

## **Measuring the Student Learning Effects of Math and Science Partnership Programs: The Appalachian Program**

### **Abstract**

This study evaluates the effects of a National Science Foundation (NSF) initiative to enhance math and science learning outcomes. The Appalachian Math and Science Partnership (AMSP) is one of the major programs funded by NSF under this initiative and we evaluate its effects on student learning. Our results show that teacher participation in AMSP is beginning to have independent, positive effects on student learning. After controlling for other variables that affect student scores, our measures of AMSP participation suggest that more hours of training and a greater share of teachers participating with the AMSP partnership program are leading to middle and high school math and science score gains. At this point, the gains are not found for elementary schools.

Keywords: teacher quality, professional development, math and science learning

## I. Introduction

Academics and policymakers increasingly emphasize that a key to improving educational outcomes in the United States lies in enhancing the quality of teachers. After many years of research that failed to show significant and systematic effects of inputs such as school expenditures, teacher salaries, or pupil-teacher ratios on student outcomes, recent work illustrates that the quality of teachers is a significant factor in explaining student achievement (Rivkin, Hanushek, and Kain, 2005). For those individuals already in the teaching world, one of the hallmarks of this new focus on teacher quality is professional development (PD) programs. While different methods of recruitment and training of teachers may affect the quality of teachers in the future, professional development programs focus on changing teacher quality and student achievement of those teachers currently in the classroom.<sup>1</sup> Unlike the growing literature on the relationship between pre-service training and student achievement (Acevedo, forthcoming; Harris and Sass, 2006), the effectiveness of professional development in terms of student learning has received less attention.

This paper examines the learning effects of a National Science Foundation-sponsored PD initiative -- the Appalachian Math and Science Partnership (AMSP). The math and science partnerships (MSPs) aim to increase the academic achievement of students in mathematics and science by enhancing the content knowledge and teaching skills of classroom teachers. We describe a model to estimate the effects of AMSP participation on student outcomes. Estimation of PD effects is complicated by the fact that, similar to many PD programs, teachers and schools voluntarily choose to participate in AMSP. We correct for the nonrandom nature of selection of schools and teachers into

---

<sup>1</sup> National Board Certification for in-service teachers is also an alternative route for enhancing teacher quality. See Goldhaber and Anthony (2007) for a description of the program and its effects on teacher quality.

AMSP and then estimate PD effects while controlling for other independent variables that would be expected to influence test scores. Our results suggest that AMSP is producing positive learning effects for math and science students although not at all grade levels. Our results also show that the lowest performing schools are those whose teachers choose to participate in AMSP. We do not evaluate the cost effectiveness of the program.

## **II. Professional Development Programs**

There is a large and evolving literature on understanding pre-service factors that contribute to higher quality teachers. While we know that teacher quality is important, we also are learning that measurable attributes of teacher quality remain hard to define. For example, years of experience, the quality of the teachers' institution of training as well as the teachers' type of certification (traditional or alternative) have been found to have some effect on student outcomes and also have been found to have no effect (Acevedo, forthcoming; Clotfelter, Ladd, and Vigdor, 2006, 2007; Darling-Hammond and Thoreson, 2001, 2002; Goldhaber and Anthony, 2007).

Of equally great interest to policymakers and certification agencies are the in-service training activities for teachers. Indeed, a common theme in education reform is support for professional development (Fullan and Hargreaves, 1996). Many state certification programs require ongoing professional development under the assumption that it contributes to teaching effectiveness. Among the recent studies attempting to evaluate in-service training are Angrist and Lavy (2001) who found positive effects of teacher professional development in non-religious schools using data in Jerusalem. They were able to conduct a natural experiment study because a "natural" experiment stemmed from a sudden infusion of money into professional development for teachers. On the other hand, Jacob and Lefgren (2002), who looked at Chicago schools, found no effects

from professional development programs. Most recently, Harris and Sass (2006) found positive effects of content-oriented teacher professional development on student outcomes in Florida but only two or three years after the actual intervention.<sup>2</sup>

In summary, the literature tells us quite clearly that teacher quality is an important ingredient in student learning outcomes. At this point in time, however, the literature is less clear on exactly how we create higher teacher quality through training programs either at the pre-service or in-service stages of teacher development. The literature also illustrates that one of the real challenges of identifying programs that might enhance teacher quality is the myriad of complicating factors that enter into estimating the effects of a particular attribute or a particular aspect of teacher quality. The remainder of this paper contributes to this literature as we estimate the effects of a teacher PD program in which schools and teachers voluntarily choose to participate.

### **III. The AMSP Program**

In the early 2000s, the National Science Foundation initiated a competitive funding program geared toward the teachers of math and science. Programs receiving funding developed partnerships between mathematicians and scientists at universities, in the community, and the K-12 teachers. While there are over 40 targeted and comprehensive funded partnerships around the country, one of the largest of these is the AMSP. The AMSP received a five year grant of \$22.5 million from the National Science Foundation (NSF) and began implementation in 2002-03. AMSP is a partnership among 38 Central and Eastern Kentucky school districts, 9 Tennessee school districts, and 5 Virginia (western part of the state) school districts, the Kentucky Science and

---

<sup>2</sup> Harris and Sass (2006) provide an excellent review of the literature on the value-added effects of professional development.

Technology Corporation, and 10 higher education institutions located in these three states. AMSP seeks to demonstrate improved student achievement in mathematics and science in the Central Appalachian region through the support of partnerships that unite the efforts of teachers, administrators, guidance counselors, and parents in local schools with administrators and faculty at area colleges and universities. AMSP provides professional development teacher training to teachers of mathematics and science through a partnership with university faculty in the same region.

There is a long history of an educational achievement gap in the Central Appalachian region. This gap is especially apparent in the areas of science and mathematics. Few students in the Central Appalachian region score at the proficiency level or above in mathematics and/or science as defined by the assessment standards developed in each state. Analysis of the assessment data consistently reveals lower performance at all K-12 levels for central Appalachian students when compared to the state averages and/or students from more affluent regions of the states. Enrollment in higher level mathematics courses, including Algebra II and calculus, is less than one-third the enrollment in lower level mathematics. Introductory science courses have more than three times the enrollment of higher-level courses such as chemistry and physics. Although 25 of the central Appalachian school districts report offering some type of AP or dual credit program, enrollment in these programs is non-existent in many schools.

The AMSP program was funded and developed on the assumption that this achievement gap may exist, in part, because teachers in Central Appalachia are less prepared to teach math and science than teachers in other areas. Demand for teachers often exceeds supply, especially in low performing schools such as those in Central Appalachia. These school districts report a major problem with attracting and maintaining a staff of highly qualified mathematics and science teachers.

Whether the MSPs generally, and AMSP in particular, are addressing the perceived deficiencies in teacher quality and enhancing student learning is only beginning to be evaluated. AMSP is currently in its fifth year of the grant and at a point where learning effects should begin to be revealed. The remainder of this paper describes a data set and a methodological framework for empirically estimating the effects of AMSP on Kentucky school-level outcomes.<sup>3</sup>

#### **IV. Data and Model**

All Kentucky public elementary and secondary schools were selected for the study. Tennessee, and West Virginia schools participate in AMSP but are not included in this evaluation because of a lack of comparable standardized test scores or outcomes data.<sup>4</sup> Schools serve as the unit of analysis.<sup>5</sup> Ideally, we would use student level data but student-teacher matching is not currently available at the state level.<sup>6</sup> School level data allow us to correct for unobservable school specific effects.

The remainder of this paper will provide an analysis of the AMSP program using the type of school level data typically available across the states as part of the reporting requirements of *No Child Left Behind* (NCLB). In the Kentucky case, math and science achievement data (Kentucky Core Content Tests) are used by the state for NCLB and are used as the accountability standards set by the Kentucky Education Reform Act of 1991. Data were collected from the Educational Professional Standards Board, the Kentucky Department of Education, and the Appalachian Math and Science Partnership. The data

---

<sup>3</sup> Ideally, the evaluation would test the effects of AMSP in the three participating states. There are not comparable achievement data available at this point in time to allow such evaluation.

<sup>4</sup> There are 56 school districts that participate in AMSP in central Appalachia and a majority (38) of the districts are located in Kentucky.

<sup>5</sup> Missing demographic data deprives us of roughly 20 percent of the elementary school observations over time. We lose 23 percent of the middle school and 10 percent of the high school observations.

<sup>6</sup> The Harris and Sass (2006) paper provides an excellent case analysis of the methodological issues that are avoided by using individual student data match to individual teacher data. Current data collections are underway that will provide a sample of student-teacher matching for a future AMSP evaluations.

collected for this study are for six school years, from 2000-2001 through 2005-2006.<sup>7</sup> This includes two years of data prior to the onset of AMSP and four-years post the implementation of AMSP. Table 1 provides a descriptive summary of the data.<sup>8</sup>

**Table 1: Descriptive Data**

Variable	Mean	Min	Max
Average Math Score	68.70743	13	126.262
Average Science Score	77.16197	26.1307	137.648
Average Reading Score	82.43509	39.7781	130.588
Average Academic Index	73.06784	33.4	116.9
Change in Math Scores 2000-2002	1.69583	-27.2707	64.1017
Change in Science Scores 2000-2002	1.688693	-31.4811	71.1786
Change in Reading Scores 2000-2002	0.9665285	-33.4342	80.4785
ARSI participation	0.0545125	0	1
% Free or Reduced Price Lunch	51.58593	0	100
% African American	8.841994	0	100
% Hispanic	0.9452214	0	66.66667
% Asian	0.5219645	0	75
Number of Teachers	32.84348	1	128

Because AMSP is a partnership between universities and schools, participation in the program is voluntary. At the onset of the professional development program, schools were invited to participate. School superintendents and principals made the initial decision in 2002-03 about whether or not to participate in AMSP in general but individual teachers within those schools later decide whether they want to participate and how much to participate in particular training programs.

<sup>7</sup> The Kentucky Department of Education (KDE) changed its CATS test in 2007 and the scores are not comparable to earlier years.

<sup>8</sup> An alternative estimating method would be the use of propensity scoring as a means of choosing the “control” schools for comparing the AMSP effects. Some recent literature illustrates that propensity scoring does not eliminate estimation biases (Wilde and Hollister, 2007). We did restrict our sample to schools in the eastern portion of the state as a robustness test and the results do not differ qualitatively.

Normally, to estimate the effects of an intervention, we compare outcome data pre- the invention to that of post- the intervention, controlling for other factors that are expected to affect the outcome. The estimated coefficient on the intervention variable illustrates the magnitude of the effects of the program. But underlying this estimation procedure is the assumption that participation in the intervention is part of a random assignment. In other words, there are no biases that would determine who is part of the treatment and who is part of the control group. We will estimate this type of model for comparison purposes, but we suspect the estimates not be consistent because it is unlikely that teacher assignment in the treatment group of AMSP follows the steps necessary for these normal, random assignment assumptions.

Recognizing that participation in this case is likely to be nonrandom, we also will estimate the effects of AMSP while simultaneously estimating participation as an endogenous variable in the model. We hypothesize that some schools are more likely to participate (i.e., be part of the treatment group) than others. For example, we expect that low performing schools are more likely to participate in professional development programs than high performing schools because the payoffs to these schools for participation will be higher.<sup>9</sup> If this participation effect is as we predict, failure to correct for the nonrandom participation will bias any possible estimated professional development coefficient downward.<sup>10</sup> The correction technique will allow us to interpret the estimated PD results as if all schools in the state were given equal probability of participating in the professional development activities.

---

<sup>9</sup> This is especially true in Kentucky where failing schools can be taken over by the State Department of Education.

<sup>10</sup> We tested the exogeneity of the participation variables by calculating the correlation coefficients for the participation variables and the residuals from the second stage regressions and then testing their significance. If participation is exogenous, then the participation measures should be uncorrelated (or at least insignificantly correlated) with the residuals from the second stage regressions. We found that the test statistics for the correlation coefficients ranged from 6.21 to 12.32. This is strong evidence that the decision to participate in AMSP is endogenous and that the OLS estimates are biased.

The first stage instrumental variable equation of our two-stage model is as follows:

$$AMSP_{it} = \alpha_0 R_{it} + \alpha_1 F_{it} + \alpha_2 T_t + \alpha_4 \Delta A_{ipre} + \varepsilon_{it} \quad (1)$$

where  $AMSP_{it}$  is a measure of participation in the professional development intervention in school  $i$  in year  $t$ ;  $R_{it}$  represents characteristics of school  $i$  in year  $t$ ;  $F_{it}$  is socioeconomic characteristics of students and their families in school  $i$  in year  $t$ ;  $T_t$  is a vector of school year fixed effects;  $\Delta A_{ipre}$  is the change in performance of the school between the two years preceding the introduction of AMSP. The change in test scores two years prior to AMSP was used to insure exogeneity. Included in the  $R$ -vector is the location of the school (i.e., whether it is located in Appalachia)<sup>11</sup> and whether the school participated in a previous NSF-funded program, the Appalachian Rural Systemic Initiative (ARSI).

These variables are intended to capture the motivation of the schools and their teachers to participate in a professional development program aimed at increasing test scores of students. Presumably, school administrators will be more likely to join AMSP and encourage teacher participation if the schools are low performing. But, perhaps unmeasured attributes such as motivation or effort can be captured by the direction of change in test scores prior to AMSP. Thus, the change in scores is included as an additional independent variable.<sup>12</sup> The measures of AMSP participation will have lower

---

<sup>11</sup> There are some non-Appalachian schools in AMSP.

<sup>12</sup> There are actually several test scores for schools that may be used as instruments for estimating purposes. Each school also has scores for specific content areas including math, science, and reading. Because different levels of schooling may focus on different content areas, we experimented with the various areas and used the content area as an instrument that provided the best fit. For elementary schools, the change in science scores from the 2000-2001 through the 2001-2002 school years was significantly correlated with the participation measures. The change in reading scores and the change in math scores provided the best fits in the middle and high school first-stage equations, respectively. We tested the exogeneity of these instruments by calculating the correlation coefficient for the instruments and the

bounds of zero but also will include continuous positive variables. For that reason, equation (1) is estimated using a tobit equation.<sup>13</sup>

In Kentucky, test scores across levels of schooling were not designed to be comparable.<sup>14</sup> As a result, equation (1) must be estimated separately for each level of schooling, elementary, middle, and high school. Each level of schooling also has two alternative measures of participation in AMSP. These measures are the percentage of teachers in the school that choose to partake of the professional development activities (with the total number of teachers in the school as an independent control variable) and the per teacher number of hours of participation in AMSP activities in the school. With the two participation measures and three levels of schooling, six tobit instrumental variables equations were estimated.<sup>15</sup>

To assess whether AMSP professional development activities have had an effect on student outcomes, we employ a standard education production function or value-added model.<sup>16</sup> These models estimate test scores, or achievement, at any time period  $t$  as a function of the cumulative influence of family and school resources and the individual characteristics of the student. Similarly, achievement in all other time periods, such as  $t - 1$ , is a function of the cumulative inputs in  $t - 1$ . Conceptually, the model to be estimated at any time period  $t$  is:

$$A_{it} = \alpha_0 R_{it} + \alpha_1 F_{it} + \alpha_2 T_t + \alpha_3 A_{it-1} + \alpha_4 AMSP_{it} + \gamma_i + \varepsilon_{it} \quad (2)$$

where  $A_{it}$  is the math or science test scores for students in school  $i$  in year  $t$ ;  $R_{it}$ , represents characteristics of school  $i$  in year  $t$ ;  $F_{it}$  is family and student socioeconomic

---

residuals from the OLS estimates of the structural equations. The test statistics were insignificant in every case which indicates the instruments are exogenous.

<sup>13</sup> Tobit estimates do not directly yield the predicted values for AMSP participation. These values must be calculated using parameter and sigma estimates generated by the tobit equation estimation.

<sup>14</sup> Since the grades covered by schools varies, we utilize data on three student groups: 4<sup>th</sup> and 5<sup>th</sup> graders, 7<sup>th</sup> and 8<sup>th</sup> graders, and 11<sup>th</sup> graders.

<sup>15</sup> These tobit results are available from the authors upon request. For the sake of efficiency, they are not presented in the paper.

<sup>16</sup> See Hanushek (1979; 1986).

characteristics in school  $i$  in year  $t$ ,  $T_t$  is a vector of school year fixed effects,  $A_{it-1}$  is student achievement of school  $i$  in the prior year  $t - 1$ , and  $AMSP_{it}$  is participation in the professional development intervention in school  $i$  in year  $t$ . Because there may be unobserved characteristics of the schools that are not captured by available data,  $\gamma_i$  is an individual school level fixed effect; and  $\varepsilon_{it}$  is a random error.

As described in detail above, for comparison purposes the model above will be estimated in two ways. First, because we expect conceptually that the measures of AMSP participation are not exogenous, equation (2) will be estimated using the instrumental variables (IV) predicted participation measures.<sup>17</sup> Because IV models may contain measurement error as well, we estimate equation (2) under the assumption that AMSP participation is exogenous; i.e., using ordinary least squares regression. A comparative analysis of the coefficients under each procedure then provides us with some understanding of the nature of the potential selection bias that influences AMSP participation.

## V. Results

Table 2 summarizes the results of our tests by listing the estimated coefficients (and standard errors) on the main variable of interest, AMSP participation. Columns (a) and (b) present the coefficients for the two alternative measures of participation when the model is estimated with ordinary least squares with school fixed effects. Because we expect schools to select into AMSP when they are low performing, the results on this set of equations should be biased downward. For comparison, columns (c) and (d) present the coefficients from the instrumental variables estimation procedure.

---

<sup>17</sup> We estimated the equations with the STATA software. We identified the predicted values of the participation variables as IVs when we estimated the second-stage equations to insure the validity of the standard errors and test statistics.

As illustrated in Table 2, the results suggest that AMSP is having positive and significant effects in both the middle and high schools but not at the elementary school level. First, column (a) shows that the more hours of AMSP participation (i.e., more hours of PD training) per teacher in middle schools raises math scores of the students in those schools. Further, column (b) illustrates that increasing the percentage of teachers

**Table 2: Effects of AMSP Participation on Math and Science Scores**

	Hours Per Teacher (OLS)	% Participating (OLS)	Hours Per Teacher (IV)	% Participating (IV)
	(a)	(b)	(c)	(d)
4 <sup>th</sup> Grade Science Scores	-0.027 (0.1)	0.02 (0.02)	1.08 (0.67)	0.2 (0.13)
5 <sup>th</sup> Grade Math Scores	0.11 (0.09)	0.01 (0.02)	1.23* (0.66)	0.23* (0.13)
7 <sup>th</sup> Grade Science Scores	0.23 (0.14)	0.05** (0.02)	1.43*** (0.5)	0.28*** (0.08)
8 <sup>th</sup> Grade Math Scores	0.23** (0.1)	0.036* (0.019)	0.99*** (0.37)	0.19*** (0.06)
11 <sup>th</sup> Grade Science Scores	0.17 (0.1)	0.046* (0.025)	1.96*** (0.5)	0.4*** (0.1)
11 <sup>th</sup> Grade Math Scores	0.3*** (0.1)	0.07*** (0.02)	1.43*** (0.5)	0.29*** (0.1)

Note: the cells contain the estimated coefficients with the standard errors in parenthesis. Heteroskedasticity-robust standard errors were utilized for the estimation of the OLS equations.

- \*-significant at the 10% confidence level.
- \*\* -significant at the 5% confidence level.
- \*\*\*-significant at the 1% confidence level.

in a middle school who participates in AMSP significantly raises both the math and science standardized scores for that school, ceteris paribus. Moving down the table to the bottom two rows, we see that the hours of participation by teachers raise math scores for high school students but do not significantly affect the science scores. But again, when AMSP participation is measured by the percentage of a school's teachers who participate in AMSP, the estimated coefficients on both math and science are significant and positive

suggesting that higher percentages of teacher participation raise scores in both content areas.

Columns (c) and (d), in contrast, summarize the estimated AMSP effects when the model is estimated using an instrumental variables procedure or under the assumption that AMSP is nonrandom. As predicted and without exception, the AMSP results are stronger in the latter case. Teacher participation in AMSP, measured under both alternatives, is now positive and significant at the .001 level in explaining math and science scores in both middle and high schools. Surprisingly, participation is also having weakly significant effects at the elementary school level in math but still has no significant effects in science. Not only are the significance levels stronger but the magnitude of the estimated coefficients is larger across all estimates in columns (c) and (d) than in the first two columns. For example, compare columns (b) and (d) in the last row of the table. Column (b) suggests that increasing the AMSP participation of teachers in a school by one percent raises math scores of 11<sup>th</sup> graders by 0.07 points. In contrast, controlling for the biases that affect which teachers choose to participate in AMSP, column (d) illustrates that the same one percent increase in teacher participation leads to an increase in average math scores for a high school of 0.29 points. While this magnitude may appear small, when added cumulatively over the school life of a student, the learning effects appear substantial.

The estimates from Table 2 were generated while controlling for other factors expected to influence student outcomes. Tables 3-5 provide the complete set of variables and their estimated coefficients.<sup>18</sup> The specifics differ slightly across the many sets of estimates but are largely consistent and in line with theoretical expectations. For

---

<sup>18</sup> The estimates are from the preferred instrumental variables model.

example, some combination of the lagged reading, math, and science scores tend to be positively related to the current year's math or science scores. Schools that performed well in the previous year are more likely to score well this year, *ceteris paribus*. Also, even controlling for these lagged scores, socioeconomic characteristics of the student body prove significant in most of the equation estimates. In particular, schools with higher percentage of free and reduced price lunch students generally perform lower on math and science and schools with higher percentages of black students also perform at lower levels. The size of the school, measured by numbers of teachers, is significant and positive in some of the estimates.<sup>19</sup> Overall, the results on these variables lend credibility to our estimates on the PD effects of AMSP in that they are consistent with findings in the education production function literature.

## **VI. Conclusion**

This work is significant for two reasons. First, and most obviously, we are beginning to answer the question of whether the National Science Foundation efforts to enhance student outcomes in math and science are producing positive results. The particular program examined in this paper is the Appalachian Math and Science Partnership (AMSP). This work suggests that the tentative answer is in the affirmative. We caution that the work could be improved with additional data that includes all AMSP schools across the 3 participating states and data at the student, rather than the school level. There are different types of professional development activities occurring within AMSP and future research will look at these activities individually rather than viewing them in the aggregate.

---

<sup>19</sup> We used a variety of other teacher characteristic measures without qualitatively changing the results. For example, we also controlled for average years of experience and the percent of the teachers with a master's degree. These results are not presented here because missing data for these variables reduced our sample size.

But independently of, and beyond our specific results, this study makes an important contribution. It is clear that policymakers think math and science outcomes should be enhanced in the United States. The National Science Foundation has funded over 40 partnerships for math and science training around the country. Such large investments of human and financial resources call for evaluation. It is not sufficient to assume the federal dollars automatically increase student learning. The partnership programs differ in size and structure across the U.S. Some may be more or less successful than others. A valuable next step would entail detailing the structural differences across these programs and contrasting their effects. Understanding these programs and whether they consistently generate positive returns for in-service professional development efforts is potentially very important in the national movement to improve teacher quality and student learning in math and science. If these programs generally indicate positive returns to learning, a final step would be a cost-effectiveness analysis of these programs relative to other successful professional development programs.

## References

- Acevedo, S. K. (forthcoming) Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement, *Economics of Education Review*.
- Angrist, J. and V. Lavy. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools. *Journal of Labor Economics*. 19(2): 343-369.
- Betts, J., A. Zau, and L. Rice. (2003). Determinants of student achievement: Evidence from San Diego. Public Policy Institute of California.
- Boyd, D., P. Grossman, H. Lankford, S. Loeb, and J. Wyckoff. (2005). How changes in entry requirements alter the teacher workforce and affect student achievement. NBER Working Paper Series. Working paper 11844. Cambridge, MA: National Bureau of Economic Research.
- Clotfelter, C., Ladd, H. & Vigdor, J. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness, *Journal of Human Resources*, 41, 778-820.
- Ehrenberg, R.G. & Brewer, D.J. (1994). Do school and teacher characteristics matter? Evidence from High School and Beyond. *Economics of Education Review*, 13(1), 1-17.
- Fullan, M. and Hargreaves, A. "What's Worth Fighting For in Your School?" New York: Teachers College Press, 1996.
- Goldhaber, D. & Anthony, E. (2007). Can teacher quality be effectively assessed? *The Review of Economics and Statistics*, February 2007, 89(1): 134–150.
- Goldhaber, D. & Brewer, D. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, 32(3), 505-523.
- Harris, D. and T. Sass. (2006). The effects of teacher training on teacher value-added. *Unpublished*.
- Jacob, B. and L. Lefgren. (2002). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. NBER Working Paper Series. Working paper 8916. Cambridge, MA: National Bureau of Economic Research.
- Monk, D. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125-145.
- Rivkin, S., Hanushek, E. & Kain, J.. (2005). Teachers, schools and academic achievement. *Econometrica*, 73(2), 417-458.
- Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2) 247-252.

Wilde, E. T., and Hollister, R. (2007). How close is close enough? Evaluation propensity score matching using data from a class size reduction experiment. *Journal of Policy Analysis and Management*, 26(3), 455-78.

## Appendix: Instrumental Variable (IV) Regression Results

**Table 3: Effect of AMSP Participation on 4<sup>th</sup> and 5<sup>th</sup> Grade Students' Math and Science Scores<sup>20</sup>**

Variables	4 <sup>th</sup> Grade Science Scores		5 <sup>th</sup> Grade Math Scores	
	Hours Per Teacher	N/A	1.08 (0.66)	N/A
% Participating	0.2 (0.13)	N/A	0.23* (0.13)	N/A
Lagged Math Score	0.05*** (0.01)	0.05*** (0.01)	0.02 (0.02)	0.03 (0.02)
Lagged Science Score	0.13*** (0.02)	0.14*** (0.026)	0.29*** (0.02)	0.3*** (0.02)
Lagged Reading Score	-0.006 (0.03)	-0.01 (0.02)	0.18*** (0.03)	0.17*** (0.03)
ARSI Participation	2.87* (1.59)	2.37* (1.32)	3.3** (1.59)	2.01 (1.32)
% Students in Free/Reduced Lunch	-0.006 (0.01)	-0.006 (0.01)	-0.03** (0.01)	-0.03** (0.01)
% Students African American	-0.05 (0.03)	-0.05 (0.03)	-0.07* (0.03)	-0.07* (0.04)
% Students Hispanic	0.07 (0.06)	0.09 (0.06)	-0.01 (0.06)	-0.01 (0.07)
% Students Asian	0.04 (0.11)	0.04 (0.1)	0.06 (0.09)	0.07 (0.1)
Number of Teachers	0.05 (0.06)	0.05 (0.06)	0.11* (0.06)	0.11* (0.06)
<i>N</i>	3,845	3,852	3,848	3,855
<i>Within R-Square</i>	0.35	0.34	0.54	0.54

\*-significant at the 10% confidence level.; \*\*-significant at the 5% confidence level; \*\*\*-significant at the 1% confidence level.

<sup>20</sup> The coefficients on the school year dummies are omitted for brevity. The cells contain the estimated coefficients with the standard errors in parenthesis.

**Table 4: Effect of AMSP Participation on 7<sup>th</sup> and 8<sup>th</sup> Grade Students' Math and Science Scores**

Variables	7 <sup>th</sup> Grade Science Scores		8 <sup>th</sup> Grade Math Scores	
	Hours Per Teacher	N/A	1.43*** (0.49)	N/A
% Participating	0.28*** (0.09)	N/A	0.19*** (0.07)	N/A
Lagged Math Score	-0.01 (0.03)	-0.02 (0.03)	-0.02 (0.03)	-0.03 (0.02)
Lagged Science Score	0.11*** (0.04)	0.1*** (0.04)	0.28*** (0.03)	0.27*** (0.03)
Lagged Reading Score	0.06 (0.05)	0.05 (0.05)	0.27*** (0.04)	0.26*** (0.03)
ARSI Participation	3.09 (2.98)	1.38 (2.83)	3.29 (2.26)	2.11 (2.14)
% Students in Free/Reduced Lunch	-0.03 (0.03)	-0.04 (0.03)	-0.05** (0.02)	-0.06** (0.02)
% Students African American	-0.07 (0.1)	-0.05 (0.1)	0.006 (0.08)	0.01 (0.07)
% Students Hispanic	-0.04 (0.09)	-0.02 (0.09)	0.09 (0.06)	0.11* (0.06)
% Students Asian	-0.08 (0.15)	-0.05 (0.1)	-0.19* (0.11)	-0.17 (0.11)
Number of Teachers	0.02 (0.08)	0.03 (0.08)	0.12** (0.06)	0.12** (0.06)
<i>N</i>	1,592	1,592	1,592	1,592
<i>Within R-Square</i>	0.34	0.35	0.48	0.49

\*-significant at the 10% confidence level; \*\*-significant at the 5% confidence level; \*\*\*-significant at the 1% confidence level.

**Table 5: Effect of AMSP Participation on 11<sup>th</sup> Grade Students' Math and Science Scores**

Variables	11 <sup>th</sup> Grade Science Scores		11 <sup>th</sup> Grade Math Scores	
	Hours Per Teacher	N/A	1.96*** (0.51)	N/A
% Participating	0.4*** (0.1)	N/A	0.29*** (0.1)	N/A
Lagged Math Score	0.05 (0.04)	0.02 (0.04)	0.04 (0.04)	0.02 (0.04)
Lagged Science Score	-0.08* (0.05)	-0.06 (0.04)	-0.05 (0.05)	0.39*** (0.03)
Lagged Reading Score	0.34*** (0.03)	0.35*** (0.03)	0.39*** (0.03)	-0.03 (0.05)
ARSI Participation	5.47*** (1.56)	5.68*** (1.69)	3.92** (1.62)	4.22** (1.73)
% Students in Free/Reduced Lunch	0.003 (0.03)	-0.008 (0.02)	-0.003 (0.03)	-0.01 (0.02)
% Students African American	-0.02 (0.03)	-0.02 (0.03)	-0.004 (0.03)	-0.004 (0.03)
% Students Hispanic	0.23 (0.19)	0.47** (0.2)	0.18 (0.21)	0.35* (0.207)
% Students Asian	-0.49** (0.25)	-0.42* (0.25)	-0.24 (0.26)	-0.18 (0.26)
Number of Teachers	0.004 (0.04)	0.02 (0.04)	0.05 (0.04)	0.06 (0.04)
<i>N</i>	1,288	1,289	1,288	1,289
<i>Within R-Square</i>	0.31	0.30	0.40	0.39

\*-significant at the 10% confidence level; \*\*-significant at the 5% confidence level; \*\*\*-significant at the 1% confidence level.