

*Working Paper, Submitted to the Journal of Public Economics*

An Evaluation of Teacher Performance Pay in Arkansas

Marcus A. Winters  
Doctoral Fellow  
Department of Economics  
University of Arkansas

Gary W. Ritter  
Endowed Chair  
Department of Education Reform  
University of Arkansas

Joshua H. Barnett  
Doctoral Fellow  
Department of Public Policy  
University of Arkansas

Jay P. Greene  
Endowed Chair, Head  
Department of Education Reform  
University of Arkansas

This research was funded by a generous grant by the Walton Family Foundation.

Acknowledgements:

We wish to thank the Walton Family Foundation for their generous support of this research. We thank Ed Williams, Hugh Hattabaugh, Jim Wohlleb, Karen Dejarnette, Maurecia Robinson, Olivine Roberts, Sadie Mitchell, and Suellen Vann from the Little Rock School District, and Lisa Black of the Public Education Foundation of Little Rock for their help with the data necessary to conduct the evaluation. Also, we are grateful for helpful comments from Bruce Dixon, Julie Trivitt, Gary Ferrier, and Jungmin Lee. Any remaining errors are, of course, our own.

*JEL classification:* H11; I2; J33; M52

**Keywords:** Teacher salaries; Incentive systems; Merit pay; Teacher incentives; Student performance

*Abstract*

Several school systems have considered adding a component to the wage structure that directly compensates teachers based upon the academic gains made by the students in a teacher's care, at least partly measured by student scores on standardized tests. This paper adds to the sparse literature on the impact of such performance pay programs. Using individual student level data, we evaluate the impact of a generous performance pay program for teachers in Little Rock, Arkansas, on student math proficiency. We find that providing teachers with bonuses based on test score improvements increased student math proficiency by between 3.6 and 4.6 Normal Curve Equivalency ranks in a year. Results are robust across three specifications: a school-level differences-in-differences (DD) estimation; a DD estimation controlling for student fixed-effects; and a student fixed-effects model accounting for the number of years a student has attended a treated school. In each evaluation we compare the performance of students in treated schools to students attending similar schools that were not offered the chance to participate in the program. This paper will help to inform the general policy debate over performance pay across the nation and add to the very limited research on the effectiveness of these programs.

## *Introduction*

The majority of public school teachers receive compensation according to a salary schedule that is nearly entirely determined by their number of years of service and their highest degree attained. This system, however, has seen increasing attacks from policymakers and researchers in recent years. Several school systems have considered adding a component to the wage structure that directly compensates teachers based upon the academic gains made by the students in a teacher's care, at least partly measured by student scores on standardized math and reading tests. The public school systems of Florida and Denver have recently adopted such "performance pay" policies. The recent push for the expansion of performance pay plans by influential political leaders, such as the Mayor of New York City and the Governor of Massachusetts, suggest that use of such policies could continue to grow in upcoming years.

The idea of paying teachers at least in part based on the measured learning gains made by students is not entirely novel. Such programs were quite common in the early twentieth century (Murnane and Cohen, 1986) and were present in some form in about 12% of public school districts in the early 1990's (Ballou, 2001). However, though performance pay programs are not new, we currently know very little about their effects on student achievement.

Several researchers have evaluated the impact of performance pay programs on reported teacher satisfaction, classroom practices, and retention (Johns, 1988; Jacobson, 1992; Heneman and Milanowski, 1999; Horan and Lambert, 1994). Some U.S. studies have found that programs providing bonuses to entire schools, rather than changing the pay of individual teachers, have a positive impact on student test scores (Clotfelter and

Ladd, 1996; Ladd, 1999). However, there is currently very little empirical evidence from the United States suggesting that direct teacher-level performance pay leads to better student outcomes.<sup>1</sup>

Figlio and Kenny (2006) independently surveyed the schools that participated in the often-used National Educational Longitudinal Survey (NELS). They then supplemented the NELS dataset with information on whether schools compensated teachers for their performance. They found that test scores were higher in schools that individually rewarded teachers for their classroom performance.

Eberts, Hollenbeck, and Stone (2000) used a differences-in-differences approach to evaluate the impact of a performance incentive for teachers in an alternative high school in Michigan. They found that the program increased completion percentages but had no effect on grade point averages or attendance rates and actually increased the percentage of students who failed the program. However, the study was unable to provide a direct evaluation of student achievement (i.e. test scores). Further, the study's focus on an alternative dropout recovery school produces difficult estimation problems and could limit its use in the discussion of traditional public K-12 education.

Finally, Keys and Dee (2005) evaluated a performance pay program in Tennessee. They took advantage of the fact that this program operated at the same time as the notable Tennessee STAR program, a random assignment experiment on the impact of class size on student achievement. Under STAR, students in these schools were randomly assigned to classrooms of different sizes. This assignment additionally meant that students were randomly assigned into classrooms led by teachers who were or were not participating in

---

<sup>1</sup> There is also limited evidence on the impact of performance pay in other countries. Lavy (2002) found that a school-based program in Israel increased student performance, and Glewwe, Ilias, and Kremer (2003) found similar results from a program in Kenya.

a state sponsored performance pay program. Importantly, however, teachers were not similarly randomly assigned to participate in the performance pay program, and thus the study cannot be considered a conventional random assignment experiment of the performance pay plan. Nonetheless, they found that students randomly assigned to classrooms with teachers participating in the performance pay program made exceptional gains in math and reading, though these results could be driven by selectivity in the teachers that choose to participate in performance pay programs, rather than the incentives of the program itself.

With this paper, we add to the limited research on the impact of teacher-based performance pay on student academic achievement with data from the United States. We utilize a differences-in-differences approach and two individual fixed-effects models to evaluate the impact of a generous performance pay plan operating in Little Rock, Arkansas on student test score gains in math. Though limited by the size of the program and the lack of random assignment, our results suggest that students attending schools where teachers directly received bonuses based upon their students' test score gains made substantially larger improvements in math proficiency than students in demographically similar control schools that did not participate in the program.

### *The Program*

In 2004, the Little Rock School District and the Public Education Foundation of Little Rock joined efforts to create a pilot performance pay program for teachers, the Achievement Challenge Pilot Project (ACPP). Under the program, teachers receive direct bonuses based on the average academic growth of students in their class as measured by

gains on the complete battery of a nationally normed standardized test, the Stanford Achievement Test (SAT), and the number of students in the class. Teachers receive a per-student bonus that is dependent upon the average test score gains in the class. If the average gain in the classroom is between 0-4% the teacher receives \$50 per student in the class; between 5-9% receives \$100 per student in the class; between 10-14% receives \$200 per student in the class; and if gains average above 15% the teacher receives a \$400 per student bonus. The potential bonus is substantial, with a maximum of \$11,200 at the end of the year.

A limited number of high minority enrollment schools with low standardized test scores were chosen and agreed to participate in ACPP. The program is now currently operating in five elementary schools across the district. At the time of our analysis, two elementary schools were participating in the program. One of the treated schools first adopted the performance pay program in 2003-04, and the other began the program in 2004-05.

### *Data*

Unfortunately, schools were not randomly assigned to participate in the program. Lacking random assignment, we identified three control schools in the Little Rock School District that had demographic and achievement characteristics that were similar to those of the treated ACPP schools but were not offered the chance to participate in the program. Table 1 compares demographic characteristics and baseline test scores of the treatment and control schools. Because, as we will see, two of our analyses only include

students in one of the treated schools we also report the descriptive stats of that school alone, listed as “Main Treated” in the table.

The table shows that control schools had slightly higher average math scores in the year before the program was adopted in the treated schools. Eventually treated schools had a slightly higher percentage of Hispanic students and a slightly lower percentage of black students than control schools. We will discuss the variable Gap later in this paper, though it is worth pointing out here that a slightly higher percentage of treated students were in this category.

[TABLE 1 ABOUT HERE]

We use the three control schools as a counterfactual in our analysis and thus assume that their experience represents the experience that would have occurred in treated school absent the performance pay program. None of the matched control schools were given the opportunity to participate in the program. Thus, though lack of random assignment continues to pose an endogeneity problem in our analysis, none of the control schools explicitly declined the opportunity to participate in the program and thus are not necessarily different from the treated schools in their willingness to adopt performance pay.

We obtained each student’s score in Normal Curve Equivalents (NCE) on a standardized math test administered across the district from 2002-03 through 2005-06. NCE scores indicate the student’s rank among a nationally representative group of students along an equal interval scale. We collected three observations of math proficiency for each student in our dataset. We then pooled the data to create a panel

format where each observation represents a particular student's math proficiency at the end of a school year.

There are two irregularities with the testing schedule that, though they are unlikely to have a particular impact on our analysis, are worth mention here. First, the standardized test administered in Little Rock changed from the SAT to the Iowa Test of Basic Skills (ITBS) beginning in the 2004-05 school year. Thus, in the evaluation described later, we measure the baseline test score gains for students by subtracting their previous NCE score on the SAT from their NCE on the ITBS a year later. The use of NCE scores makes this test switch possible, however, since both represent the student's performance relative to a nationally representative group of students. Further, this problem exists equally for members of the treatment and control groups, so it should produce no systematic differences.

As described above, students in the treated schools continue to be administered the complete battery of the SAT, which determines the bonuses that teachers receive. Students in our control schools, however, were only administered the SAT in the baseline years, when it was the state mandated exam.

However, use of the ITBS is also potentially useful in our analysis because scores on that exam in no way determine the bonuses teachers earn under the performance pay program in the treated schools. We might worry that any increase in the test scores of students on the test that produces salary bonuses in performance pay schools could occur because teachers find ways to increase student test scores without actually increasing their academic proficiency. Such worries about "teaching-to-the-test" are common to all education reforms that rely on the results of standardized tests. Since scores on the ITBS

are not part of the performance pay program, teachers in the treated schools have no more incentive to manipulate the results of this test than do teachers in the comparison schools.

Another difficulty is that, according to the regular testing schedule in Little Rock, students were not tested in each grade and year. Table 2 summarizes the mathematics testing schedule in Little Rock by grade and year during this time. Our evaluation design requires an observation of student proficiency at three points in time. The existence of three observation points allows us to calculate two observations of test score improvements: one test score improvement before implementation of the policy and one test score gain after implementation of the policy. As Table 2 illustrates, we can observe three test scores only for students who were in the fourth or fifth grades during the 2005-06 school year. Assuming they were cumulatively promoted, for students in the fourth grade in the end year of the evaluation, we have math scores from their third, fourth, and fifth grade years, allowing us to directly calculate one-year test score gains once before and once after implementation of the policy in the 2004-05 school year.

[TABLE 2 ABOUT HERE]

The situation is different for students in the fifth grade in 2005-06. While we have test scores for these students cumulatively in the fourth and fifth grades, we do not similarly observe their third grade proficiency because the test was not administered to third grade students in 2003-04. However, we do observe these students' test score in the second grade in 2002-03. We use this second grade test score as the baseline proficiency for these students. Thus, the calculation of test score gains in the period before adoption of the policy represents a two rather than a one year of gain for these students.

Though less than ideal, the gap in the observation of proficiency for students who were in the fifth grade in 2004-05 is a surmountable problem. First, Table 1 shows that the gap in baseline test score gains occurs for a similar proportion of students in both the treatment and control schools. Secondly, since no students were treated in 2004, this data transformation has no effect on the analysis if we make the reasonable assumption that there was no systematic change in the Normal Curve Equivalent (NCE) score made by students with this gap in 2003 that depended upon whether the school was eventually treated.

Students do appear to have made very substantial math gains in the year for which we cannot observe their math proficiency, about 20 NCE ranks. Similar gains were made by Gap students who were in the treatment and control groups, indicating that the gains likely were not conditioned on eventual treatment. The reason for these gains remains unclear.<sup>2</sup> However, it is worth noting that since a slightly higher percentage of students in eventually treated schools had a gap in their test scores, the larger gain in the baseline year should tend to understate the gains made in the treatment year for these students.

We can account for any disparity caused by the gap in observations of proficiency by incorporating a binary variable indicating whether the student's baseline test score gain was over one or two years. This effectively measures the test score gains for these students on a separate intercept. We also estimated each of our models only including those students who were in the fourth grade in 2005-06 and thus did not have a gap in their test scores. We briefly discuss the results of these estimations below, but for space considerations do not report them directly.

---

<sup>2</sup> It had occurred to us that this large difference was caused by a data error. However, we have been assured on several occasions by the Little Rock School District that the numbers are correct. Further, the District reported to us that this gain was found across the district, not simply for students in our schools of interest.

### *Method*

We conduct three separate estimations to evaluate the impact of the performance-pay program on student math proficiency. In our first estimation, we follow a relatively conventional differences-in-differences approach. Differences-in-differences is a simple panel data method that has been utilized in literally hundreds of studies. This methodology is useful in cases, such as ours, concerned with changes across two periods in time and where matched comparison groups are available with which to compare treated groups.

Differences-in-differences requires only two observations of the dependent variable for each member of the sample across time. In our case, we require one observation of test score gains in math before the performance pay intervention and one observation in the first year of intervention. We use OLS with heteroskedastic robust standard errors clustered by school to estimate a regression taking the form:

$$(1) \quad \Delta Math_{i,t} = \beta_0 + \beta_1 Gap_{i,t} + \beta_2 Year^{pre}_t + \beta_3 Year^{post}_t + \beta_4 School^c_i + \beta_5 School^{treat}_i + \beta_6 Demographics_i + \beta_7 Treat_{i,t} + \varepsilon_{i,t}$$

The dependent variable is student math gains in NCE rank. That is,  $\Delta Math_{i,t} = Math_{i,t} - Math_{i,t-1}$ , where  $i$  indexes individuals and  $t$  indexes the year of the observation.<sup>3</sup>  $Year^{pre}$  indicates that the observation is of gains in the year prior to the policy and  $Year^{post}$  indicates that the observation is of student test score gains in the year after the

---

<sup>3</sup> Use of a difference as a dependent variable is not the normal procedure in differences-in-differences estimation. That is, our model is essentially a differences-in-differences of differences. However, in the case of schools we are particularly concerned with gains students make in proficiency, not their overall level. However, though unusual in the differences-in-differences approach, the use of a difference as a dependent variable should not pose any particular problem for the analysis.

policy was implemented.<sup>4</sup>  $School^c$  is a vector of indicator variables representing that the student attended a particular one of the control schools, while  $School^{treat}$  indexes that the student attended an eventually treated school.<sup>5</sup> The variable *Demographics* is a vector of time invariant variables controlling for student race, gender and free-lunch status.

The binary variable *Gap* indicates whether the student was in the cohort for which the observation of the baseline test score gain was across two years rather than only one year (as described above). That is, *Gap* equals 1 for observations where the student was in the fourth grade in 2004-05 and zero otherwise (including the observation of the student in the next year). Inclusion of this variable allows us to estimate the baseline test score gains for these students on a separate intercept and should thus adequately account for any gains that were made during the missing test score year.

*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  $Year^{post}$  and  $School^{treat}$ . When Equation (1) is estimated using OLS, it can be shown that the coefficient on *Treat* ( $\beta_7$ ) becomes an estimate of the changes in the conditional expectations of test score gains resulting from the performance pay treatment. That is,  $\beta_7$  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. Formally, it can be shown that:

$$\beta_7 = [(\Delta\text{Math} \mid \text{School}^{treat}, \text{Year}^{post}) - (\Delta\text{Math} \mid \text{School}^c, \text{Year}^{post})] - [(\Delta\text{Math} \mid \text{School}^{treat}, \text{Year}^{pre}) - (\Delta\text{Math} \mid \text{School}^c, \text{Year}^{pre})].$$

---

<sup>4</sup> We use the pre and post designations, rather than t and t-1 because, as described above, for some students the initial gain is measured across two instead of one year.

<sup>5</sup> Since both are vectors of school indicator variables, there is no particular mathematical reason to disaggregate  $School^c$  and  $School^{treat}$  in the equation. We do so here to aid in the explanation of our differences-in-differences methodology below.

Unfortunately, in this analysis we are forced to exclude students in the school that began the performance pay treatment in 2003-04. The reason for the exclusion is that students in this school were treated in both years for which we have the test score data necessary to calculate gains in math. After taking first differences of Equation (1) it is easy to see that the estimate on *Treat* would only include those students who went from non-treatment to treatment in that particular year of observation. Students in the school that was always treated would be used as part of the comparison group in the evaluation rather than in the treatment group. This is a common problem in estimations using panel data. However, we are able to include students in these schools in a later estimation performed in this paper.

In a second model, we further expand upon Equation (1) by including a student fixed-effect. This fixed-effect allows us to account for unobserved factors that affect student performance and often are the bane of education policy research. Inclusion of this fixed-effect is equivalent to adding an indicator variable indexing each individual student. Since students do not leave one school to attend another school in our sample, inclusion of this variable disallows the use of the school indicator variables ( $School^c$  and  $School^{treat}$ ) as well as the time invariant demographic characteristics from Equation (1). Thus, our analysis becomes a differences-in-differences estimation across students, rather than schools. Using the within fixed-effects estimator in STATA with heteroskedastic robust standard errors clustered by school we estimate:

$$(2) \Delta Math_{i,t} = \beta_0 + \beta_1 Gap_{i,t} + \beta_2 Year^{pre}_t + \beta_3 Year^{post}_t + \beta_4 Treat_{i,t} + \gamma_i + \varepsilon_{i,t}$$

The student fixed-effect is represented by  $\gamma_i$ , and all other variables are as previously defined. As in the previous equation, the coefficient on *Treat* ( $\beta_4$ ) represents

the difference-in-differences estimate of the impact of the performance pay treatment, but in this equation the estimation is at the student rather than the school level.

In a final analysis, we perform another student fixed-effects evaluation that allows for the inclusion of students in the school that began treatment in 2003-04 and thus were treated in both years of our sample. We estimate an equation similar to (2) but change the Treat variable to instead represent the number of years that the student has attended a school with performance pay, Years\_Treat. Again using OLS with heteroskedastic robust standard errors clustered by school we estimate:

(3)

$$\Delta Math_{i,t} = \beta_0 + \beta_1 Gap_{i,t} + \beta_2 Year^{pre}_t + \beta_3 Year^{post}_t + \beta_4 Years\_Treat_{i,t} + \gamma_i + \varepsilon_{i,t}$$

Model (3) allows the inclusion of both treated schools because the always treated students were treated for 1 year in the first and 2 years in the second year of the sample. After taking a first difference we can see that these students remain in the treatment rather than the control group in the estimation. This model makes the assumption that the impact of performance pay is linear across years.

### *Results*

The results of our school-level differences-in-differences estimation, represented by Equation (1), are reported in Table 3. The coefficient on Treat indicates that the performance pay treatment improved math test score gains by about 3.64 NCE ranks relative to students in untreated schools in the first year of treatment. This result is statistically significant at any convention level.

[TABLE 3 ABOUT HERE]

We report the results of the individual fixed-effects differences-in-differences estimation in Table 4. Again, the coefficient on the treatment variable is significantly positive, though here the estimated impact from the treatment is larger at 4.19 NCE ranks after the first year of treatment. Again, the Gap variable is quite large and statistically significant.

[TABLE 4 ABOUT HERE]

Finally, Table 5 reports the results of estimating Equation (3), which alters the independent variable of interest to represent the number of years that the student has attended a treated school. Recall that this estimation allows use of the school that began treatment in the first year in our sample, which accounts for the increase in the number of observations. Here we again find a positive impact from the performance pay treatment. The coefficient on Years\_Treat indicates that for each year the student attends a school with performance pay he makes an additional 4.63 NCE rank point improvement relative to similar students in untreated schools. This yearly estimate is higher, but similar to the one-year estimate reported in Table 4, which indicates that the assumption of linearity across years is likely appropriate.

[TABLE 5 ABOUT HERE]

The large coefficient on the Gap variable is somewhat concerning. The size of this estimate, between 24 and 32 NCE ranks, is consistent with descriptive statistics on the gains made by Gap students. However, there is no theoretical reason for students to have made such a large increase. As discussed previously, as long as there is no impact from Gap conditional on the student eventually being treated, estimating the gains made by these students on a separate intercept, as we did above, should sufficiently account for

any contamination of results. For a further robustness check, however, we estimated each of our three models including only those students who were in the fourth grade in 2005-06 and thus had no gap in their test score performance. The coefficient on the treatment variable in the first two models, the more conventional school and student-level difference-in-differences estimations, were very similar to those reported above. However, excluding these students, and thus the Gap variable as well, caused the R-square dropped considerably, to about .05. This clearly indicates that that Gap is a large portion of our R-square and thus those R-squares reported in the tables are misleadingly high. Without the Gap variable the model is explaining relatively little of the variation in test scores among student in these schools, which is not uncommon in regressions predicting test score gains. Overall, the main results hold in absence of the Gap students. In the third model, however, the coefficient estimate on the Years\_Treat variable remained similar in size but became statistically insignificant, ( $p = .071$ ).<sup>6</sup>

Lastly, it is worth noting that the Constant in each model reports a negative coefficient. Though at first blush this appears odd, the negative constant impact is not completely unanticipated because our scores are reported in NCE ranks rather than a vertical proficiency scale. A student would lose ground on NCE scores even if she made proficiency gains in math if those gains were smaller than the average student in the nation. The negative coefficient indicates that students in these historically low performing schools are, on average, losing ground compared to a nationally representative group of students, not that they are actually losing proficiency in math. Thus, in the positive impact of the performance pay policy, represented in the treatment

---

<sup>6</sup> Full results are available from the authors by request.

variables, appears to be to slow the decline of students in these schools relative to the rest of the nation as compared to the counterfactual.

### *Conclusion*

This study provides new evidence that providing additional compensation to teachers based upon the measured academic performance of students in their classroom substantially improves academic proficiency. These results add to the sparse literature on the impact of performance pay programs on student achievement.

However, it is important to recognize that there are some important limitations to the present analysis. First, the size of the performance-pay program evaluated in this paper is quite small. Two of the three analyses reported in this paper rely on only the performance in a single treated school, and only two schools were treated by the program in the third analysis. Further research on the ACPP in Little Rock and in larger programs across the nation is necessary before we can comfortably draw general conclusions about the potential benefits of performance pay policies.

Secondly, the analysis is limited by the lack of random assignment of schools into the treatment and control group. At least part of the positive result found in this paper could be driven by factors about the treated schools that led them to participate in the program.

It is also important to keep in mind that this study only evaluates one portion of the potential impact of performance pay programs. There are a variety of factors other than improvement in test scores that policymakers and researchers should consider. Several writers have worried about the potential for these programs to discourage

collaboration among teachers,<sup>7</sup> and the possibility that teachers could alter their practices in such a way that they could increase student test scores without similarly improving student academic proficiency. On the other hand, some proponents of these policies suggest that altering the payment structure could also improve instruction by recruiting higher quality workers into the teaching profession. To be certain, we still have a great deal to learn about the overall impact of performance pay policies on student proficiency and on the teaching workforce.

This paper provides new estimates indicating that rewarding teachers for the academic gains of their students improves academic proficiency. Our results are consistent with the limited previous research and add to this small but growing literature and suggest that performance pay is a promising avenue for improving student proficiency.

---

<sup>7</sup> A survey of teachers participating in the ACPP suggests that the program did not lead to decreases in collaboration in treated schools (Barnett, et. al., 2007).

## References

- Ballou, D. 2001. Pay for Performance in Public and Private Schools. *Economics of Education Review*, 20, p. 51-61.
- Barnett, J., Ritter, G., Winters, M., and Greene, J., 2006. Evaluation of Year One of the Achievement Challenge Pilot Project in the Little Rock School District. Department of Education Reform, University of Arkansas. Unpublished working paper.
- Clottenfelter, C., and Ladd, H., 1996. "Recognizing and Rewarding Success in Public Schools" in H. Ladd, ed. *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, D.C., Brookings Institution.
- Eberts, R., Hollenbeck, K., and Stone, J., 2002. Teacher Performance Incentives and Student Outcomes. *Journal of Human Resources*, 37, p. 913-27.
- Figlio, D., and Kenny, L., 2006. Individual Teacher Incentives and Student Performance. *Journal of Public Economics*, doi: 10.1016/j.jpubeco.2006.10.001, forthcoming.
- Glewwe, P., N. Ilias, and M. Kremer 2003. *Teacher Incentives*. NBER working paper 9671.
- Heneman, H. G., and Milanowski, A. T., 1999. Teachers' attitudes about teacher bonuses under school-based performance award programs. *Journal of Personnel Evaluation in Education*, 12, p. 327-41.
- Horan, C. B., and Lambert, V., 1994. Evaluation of Utah career ladder programs. Beryl Buck Institute for Education. Utah State Office of Education and Utah State Legislature.
- Jacobson, S. L. 1992. Performance-related pay for teacher: the American experience. In Tomlinson, H. (Ed.) *Performance-related pay in education* (pp. 34-54). London: Routledge.
- Johns, H.E. (1988). Faculty perceptions of a teacher career ladder program. *Contemporary Education*, 59(4), 198-203.
- Keys, B., and Dee, T., 2005. Dollars and Sense. *Education Next*, 5, p. 60-67.
- Lavy, V. 2002. Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. *Journal of Political Economy*, 110, p. 1286-1317.
- Murnane, R., and Cohen, D., 1986. Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and Few Survive. *Harvard Educational Review*, 56, p. 1-17.

## Tables

Table 1  
Mean Descriptive Statistics: Comparison of Treatment and Control School Demographics and Baseline Test Score Performance

	All Schools	All Treated	Main Treated	Control
Black	0.84	0.81	0.77	0.87
White	0.04	0.05	0.03	0.03
Hispanic	0.10	0.12	0.18	0.08
Other Race	0.02	0.01	0.01	0.02
Male	0.51	0.50	0.50	0.51
Participates in Free Lunch	0.18	0.19	0.17	0.18
Gap	0.5	0.53	0.54	0.48
Baseline Math Test Score (NCE)	36.13	35.5	35.13	36.59
N	374	158	176	216

Table 2  
Illustration of Little Rock Testing Schedule During Evaluation Period

Grade, 05-06	Grade, 04-05	Grade, 03-04	Grade, 02-03
5	4	not tested	2
4	3	2	not tested
3	2	not tested	not tested
2	1	not tested	not tested
1	K	not tested	not tested

Table 3  
School-level Difference-in-Differences Estimation

	Coefficient (Robust Error)	P-Value
Constant	-8.98 (1.07)	0.004**
Control School 1	4.60 (0.38)	0.001**
Control School 2	-3.71 (0.33)	0.002**
Treat School	0.03 (0.40)	0.947
Year 2005	-2.80 (0.70)	0.028*
Black	8.29 (1.29)	0.008**
Hispanic	6.84 (0.81)	0.003**
Other Race	6.79 (5.21)	0.284
Male	-.058 (1.24)	0.673
Free Lunch Participant	1.81 (1.41)	0.289
Gap	24.28 (1.64)	0.001**
Treat School	3.64 (0.42)	0.003**

N = 608

R-Squared = 0.1776

Estimated via OLS with robust standard errors clustered by school.

\* indicates significant at .05 level

\*\* indicates significant at .01 level

Table 4  
Differences-in-Differences Estimation with Student Fixed-Effects

	Coefficient (Robust Error)	P-Value
Constant	-1.66 (0.14)	0.001**
Year 2005	-6.94 (0.72)	0.002**
Gap	32.88 (1.43)	0.000**
Treat School	4.19 (0.23)	0.000**

N = 608

Within R-Squared = 0.2351

Estimated with the Within estimator in STATA with robust standard errors clustered by school.

\* indicates significant at .05 level

\*\* indicates significant at .01 level

Table 5  
Individual Fixed-Effects; Number of Years Treated

	Coefficient (Robust Error)	P-Value
Constant	-3.52 (0.53)	0.003**
Year 2005	-5.11 (0.72)	0.058
Gap	30.34 (2.83)	0.000**
Years_Treat	4.63 (0.49)	0.001**

N = 748

Within R-Squared = 0.2227

Estimated with the Within estimator in STATA with robust standard errors clustered by school.

\* indicates significant at .05 level

\*\* indicates significant at .01 level