Appendix G: Technical Report for Year Two Evaluation of the Achievement Challenge Pilot Project in the Little Rock Public School District

The Technical Appendix of the evaluation provides a more thorough explanation of the student effect portion of the evaluation. This Appendix is divided into three sections. Section I discusses the data analyzed in the paper and develops the empirical models used to measure the impact of the Achievement Challenge Pilot Project (ACPP) on student achievement. Section II reports the results of the empirical models. Section III explains the findings from the analysis.

Data and Methods

We acquired individual data for the universe of public school students enrolled in Little Rock, Arkansas elementary schools in the 2005 through 2007 school years, providing us with two observations of student test scores gains.[13] For each elementary student in Little Rock, this dataset included demographic information, test scores, an identifier for the student's classroom teacher, and a unique student identifier that allows us to track each student's performance over time. We evaluate the impact of the adoption of the performance pay program on student proficiency in math, reading, and language.

Test scores are reported in our dataset in Normal Curve Equivalent (NCE) units. NCE's rank the student on a normal cure compared to a nationally representative group of students who have taken the test. NCE's are similar to percentile scores, but differ in that they are equal-interval scaled, meaning that the difference between two scores on one part of the curve are equivalent to the difference of a similar interval on another part of the curve. NCE scores are scaled between 1 and 99 with a mean of 50.

We utilize the differences-in-differences procedure to study the impact of performance pay. Unfortunately, we are forced to exclude students in the schools that began the performance pay treatment prior to 2007. The reason for the exclusion is that since these schools were treated in each year for which we have data, in the analysis they would become part of the comparison group.

We use an ordinary least squares (OLS) regression to estimate a model taking the form:

$$Y_{i,a,t} = \beta_o + \beta_1 Y_{i,a,t-1} + \beta_2 Student_{i,t} + \beta_3 School_{i,t} + \beta_4 Year_t + \beta_5 Treat_{i,t} + \varepsilon_{i,t} \quad (1)$$

Where $Y_{i,a,t}$ is the test score gain of student i in subject a in the spring of year t; Student is a vector of observable characteristics about the student; School is vector indicating the school that the student attended; Year is an indicator variable for the year; and ε is a stochastic term clustered by teacher.[14]

---

[13] Here and throughout this paper we use the spring term year to identify the school year. That is, the 2004-05 school year is referred to as 2005.
[14] Results are similar if standard errors are clustered by school. Results available from authors by request.

Treat is an indicator variable for whether the observation occurred for a student attending the treatment school during the treatment year. That is, this variable is an interaction between Year = 2007 and the indicator variable for each school that was eventually treated. When Equation (1) is estimated using OLS, it can be shown that the coefficient on Treat (β_5) becomes an estimate of the changes in the conditional expectations of test score gains resulting from the performance pay treatment. That is, β_5 represents the impact of the performance pay treatment after having accounted for the differences in the test scores that occur naturally over time and within the individual schools.

We are able to estimate these equations in math, reading, and language in elementary schools. However, the grades included in the analyses of each subject differ due to limitations of the testing scheduled in Little Rock. Students were administered the math version of the ITBS in all grades K-5 in each of the three years from 2005 - 2007, and so each of these grades are included in the analyses. However, Little Rock students were not administered the ITBS language or reading test in grades 3, 4, or 5 until 2006. Further, students were not administered the ITBS reading test in Kindergarten until 2007. These data limitations lead us to only include students in grades 2 and 3 for the reading analyses and students in grades 1, 2, or 3 in the language analyses -- the only grades for which we have both a pre- and post test score for students in both the baseline and treatment eligible year.

A potential limitation of our approach is that we may have an endogeneity problem since schools were not randomly assigned to the performance pay treatment. In particular, as discussed above, the treatment was made available to schools non-randomly and treated schools had higher minority populations and lower income students on average.

We are able to partially account for this endogeneity bias by including school to account for heterogeneity in school quality. However, it is also worth noting that summary statistics indicate that any endogeneity bias should likely tend to underestimate the impact of the performance pay treatment. Table 2 provides the baseline descriptive comparisons between the ACPP and comparison schools.

Table 1: Baseline Descriptive Statitics for ACPP and Comparison Schools, 2006

	All		Comparison		ACPP	
Variable	Mean	Std.	Mean	Std.	Mean	Std.
Black	0.69	0.46	0.67	0.47	0.88	0.33
Asian	0.02	0.12	0.02	0.13	0.00	0.00
Hispanic	0.04	0.19	0.04	0.19	0.06	0.23
Indian	0.00	0.06	0.00	0.06	0.00	0.05
Male	0.50	0.50	0.50	0.50	0.52	0.50
Eligible for Free or Reduced Lunch	0.65	0.48	0.63	0.48	0.88	0.33
Baseline Math	50.41	21.54	51.15	21.57	38.57	17.27
Baseline Reading	50.16	21.53	51.12	21.55	40.53	18.87
Baseline Language	49.87	21.13	50.88	21.18	40.21	18.02
Math Gain 2006	1.94	14.37	2.14	14.25	-1.29	15.83
Reading Gain 2006	1.83	14.51	1.89	14.53	1.19	14.29
Language Gain 2006	0.00	16.07	0.18	15.90	-1.75	17.45

Note: Only students included in overall math regression are included in above summary statistics for demographic variables. Reading and language test descriptive statistics include only students used in those regressions.

Note that Table 1 shows that in 2006, the year before the policy was available, on average students in ACPP schools made smaller test score improvements in each of the three subjects used in our analyses. That is, we should expect that in the absence of treatment these schools should have made smaller test score improvements than the control schools, which would tend to bias the estimation of the treatment effect downward. Nonetheless, we recognize that lack of random assignment is a concern with any results.

Results

The results from estimation of equation (1) are reported in Table 2. Recall that we are forced to use a more restricted group of grades in the reading and language analyses, which accounts for the variation in the number of observations across subjects.

In each subject we find a statistically significant positive relationship between the performance pay treatment and student achievement. The analyses suggest that the performance pay treatment led to an increase of about 3.52 NCE points in math, 3.29 NCE points in reading, and 4.56 NCE points in language

The size of these effects is economically substantial. We can use the summary statistics for baseline achievement in these subjects reported in Table 1 to put our results into terms of standard deviation units. Dividing the effect size by the standard deviation of the baseline test score in the subject, our results suggest that performance pay increased student proficiency by 0.16 standard deviations in math, 0.15 standard deviations in reading, and 0.22 standard deviation units in language.

Table 3 reports the results of estimation of the overall treatment effect when we include a fixed effect for each individual teacher. The table shows that the results are qualitatively similar to those without a teacher fixed effect, with the exception that the impact of performance pay in language becomes statistically insignificant. Somewhat surprisingly, the small gain in the R-Squared value between the analyses reported in Tables 2 and 3 suggest that the teacher fixed-effect is explaining very little of the variance in student achievement.

Table 2: Regression Results – Overall Treatment Effect

	Math		Reading		Language	
Variable	Coef.	t-value	Coef.	t-value	Coef.	t-value
Math t-1	0.70	82.60***				
Reading t-1			0.68	68.72		
Language t-1					0.68	60.12
Black	-4.60	-12.34***	-4.69	-10.21***	-2.75	-6.19***
Asian	3.65	4.28***	1.04	0.76	5.81	5.12***
Hispanic	-1.14	-1.66*	-1.62	-1.89*	1.18	1.18
Indian	-1.80	-1.15	-3.78	-2.01**	-3.19	-1.27
Male	0.03	0.12	-0.41	-1.41	-2.87	-10.12***
Lunch Eligible	-2.47	-8.31***	-2.88	-7.24***	-3.19	-8.02***
Treat	3.52	2.84***	3.29	2.35**	4.56	2.77***
Constant	23.11	18.82***	19.40	19.02***	20.04	12.56***
N	13389		5948		8933	
Adjusted R-Squared	0.65		0.71		0.62	

Estimated via OLS. Models also control for school, grade, and year fixed effects. Standard errors clustered by teacher.

*** Significant at p<= .01
** Significant at p<= .05
* Significant at p<= .10

Table 3: Regression Results – Including Teacher-Fixed Effect

	Math		Reading		Language	
Variable	Coef.	t-value	Coef.	t-value	Coef.	t-value
Math t-1	0.71	83.30***				
Reading t-1			0.69	68.91***		
Language t-1					0.68	59.88***
Black	-4.41	-11.71***	-4.56	-9.77***	-2.70	-5.97***
Asian	3.64	4.19***	1.33	0.97	5.92	5.55***
Hispanic	-0.86	-1.26	-1.27	-1.46	1.68	1.80*
Indian	-1.34	-0.78	-2.89	-1.61	-3.11	-1.40
Male	0.06	0.25	-0.43	-1.41	-2.71	-9.51***
Lunch Eligible	-2.24	-7.52***	-2.82	-6.90***	-2.90	-7.21***
Treat	5.23	3.18***	3.05	3.31***	2.04	1.06
Constant	49.77	26.42***	22.60	5.83***	24.54	11.73***
N	13389		5948		8933	
Adjusted R-Squared	0.70		0.73		0.65	

Estimated via OLS. Models also control for school, grade, and year fixed effects. Standard errors clustered by teacher.

*** Significant at p<= .01
** Significant at p<= .05
* Significant at p<= .10

Conclusions

This paper has made a variety of contributions to the literature through the lens of performance pay for teachers. We have added to the limited empirical research on performance pay programs. We find that adoption of performance pay led to substantial improvements in student math, reading, and language proficiency.