing on both mathematics and science instruction at all levels of schooling and, one targeting mathematics middle and high school teacher development.

The same categorization of MSPs was done for the second group of states analyzed in Phases II & III. State D has one operating MSP that is a targeted mathematics initiative. State E has one operating MSP that only focuses on mathematics instruction improvement at the elementary and middle school levels. State F has two operating MSPs, but one is a Cohort II grantee so it was dropped. Also, due to some issues, the Cohort I grantee was dropped after the 2005-06 school year. Since we randomly chose the states we decided to keep the state as part of the study.

According to the MSP-MIS data, and using the definition of “intensity” of MSP participation as previously discussed, the following table, Table 2, shows the break down of Participating (schools that meet at least one of the three 30% criteria for

Table 2
Number of Participating and Partnership Schools in Each MSP

Summary of MSPs Included in Pilot Study					
ID	Total # Schools	# Participating Schools	% Participating Schools	# Partnership Schools	% Partnership Schools
State A: MSP (1)	40	30	76.6	10	23.4
State A: MSP (2)	81	0	2.4	81	97.6
State B: MSP (3)	24	23	89.3	1	10.7
State C: MSP (4)	50	39	79.2	11	20.8
State C: MSP (5)	73	11	15.1	62	84.9
State C: MSP (6)	224	72	31.9	152	68.1
State D: MSP (7)	124	90	71.4	34	28.6
State E: MSP (8)	39	12	30.8	26	69.2
State F: MSP (9)	112	109	97.6	3	2.4
Pilot Study Total	770	390	50.6	380	49.4
Program Total	6,020	3,181	52.8	2,839	47.2

any given year) and Partnership schools (all other schools that are a part of the MSP program) in each of the MSPs in our study.

Once the states were selected and the Participating schools were sorted from the Partnership schools, the following characteristics listed in Table 3 were used to match MSP schools to non-MSP schools:

Table 3

Mahalanobis Matching Variables

Matching Variables		
Variable	Source	Abbreviation
Total Enrollment	NCES CCD	ENROLL
Percent Black/ African-American Students	NCES CCD*	PCT_BLK
Percent Latino Students	NCES CCD*	PCT_LATN
Percent Students with Disabilities**	State Data	PCT_SWD
School Title I Eligibility	NCES CCD	TITLE1
% Students Free/ Reduced Lunch Eligible	NCES CCD*	PCT_LUNCH
Students to Teacher Ratio	NCES CCD	PPT
Locale of School	NCES CCD	LOCALE

*Percent calculated from a raw number

**Not available in states B or C

Once the characteristics have been collected, we match the entire universe of each state’s traditional public non-MSP schools to the MSP schools using Mahalanobis distance matching. This methodology calculates a distance between each MSP school and non-MSP school signifying how well the non-MSP school matches the MSP school profile. In short, the Mahalanobis method does this by determining the range between the minimum and maximum values for each variable and then summing-up all of those values. That number becomes the maximum distance value. So, a perfect match would have a distance value of 0 and the farthest observation from the average MSP school profile would have a distance value equal to the maximum distance value.

Results

Originally for States A-C in Phase I, we performed a regression analysis for each grade level for each MSP in spite of the small sample size. For the analyses, we used school year 2002-03 as the baseline and tracked student performance on mathematics and science state standardized tests through school year 2005-06. A total of 73 comparative regressions were run and out of all the comparisons, only two demonstrated statistically significant results. The findings are presented in Table 4.

Table 4

Phase I Results (Participating Schools)

Phase I						
School Year	State A		State B		State C	
	Mathematics	Science	Mathematics	Science	Mathematics	Science
2003-2004	8.00% change (99% CI) # students proficient in grade 11	n.s.	n.s.	n.s.	n.s.	n.s.
2004-2005	n.s.	n.s.	n.s.	n.s.	19.06% change (95% CI) # students proficient in grade 5	n.s.
2005-2006	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

The table shows two statistically significant differences and no clear trend over time. Given the large number of regressions that were run, we consider these two as being chance occurrences, especially because one result was positive and the other was negative.

For Phase II, we looked to increase the sample size by matching each MSP school to a non-MSP school instead of matching an “average” MSP school to ten non-MSP schools as in Phase I. We also looked at grade spans (elementary, middle, and high school) instead of the individual grade levels. For each of the grade spans, we only considered those in which the MSP activities were concentrated. Unfortunately, we still encountered many instances where the sample size was too small. For example, there are only 4 participating middle schools and 2 participating high schools in State D, and in State E there are only 10 participating schools total – 5 at the elementary school level and 5 at the middle school level. Given the situation, we decided to only perform regression analyses for the MSP groups that had a large enough sample size and do a trend analysis for the MSP groups that had a smaller sample size.

Following this plan, we did a regression analysis on the elementary schools in State D (85 schools) and the elementary schools in State F (74 schools). Like Phase I, for these analyses we used school year 2002-03 as the baseline. However, because Phase II occurred many months after Phase I, we were able to include an additional year of test scores and tracked student performance on mathematics and science state standardized tests through school year 2006-07. We used the Mahalanobis analysis to find the best

match for each MSP school on the eight variables that were discussed earlier in the paper and then all the MSP schools were regressed against their matching non-MSP schools. The results from this analysis in most cases were not statistically significant, and the ones that were statistically significant showed very little effect from the MSP treatment. Table 5 documents the findings.

Table 5
Phase II Results (Participating Schools)

Phase II				
School Year	State D		State F	
	Mathematics	Science	Mathematics	Science
	Elementary		Elementary	
2003-2004	n.s.	n.s.	n.s.	n.s.
2004-2005	2.584 (<i>SD</i> = 1.255, <i>p</i> < .05)	n.s.	n.s.	n.s.
2005-2006	n.s.	n.s.	n.s.	n.s.
2006-2007	n.s.	n.s.	n.s.	n.s.

The above results again do not show any meaningful results, even though one of the comparisons was statistically significant.

We conjectured that one of the reasons that most of the results did not show an effect may be because we were matching on too many variables. Some of the variables are often highly correlated and may be interfering with the results. In order to test this theory, we redid the Mahalanobis matching for the elementary schools in State D and only matched on two variables: free- or reduced-price lunch and locale. The results from this analysis showed no significant results. Therefore, we put aside this line of thinking and went back to matching on all eight variables.

For the MSP groups that have a small sample size we did a trend analysis comparing MSP schools to the best-matched non-MSP school. The following table shows this analysis for State E.

The trend analysis was not conclusive either, in any state or grade span. We do not show the results from the other trend analyses since they are fairly similar to what is shown above for state E.

In an effort to have a large enough sample size to be able to do a regression analysis for each grade-span, we normalized the percent of students scoring at or above proficient across States C-F, using a Z-score. We felt it was necessary to keep the grade-spans separated (elementary, middle and high school) not only because of the differences of the grade-spans but also because of the differences of program implementation at each grade-span.

Table 6
One-on-one Matching Trend Analysis Results for State E

School	Grade	Percent of Students Scoring At or Above Proficient						Match Score
		2002	2003	2004	2005	2006*	2007*	
MSP Primary School A	4	N/A	N/A	N/A	43.6	37.8	43.8	4.722
Match Primary School A	4	N/A	N/A	N/A	27.5	37.8	21.5	
MSP Primary School B	4	33.9	62.5	41.5	58.0	6.0	21.4	0.304
Match Primary School B	4	16.7	35.3	36.5	25.5	16.8	25.8	
MSP Primary School C	4	43.9	47.8	43.4	54.3	46.0	59.2	2.211
Match Primary School C	4	60.0	56.3	59.6	82.9	84.6	79.5	
MSP Primary School D	4	33.3	33.8	49.2	45.7	60.9	43.6	0.888
Match Primary School D	4	23.3	25.0	21.3	36.7	23.5	29.7	
MSP Primary School E	4	25.8	23.3	48.3	44.2	19.5	27.7	2.16
Match Primary School E	4	7.8	5.4	19.1	18.2	15.3	14.9	
MSP Middle School A	8	1.6	3.4	1.9	0.5	5.5	11.2	11.809
Match Middle School A	8	36.1	4.0	13.6	28.6	11.5	8.7	
MSP Middle School B	8	3.0	7.5	3.5	7.7	14.8	23.3	9.415
Match Middle School B	8	25.7	19.6	22.6	33.7	60.8	67.2	
MSP Middle School C	8	26.1	39.0	34.9	32.9	58.6	60.8	0.799
Match Middle School C	8	28.0	20.1	27.1	24.7	60.7	65.5	
MSP Middle School D	8	7.7	3.0	3.3	4.9	22.5	22.9	10.016
Match Middle School D	8	12.8	12.1	5.9	12.9	29.2	38.1	
MSP Middle School E	6	N/A	N/A	N/A	N/A	65.6	65.0	5.304
Match Middle School E	6	N/A	N/A	N/A	N/A	25.0	23.9	

Note. *Test changed in 2005-06 School Year so scores are not comparable to previous years.

Another change we made in Phase III was to examine Partnership schools along with Participating schools. Currently, as a trial, we have only looked at Partnership schools in States C and E. We are in the process of adding Partnership schools in States D and F to our analysis. We looked at both Partnership and Participating schools by themselves when compared to their matched non-MSP schools, as well as Partnership and Participating schools combined, labeled as “MSP” schools.

Tables 7.1, 7.2, and 7.3 show the results from the combined analysis. Looking at the following three tables (7.1-7.3), you will notice that we did not analyze science test scores when we combined just States C and E. This is because State C did not start testing for science until the 2006-07 school year, which means we did not have a previous year’s score to include in our regression. Also, as mentioned above, State E only has a few Participating and Partnership schools involved in the MSP program, so the sample size is not big enough, not to mention that they did not start testing for science until the 2005-06 school year, which, even if there were enough schools to perform a regression, would only give us one year of data. The lack of science assessment has to do with the fact that under No Child Left Behind (NCLB), science assessments did not have to be developed until the 2005-06 school year and states were not required to use the assessments until the 2007-08 school year. This is not to say that this analysis is part of the No Child Left Behind Act, but that our access to test scores is linked to NCLB because of the mandate in the Act for states to perform assessments.

For the high school, since we only used States C and E when examining Partnership schools, we were unfortunately not able to get any results. This is due to the fact that in State C, there are only two Partnership schools and State E did not start standardized testing in mathematics until the 2006-07 school year, which means we did not have a previous year’s scores to use in our regression analysis.

Looking over the results, Tables 7.1-7.3 again show no particular differences between the MSP and non-MSP schools. There are no statistically significant differences at the elementary and high school levels, and only one significant difference at the middle school level. Furthermore, the single significant difference occurs in an early year of the MSP program. One possibility for this single difference is that schools struggling with mathematics or science may have been chosen to be part of the program in order to give those schools increased assistance to catch-up to the state standard, and the early year still showed this lagging baseline condition.

Overall, and across all three Phases, the absence of nearly any differences between the MSP and non-MSP schools may be a reflection of three conditions that defined the current research design. First, the extent of MSP intervention in the MSP schools was only weakly defined, and the non-MSP schools may have been equally or even more engaged in MSP-like activities, but not funded by the MSP Program. For instance, the “Participating” schools only had to meet one of three “30 percent” criteria, and for only one of the four monitored years in Tables 7.1-7.3.

Table 7.1
 Results for Phase III – Elementary Schools

School Year	<u>Combined States C, D, E & F</u>		<u>Combined States C & E</u>		
	Mathematics	Science	Mathematics		
	Particip	Particip	Partner	Particip	MSP
2003-2004	0.392 (<i>SD</i> = 0.396)	N/A	0.174 (<i>SD</i> = 0.192)	0.392 (<i>SD</i> = 0.396)	0.211 (<i>SD</i> = 0.179)
2004-2005	0.023 (<i>SD</i> = 0.055)	-0.046 (<i>SD</i> = 0.114)	-0.313 (<i>SD</i> = 0.222)	-0.09 (<i>SD</i> = 0.425)	-0.241 (<i>SD</i> = 0.199)
2005-2006	0.041 (<i>SD</i> = 0.059)	0.125 (<i>SD</i> = 0.079)	-0.011 (<i>SD</i> = 0.183)	-0.506 (<i>SD</i> = 0.529)	-0.078 (<i>SD</i> = 0.170)
2006-2007	0.035 (<i>SD</i> = 0.062)	-0.09 (<i>SD</i> = 0.104)	0.015 (<i>SD</i> = 0.085)	-0.238 (<i>SD</i> = 0.312)	-0.022 (<i>SD</i> = 0.084)
	<i>N</i> = 199	<i>N</i> = 199	<i>N</i> = 192	<i>N</i> = 32	<i>N</i> = 224

Note. “Particip” is the effect of Participating Schools compared to non-MSP Schools; “Partner” is the effect of Partnering Schools compared to non-MSP Schools; “MSP” is the effect of MSP Schools (Partnering combined with Participating) compared to non-MSP Schools.

Second, the comparisons were made as a series of annual comparisons, rather than calculating a single multi-year trend for each school and then comparing the two groups of schools. Such multi-year trends may yet reveal differences between the two groups, and therefore the estimation of such trends is among our ongoing research priorities.

Third, the “light” participation definition and the series of annual comparisons has a potential and undesirable interaction: the year of comparison between a school and its matched counterpart may have been a year in which the MSP school did not achieve even the 30 percent criterion. As more states and MSPs are added to the Phase III analysis, the sample sizes will increase so that this interaction can be avoided in the future. For instance, the criterion for a “participating” school might be made more stringent so that a school has to show 30 percent participation in at least two if not all of the monitored years in order to be defined as a participating school.

Table 7.2
Results for Phase III – Middle Schools

School Year	<u>Combined States C, D, E & F</u>		<u>Combined States C & E</u>		
	Mathematics	Science	Mathematics		
	Particip	Particip	Partner	Particip	MSP
2003-2004	-0.234 (SD = 0.117)	N/A	-0.194 (SD = 0.936)	-0.234 (SD = 0.117)	-0.291 (SD = 0.113, <i>p</i> < 5%)
2004-2005	0.044 (SD = 0.121)	-0.045 (SD = 0.138)	-1.612 (SD = 1.323)	-0.186 (SD = 0.117)	-0.092 (SD = 0.149)
2005-2006	0.058 (SD = 0.140)	0.11 (SD = 0.114)	0.516 (SD = 0.300)	0.006 (SD = 0.158)	0.113 (SD = 0.135)
2006-2007	-0.161 (SD = 0.087)	-0.323 (SD = 0.161)	-0.063 (SD = 0.187)	-0.139 (SD = 0.085)	-0.118 (SD = 0.069)
	<i>N</i> = 77	<i>N</i> = 77	<i>N</i> = 37	<i>N</i> = 58	<i>N</i> = 95

Note. “Partner” is the effect of Partnering Schools compared to non-MSP Schools; “Particip” is the effect of Participating Schools compared to non-MSP Schools; “MSP” is the effect of MSP Schools (Partnering combined with Participating) compared to non-MSP Schools

Implications, Lessons Learned, and Future Work

In this study, we focus on a sample of MSPs participating in six states. The MSP schools were carefully matched with non-MSP schools on eight demographic variables to form a comparison group. This paper offers detailed documentation on how we operationalize two matching methods for comparative purposes. This is compliant with the U.S. Department of Education’s Academic Competitive Council’s (ACC) charge to evaluate the effectiveness of STEM education interventions under rigorous conditions. In a hierarchy with “Experimental Methods such as Randomized Controlled Trials (RCTs)” at the top and “Other designs, such as Pre- and Post-Test Studies, and Comparison Group Studies without careful matching” at the bottom, our matching methodology falls in between as one that is a “Quasi-experimental

Table 7.3

Results for Phase III – High Schools

School Year	Combined States C, D, E & F	
	Mathematics	Science
	Particip	Particip
2004-2005	0.068 (SD = 0.134)	-0.204 (SD = 1.400)
2005-2006	0.12 (SD = 0.142)	-0.03 (SD = 0.003)
2006-2007	0.059 (SD = 0.200)	0.054 (SD = 0.000)
	N = 52	N = 52

Note. “Particip” is the effect of Participating Schools compared to non-MSP schools.

Method such as Well-Matched Comparison Group Study.” In the absence of having the prime condition of being able to conduct a randomized controlled trial of MSP-funded schools, we will continue to refine our matching methodology to provide the most appropriate quasi-experimental method so that it may act as a model for similar program analyses.

As observed in this pilot study, we are unable to find any effect in the statistically significant results. Then again, this outcome is from combining six MSPs out of a possible twenty-two (22) Cohort I MSPs. We think this could change if we added more MSPs and increased the standard for what it means to be classified as a Participating school. We also need to look into discovering which of our non-MSP schools have MSP-like activities and to make them into their own category of school.

Our matching results suggest that carefully executed matching methods are promising for large-scale comparative analysis on the effects of the MSP Program across different states. Our next step is to do a closer look at the states we’ve already analyzed to see if there’s an effect we’re missing, and expand the methods to include additional states with operating Cohort I MSPs and additional years of data. We will go back to try and take into account other mathematics and science programs that other schools may be participating in that may be contributing to an evident lack of effect from the MSP. By doing this, we may actually create four categories of schools in each state: 1) participating in the MSP, 2) a partnership school in the MSP, 3) participating in another mathematics/science partnership outside of NSF, and 4) no extra outside mathematics/science help. We will also consider looking at each classification of passing on the state standardized tests, such as “below basic,” “basic,” “proficient,” and “above proficient.” It may be that by not looking at each category separately we are missing large gains in individual classifications, like large groups of students

moving from proficient to above proficient, or even many students moving from below basic to basic.

Ultimately, the goal is to analyze the relationship between MSP school participation and state standardized achievement test gains. We shall do this through our matching methodology, which controls for various extraneous factors that may affect student test scores. We hope that our effort in defining and operationalizing an appropriate comparison group in the MSP program evaluation will contribute to a broader discussion in program evaluation.

Endnotes

1. It should be noted that Dimitrov's study only uses the condition that 30 percent or more of a school's targeted teachers participated in 30 or more hours of MSP-sponsored activities.

2. Title I was established by the Elementary and Secondary Education Act of 1965 in order to "distribute funding to schools and school districts with a high percentage of students from low-income families." In order to qualify for Title I funding at least 40% of the students must come from low-income families as defined by the U.S. Census. (Encyclopedia Britannica, "Elementary and Secondary Education Act." <http://www.britannica.com/EBchecked/topic/184196/Elementary-and-Secondary-Education-Act>, Accessed October 20, 2008.

3. See <http://nces.ed.gov/ccd/commonfiles/glossary.asp>

Acknowledgments

This article is one in a series of studies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's Math and Science Partnership Program (NSF MSP). The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the NSF Program Officer.

The MSP-PE is led by COSMOS Corporation. Robert K. Yin (COSMOS) serves as Principal Investigator (PI) and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators are Patricia Moyer-Packenham (Utah State University) and Kenneth Wong (Brown University).

The authors would like to express gratitude to the Brown University research team consisting of Pamela Denton, Ashley Weaver, and Melanie Duch who contributed to this project.

Any opinions, findings, conclusions, and recommendations expressed in this article are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Angrist, J. D., & Lavy, V. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools. *Journal of Labor Economics*, 19, 343-369.
- Burtless, G. (Ed.). (1996). *Does Money Matter? The effect of school resources on student achievement and adult success*. Washington, DC: Brookings Institution Press.
- Coleman, J. S. (1966). *Equality of educational opportunity (COLEMAN) study (EEOS)*. Retrieved from U.S. Department of Education Web site: <http://dx.doi.org/10.3886/ICPSR06389>
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38, 915-945.
- Good, T. L., Burross, H. L., & McCaslin, M. M. (2005). Comprehensive school reform: A longitudinal study of school improvement in one state. *Teachers College Record*, 107, 2205-2226.
- Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2, 405.
- Hamilton, L. C. (2006). *Statistics with STATA* (6th ed.). Pacific Grove, CA: Duxbury Press.
- Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, Summer, 351-388.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, September, 1141-1177.
- Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *The Review of Economics and Statistics*, 78, 611-627.
- Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, 39, 50-79.
- Jencks, C., & Phillips, M. (Eds.) (1998). *The black-white test score gap*. Washington, DC: Brookings Institution Press.
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- National Center for Education Statistics, Common Core of Data (CCD). (n.d.). Retrieved from <http://nces.ed.gov/ccd/>
- The National Science Foundation. (2005). *Math and science partnership program: Strengthening America by advancing academic achievement in mathematics and science* (NSF-05-069). Retrieved from <http://www.nsf.gov/pubs/2005/nsf05069/nsf05069.pdf>

- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417-458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, *94*, 247-252.
- Sherrill, J. A. (1999). Preparing teachers for leadership roles in the 21st century. *Theory into Practice*, *38*, 56-61.
- Takeshita, T., Nozawa, S., & Kimura, F. (1993). On the bias of mahalanobis distance due to limited sample size effect. *Second Conference on the Institute of Electrical and Electronics Engineers*, Tsukuba Science City, Japan.
- York-Barr, J., & Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research*, *74*, 255-316.