Teacher Performance Incentives
And Student Outcomes

Randall Eberts
Kevin Hollenbeck
Joe Stone

ABSTRACT

Although merit pay systems have been established in many school districts across the United States, little empirical evidence exists concerning their influence on student achievement. This paper reviews that evidence and presents case study evidence from a county where one high school piloted a merit pay system to reward student retention while another comparable high school maintained a traditional compensation system. A difference-in-differences analysis implies that merit pay increased retention, had no effect on grade point averages, reduced average daily attendance rates, and increased the percentage of students who failed.

I. Introduction

Public discontent with the performance of U.S. public elementary and secondary schooling in raising student achievement is deep. International comparisons of test scores place the United States in the lowest echelons, and comparative analyses of student achievement across regions and demographic groups point out sizeable gaps. Urban schools appear to be the worst performers. In response, reformers have advocated incentive-based or market-driven educational reforms to
improve school quality, such as merit pay for teachers. Merit pay plans have been implemented in many places and the concept has been around for many years. Yet, there is surprisingly little evidence of their effectiveness in raising student achievement.

Why haven’t incentive-based reforms such as merit pay produced better results? Some would say that the implementation has not gone far enough. Merit pay increases have not been large enough to compensate teachers for the risk involved or to induce them to excel in the classroom. Teacher unions, which represent about two-thirds of the nation’s elementary and secondary teachers, may also thwart the success of market-based reforms.¹ There is strong sentiment, particularly among those advocating market-based reforms, that the goals of teacher unions, as manifested in collective bargaining agreements, are not aligned with the goal of improving student performance.²

A third explanation may lie in the inherent nature of the educational process. Education involves multiple stakeholders, disparate and conflicting goals, complex and multitask jobs, team production, uncertain inputs, and idiosyncratic elements contingent on the attributes of individual students, efforts and attitudes of fellow teachers, and classroom environments. The complexity of the process may tend to mitigate student achievement effects of reforms based on individual incentive pay.³

A sizeable literature on incentives demonstrates that education does not differ from the private sector all that much in the incidence of performance-based compensation. Only a small proportion of jobs in the private sector base compensation on explicit contracts that reward individual performance. Rather, private sector companies prefer to reward individuals based on discretionary subjective measures of performance or to follow bureaucratic rules that establish job grades and promotion criteria.⁴ Furthermore, private sector businesses reward workers more through promotions and group-based merit systems than through individual merit rewards (Prendergast 1999).

The purpose of this paper is twofold: (1) to review the scant evidence on the effectiveness of individual merit pay on student achievement and (2) to present new empirical evidence from a case study. Merit pay reforms attempt to incorporate individual pay-for-performance systems into school compensation systems. Many of these systems have been introduced in nonunion schools; a few have been successful in gaining union approval. Even so, our review finds very little empirical evidence concerning the influence of merit pay on student achievement. To begin to fill this gap, we have acquired student achievement data from a high school that is operating

---

¹ For example, Ballou and Podgursky (1997), pp. 107–108, present this argument.
² See Eberst and Stone (1984, 1986, 1987, 1991) and Stone (1998) for empirical evidence that counters this sentiment. These studies find evidence that collective bargaining agreements are positively correlated with average student achievement, although negatively correlated with achievement for low- or high-achieving students. Hoxby (1996) reports similar results, when one considers those students likely to drop out as low academic achievements. Milkman (1989) follows a similar approach for twelfth-grade high school students. Argys and Rees (1995) also find small but significantly positive union effects on student achievement.
³ See Murnane and Cohen (1986).
⁴ Others have cited problems with successfully implementing a merit pay system even in private business. A study in the early 1980s on the topic found that the practices of merit pay in private industry are neither as common nor effective as many believe (Lawler 1983).
a merit pay system and a high school that uses a more traditional compensation system that we argue has a comparable student population. We describe the structure and operation of the merit pay plan and use a difference-in-differences approach to examine its effects on student outcomes. Although the sample size is small, we submit that the juxtaposition of a merit pay system and a traditional compensation system within similar high schools drawing from the same population offers a rare opportunity to begin to evaluate the effects of such a system.5

II. Performance-based Compensation Systems for Individual Agents

The nature of the educational process complicates and confounds the effectiveness of individual performance-based compensation systems.6 The simple, static principal-agent model that Prendergast (1999) explicates rewards agents for taking on additional risk through a pay-for-performance contract with higher (mean) wages. In his model, the performance measures used are noisy and the efficacy of the incentives depends on the risk aversion of the agents. As is well known, incentives may result in unintended, sometimes perverse, consequences. Prendergast uses the term “dysfunctional behavioral responses”; Murnane and Cohen (1986) call it “opportunistic behavior.” We explore these types of consequences below in the context of teaching, although it is the case that they do not necessarily outweigh the potential efficacy of incentive-based pay schemes.

Institutional factors that may result in such dysfunctional responses include poorly defined or poorly measured outcomes leading to a reliance on subjective evaluations that may be flawed, multitasking by job incumbents, team production, and multiple principals/stakeholders. An obvious condition that must hold for pay-for-performance schemes is that performance is measured and that both the agent and principal understand the measure. But measurement may be costly or difficult. Occupations where “output” has an important quality dimension that is subjectively measured (for example, design, or arts and entertainment) or whose output is quality of life (for example, most medical or related occupations) are most likely to rely on subjective evaluations of performance.

Subjective evaluations may be flawed because (a) evaluators may be subject to a moral hazard problem if they are being compensated based on their performance, and they can claim part of the evaluatee’s performance as their own; (b) individuals being evaluated may engage in nonproductive activities to curry favor with their evaluators; or (c) evaluators may end up with distributions of ratings that are compressed because of a reluctance to give very high or very low ratings (Prendergast 1999, pp. 29–31). Jobs in which the incumbents perform many different tasks also strain an incentives-based compensation contract. First, multiple tasks imply multiple performance measures, some of which may have measurability problems. Second, if performance measures are skewed in how they relatively weight various tasks,

5. The empirical analyses of our case study data are meant to be suggestive only. A rigorous evaluation of the efficacy of merit pay would require a careful ex ante design and a much larger sample.

6. Much of the argument presented here was also presented in Murnane and Cohen (1986).
then the agent may respond by investing too much effort into the tasks that receive the most weight in the performance measurement system.

Team production introduces the ‘‘1/n’’ problem, in which each individual’s contribution (and reward) is diluted by the size of the team. Furthermore, if the individuals’ contributions to the team are costly to observe or measure, then team-based incentives may lead to free riders. The problems for an incentive-based compensation system when there are multiple stakeholders come from a potential for misalignment of organizational goals. In effect, the principal-agent arrangement becomes a ‘‘principals’’-agent problem, and multiple principals may have different, and conflicting, goals that they want followed. For example, training directors and production supervisors may conflict with each other on how to reward an individual’s (paid) time spent in training activities.

Each of these four constraints on the effectiveness of incentives-based compensation—need for reliance on subjectively measured outcomes, multiple tasks undertaken by incumbent workers, team production, and multiple stakeholders—appears to characterize the teaching and learning process in schools. Learning outcomes can be and are assessed through standardized tests, which are amenable to performance-based contracts (particularly if value-added measures are available). However, there are many additional dimensions to student learning and development that are either not assessed or are assessed but not with standardized instruments, so evaluations must be inherently subjective.

Schools (at all levels of the K–12 system) typically have dozens of learning processes or programs going on simultaneously. These might include core academic subjects; noncore academic subjects such as art, physical education, music; acquisition of technology skills; career development; special education; extracurricular offerings; gifted and talented programs; human growth and development; and remediation or developmental education. At some levels of education (for example, elementary grade levels), teachers may individually be involved in virtually all of these programs. Even at secondary school levels, which are typically organized by academic discipline, teachers get involved in areas such as career development, extracurricular leadership, and special education (as mainstreaming becomes the status quo). But the myriad of programs or processes that occur in schools is only one source of multitasks. Even within a teacher’s discipline, multiple tasks constitute the teaching and learning process—curriculum development and planning, instruction, and assessment, for example. Furthermore, good teaching requires attention to students’ learning styles, which may mean multiple modes of instruction.

Education, to some extent, requires team production. For example, many elementary and middle schools are organized into teams of teachers. The notion at the elementary level is to rely on teachers’ comparative advantages in core academic areas so that each teacher in a team is responsible for the subjects with which they are most comfortable. Apart from explicit team teaching, however, departmentalized secondary schools result in team production because students’ performances on standardized tests depend on learning in several courses taught by different teachers.

Finally, school governance and control is characterized by many different stakeholders with differing, and sometimes, conflicting goals. Administrators who are accountable for direct student achievement may be most responsive to levels of test scores. School board members who are accountable for resource decisions may be
most interested in changes (value added) over time in test scores. Parents may be 
most concerned about postsecondary education attendance rates, whereas employers 
may be most concerned about "soft" employability skills such as problem-solving, 
attendance, and attitude.

Another characteristic of most school districts is that they have very little control 
over their revenue streams. As noted, incentive-based contracts allocate part of the 
production risk to the employees in return for higher rewards (wages). Because 
school administrators have little revenue to share, however, they cannot offer size-
able increases in pay, even if teachers were willing to accept the risk inherent in a 
merit pay system.\(^7\) Furthermore, increased teacher performance does little to add to 
the revenue of a school district, as meeting or exceeding sales quotas would do for 
a private company, except perhaps for attracting more students into a district or 
 improving attendance rates, which may be the basis for revenue.\(^8\)

In short, while empirical evidence and common sense show clearly that economic 
actors respond to incentives, there may be several wedges between performance 
measures and the actions of teachers that tend to mitigate against individual level, 
incentive-based compensation schemes—just as they do in large parts of the private 
sector, as well. In the end, of course, the net result of these forces is an empirical 
issue.

There is little empirical evidence on the effects of merit pay on student achieve-
ment. Most of the literature on merit pay systems documents the institutional expe-
riences in districts. Those experiences, for the most part, have been rather short-lived 
and usually negative. For example, a major study of merit-based pay (Hatry, Greiner, 
and Ashford 1994) found that most (75 percent) merit pay programs that had existed 
in 1983 and had been studied by the researchers were no longer operational in 1993.\(^9\) 
An interesting self-described limitation of the Hatry et al. (1994) study is that they 
did not examine student achievement. They note:

We would especially have liked to have performed an in-depth analysis of the 
impact of incentive programs on student achievement. However, very few of 
the participating districts had attempted any systematic evaluation of the effects 
of their incentive plans on student achievement, even though a basic assumption 
behind incentive plans is that teachers can indeed significantly affect learning 
(pp. 7–8).

In a study involving one district in Pennsylvania, Tulli (1991) found no correlation 
between gains in student achievement and teachers awarded merit bonuses under 
this district’s plan. The author noted, however, that this district heavily weighted 
inputs such as attendance, participation in professional development, and supervision 
of extra activities relative to student outcomes.

---

7. Since collective bargaining seems to impose added costs to a district, perhaps these costs could be identified, captured, and reallocated as merit pay (Eberts and Stone 1991). Also market-based alternatives to traditional public schools may be less reliant on state or local government funding, and therefore may be in more control of their revenue.
8. A reviewer suggested that improved teacher performance may reduce costs so that revenue could then be allocated more productively. The reviewer cited less remedial learning and lower discipline costs.
9. Murnane and Cohen (1986) also emphasize the short-lived nature of merit pay systems.
III. A Case Study of a Merit Pay System

We have acquired data from a particular high school that implemented a merit pay system in 1996 and a "comparable" high school that maintained a traditional compensation system. What is interesting about the school that adopted merit pay is that it is the only nonunionized school building in its local public school system. The two schools make an interesting case study for two reasons: (1) they are comparable, as described below, yet one school adopted a merit pay system and the other didn't, and (2) the school that adopted merit pay is nonunionized whereas the other school is unionized. The Michigan district where the merit pay high school is located has an enrollment of about 9,000 students and has 15 school buildings; three of which are high schools. Community High School, which implemented the merit pay system, is an alternative education facility that has an enrollment of approximately 500 students pursuing a high school diploma. The school also has an adult basic education enrollment of about 100 students pursuing GED preparation, a high school degree, or English as Second Language programs. Community High School has approximately two dozen faculty members.

This school implemented its merit pay system at a time of great transition in adult education and K–12 educational funding in Michigan. In 1995, the state legislature took two actions that influenced adult and alternative education institutions in the state. First, in the state budget, the legislature virtually "zeroed out" all state funds for adult education and transferred those funds into the state economic development agency to be used for customized training. Second, the state passed a comprehensive public education finance reform that moved the state from a system funded primarily at the local level by property taxes to a system which is state-funded, primarily through sales taxes. The new state finance system funds "pupils" (students under age 21 in elementary or secondary schooling and pursuing a program leading to a high school diploma) at a foundation allowance that is nearly identical for all districts. Given that the state funds alternative education pupils at approximately $6,000 per year (expressed in 2000 dollars), Community High School decided to disengage from the adult education consortium it participated in prior to 1995 and to focus primarily on alternative education for high school students. In the same year that the legislature revised the school finance system, it also passed a law that allows school districts to operate pilot programs that may be exempt from collective bargaining requirements.

The district decided to operate Community High School as a "pilot" program with a performance-based compensation scheme for its teachers, who collectively decided to remain separate from the local district's education association (union). Alternative education settings are characterized by students who have often not succeeded in traditional school settings, and usually experience attendance problems and intermittent dropping out and reenrollment episodes. Consequently, the performance-based incentives are targeted on student retention.

10. Unfortunately, the data only contained course-related information such as grades and daily attendance. They did not include any information about the students other than ID number.
A. Description of the Performance-based Compensation System

Teachers are paid a base wage for each 60-minute class that they teach. The base wage depends solely on their own educational attainment (not on experience or tenure). Teachers with a master’s degree or higher receive a 5 percent higher base pay than teachers with a bachelor’s degree. The merit pay system offers two supplements that may be earned and added to the base pay. A retention bonus is paid if 80 percent or more of the students assigned to the class (as of the end of the second week of the quarter) are still enrolled and attending at the end of the quarter. The student count at the end of the course is taken on a day during the last week of the quarter determined by the principal and supposedly unbeknownst to the teacher. The bonus is the same for all teachers, no matter what their educational background, and it is approximately 12.5 percent of the base for teachers with bachelor’s degrees (12 percent for master’s plus).

The second supplement is based on student evaluations. Students rate the following 15 factors on a 1–5 scale:

- Objectives, requirements, and expectations for the class were clearly started.
- The instructor was well-informed and had current knowledge of subject matter.
- The instructor was well-prepared for each class.
- The instructor presented the material clearly.
- The material was presented in an organized manner.
- The instructor used class time well.
- The instructor was interested in students and willing to listen to them.
- The instructor encouraged student participation and welcomed questions, discussion, and different points of view.
- The instructor encouraged a high level of student attendance in the class.
- The methods of evaluating student progress and performance were clearly stated.
- The instructor encouraged students to think critically.
- The instructor was enthusiastic.
- The instructor was aware of the varying levels and abilities of students.
- The materials used were appropriate for the subject taught.
- This class was worthwhile.

11. The initial enrollment in the class for purposes of calculating retention is capped at 20, so to earn the retention bonus, teachers must have 16 students or 80 percent of the initial enrollment at the end of the term, whichever is less. Sometimes actual class sizes exceed 20. In these cases, the retention bonus is still earned if the ending enrollment is 16 or more. This was seen as an incentive for teachers to allow larger enrollments in their classes, when it was warranted by scheduling and overall enrollment concerns.
Teachers who receive an average rating of 4.65 or higher\textsuperscript{12} on the five-point scale for all 15 items in all of their classes (weighted by class enrollment) in each quarter for four (4) consecutive quarters receive the performance bonus, which increases their base pay by about 5 percent \textit{and} increases their retention bonus by 10 percent.\textsuperscript{13} To give a sense of the size of these bonuses, during school year 1998–99, the base pay for a beginning teacher with a bachelor’s degree was $816 per class ($22,848 for nine months; four quarters with seven classes).\textsuperscript{14} With the performance bonus and retention bonuses in all classes, the per-class pay would be $979 ($27,412 for nine months; four quarters with seven classes).\textsuperscript{15}

\textbf{B. Impacts on student outcomes}

Did the performance incentives affect student outcomes at this school? To answer this question, we analyzed data from students at this school and at a similar alternative education high school in the same county that relies on a traditional experience/education compensation scheme. Both schools have open enrollment within the county, and so they draw from the same population of students. In particular, we have obtained data for a five-year period (1994–95 through 1998–99) that encompasses two years prior to and two years after the implementation of the performance incentive system. We perform a difference-in-differences analysis of several student outcomes including course completion, class attendance, grade point average, and passing rates conditional on course completion.\textsuperscript{16} The grade point average (GPA) is calculated from student-level data; the other three outcomes—attendance, completion, and conditional passing—are calculated from course-level data. The data are for alternative education high school students only; they do not include adult education students.

With no data on detailed student characteristics, we relied on the judgement of building administrators and district educators in selecting the best local alternative school to use as a comparison site. Both schools are located in the same county, but not in the same district. The districts are both suburban districts. Both schools enroll students from throughout the county, but the majority of their students come from within the district boundaries where the schools were located. Educators familiar with both schools indicated that the schools were comparable in course offerings, student socioeconomic characteristics, and funding levels. The individuals making this comparison consistently pointed out how the student characteristics at both

\textsuperscript{12} This was the average rating in the previous school year.

\textsuperscript{13} Hary et al. (1994) found a range of merit pay awards in their study from at most 25 percent of salary to 5 percent or less. See also Lawler (1983).

\textsuperscript{14} Community High School is on an eight-period per day schedule, and the average teaching load is 7 classes.

\textsuperscript{15} Many teachers have more than six classes per term. With at least six, the teachers receive full benefits equivalent to the unionized teachers in the district.

\textsuperscript{16} The difference-in-differences technique “differences out” time-invariant causal variables and assumes that there is no interaction between the “treatment,” in this case merit pay, and time-varying causal variables. In short, it is appropriate in this case only if both high schools' student characteristics, curriculum and instruction, and outside external factors such as the local economies changed similarly. Unfortunately, the small sample size and data deficiencies did not allow formal testing of these assumptions.
### Table 1

<table>
<thead>
<tr>
<th></th>
<th>School A (Merit Pay)</th>
<th>School B (Traditional Pay)</th>
<th>Difference (School A–School B)</th>
<th>Difference-in-Differences (Post–Pre or School A–School B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre (1994–95)</td>
<td>50.92</td>
<td>39.02</td>
<td>11.90</td>
<td>(1.29)</td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.67)</td>
<td>(1.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>5,316</td>
<td>2,004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post (1998–99)</td>
<td>71.66</td>
<td>54.47</td>
<td>17.19</td>
<td>(0.95)</td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.58)</td>
<td>(0.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>6,171</td>
<td>4,318</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (Post–Pre)</td>
<td>20.74</td>
<td>15.45</td>
<td>5.29</td>
<td>(1.60)</td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.89)</td>
<td>(1.33)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Archive files from attendance and gradebook software used at School A and School B.  
Note: Standard errors are calculated under the assumption that there is no covariance between the two districts. This assumption places an upper bound on the standard errors, since any positive covariance which would be expected, would lower the standard errors.

schools differed from the characteristics of students at the third alternative high school in the county that drew from an urban population.

The results are consistent with the expectations regarding the effect of incentives on teacher behavior. As previously noted, the merit pay system directly rewards teachers for the number of students still enrolled at the end of the course. Table 1 presents the analysis of this outcome. The pre- and post-implementation data are for the 1994–95 and 1998–99 school years, respectively, which are two years prior to and two years after implementation of the merit pay system in School A.17 The entries in Table 1 are the percentage of students who completed a class after enrolling in it. The number who did not complete a class may have included students who transferred to another class. There is no way to trace transfers in the student data base system. Interviews with administrators, however, suggest that transfers were quite rare at both high schools.

While the course completion percentages increased in both schools over the five years, the increase was quite dramatic in School A, as would be predicted. Prior to the merit pay system, in which compensation is partially based on the number of students who complete courses, about half the School A students completed their courses (note that both schools are on the quarter system). After implementing the merit pay system, the fraction increased by more than 40 percent to almost three-quarters.

Daily attendance, on the other hand, was not rewarded (except that students had to be present during the last week of classes to be considered a completer). Table 2

---

17. Results from analyses using data that are one year prior to implementation of the merit system and one year after implementation are similar in magnitude, sign, and statistical significance.
Table 2
Difference-in-Differences Analysis of Course Daily Attendance

<table>
<thead>
<tr>
<th></th>
<th>School A (Merit Pay)</th>
<th>School B (Traditional Pay)</th>
<th>Difference (School A–School B)</th>
<th>Difference-in-Differences (Post–Pre or School A–School B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre (1994–95)</td>
<td>59.02</td>
<td>34.24</td>
<td>24.78</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.41)</td>
<td>(0.69)</td>
<td>(0.86)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5,316</td>
<td>2,004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post (1998–99)</td>
<td>58.62</td>
<td>36.33</td>
<td>22.29</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.35)</td>
<td>(0.47)</td>
<td>(0.58)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>6,171</td>
<td>4,318</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (Post–Pre)</td>
<td>-0.40</td>
<td>2.09</td>
<td>-2.49</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.54)</td>
<td>(0.83)</td>
<td>(1.04)</td>
<td></td>
</tr>
</tbody>
</table>

Source: Archive files from attendance and gradebook software used at School A and School B.

Note: Standard errors are calculated under the assumption that there is no covariance between the two districts. This assumption places an upper bound on the standard errors, since any positive covariance which would be expected, would lower the standard errors.

presents our analysis of the daily attendance of students. The entries in the table are percentages of students attending class among those who are still enrolled. The entries represent class average daily attendance rates (adjusted for enrollment) averaged over all classes in the school for the entire academic year. Again, interviews with administrators confirmed that the very low rates of attendance are correct. Unlike its effect on course completion, the merit pay system appears to have little effect on daily attendance rates. The comparison site’s (School B) average daily attendance rate went up slightly, while School A’s attendance rate stayed approximately the same in the two analysis years. Note that as shown in Table 2, course completion rates were higher in School A, so the difference-in-differences in actual average levels of classroom attendance may be positive. Nevertheless, in Table 2, which shows rates and not levels, the results are the opposite of what one would expect if the merit pay system had been based on average daily attendance rates and not simply on whether students were in class during the final week of the term when the head count was taken.

Table 3 shows the analysis of student achievement as measured by individual grade point averages (GPA). The (student) average GPA in both schools declined over the five-year period, but the decline in School A of 0.53 points was greater in magnitude than the decline of 0.37 points in School B, although this difference is statistically significant.18 Interviews with administrators in both schools confirmed that the change in the funding mechanism as well as secular trends have caused an increase in “harder-to-serve” students at both alternative high schools, so they were well aware of the declines in GPA. The fact that the decline in School A was greater

18. The pre-merit pay difference in student GPA levels is substantial and weakens our confidence in the comparability of the schools. Nevertheless, we note that the difference is consistent with the Eberts and Stone (1984) evidence that unionization may have a negative impact on lower achieving students.
Table 3

<table>
<thead>
<tr>
<th></th>
<th>School A (Merit Pay)</th>
<th>School B (Traditional Pay)</th>
<th>Difference (School A–School B)</th>
<th>Difference-in-Differences (Post–Pre or School A–School B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre (1994/95)</td>
<td>2.71</td>
<td>2.19</td>
<td>0.52</td>
<td>-0.16</td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.04)</td>
<td>(0.08)</td>
<td>(0.09)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Number of</td>
<td>392</td>
<td>324</td>
<td></td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post (1998/99)</td>
<td>2.18</td>
<td>1.82</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>Number of</td>
<td>578</td>
<td>304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (Post–Pre)</td>
<td>-0.53</td>
<td>-0.37</td>
<td></td>
<td>-0.16</td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.06)</td>
<td>(0.09)</td>
<td></td>
<td>(0.11)</td>
</tr>
</tbody>
</table>

Source: Archive files from attendance and gradebook software used at School A and School B.

Note: Standard errors are calculated under the assumption that there is no covariance between the two districts. This assumption places an upper bound on the standard errors, since any positive covariance which would be expected, would lower the standard errors.

than the decline in School B is consistent with the hypothesis that the merit pay incentive resulted in higher retention of lower-achieving students, who were most likely to drop out.

Finally, we looked at the percentage of students who passed the course given that they completed the course, defined as not receiving a failing or incomplete grade. Table 4 displays our analysis of these data. Consistent with the GPA data, the percentage of students actually passing their courses declined over the period of analysis. Again, the decline was far larger for School A, which went from approximately 93 percent to 75 percent. That school’s decline in the percentage of students passing the course conditional on completion is more than 6 percentage points greater than School B’s. Again, this is consistent with the hypothesis that School A is retaining, on average, more low-achieving students.

C. Discussion of the results

The merit pay system, established in School A, was intended to increase student retention. In discussion with the school’s administrators, we determined that a further goal was to increase student achievement. Their reasoning was that increased achievement would result from increased retention.

The outcomes of this merit pay system illustrate the difficulty of instituting such a compensation system in schools. First, the output measure has to be easily, inexpensively, and accurately determined and it has to be agreed upon up front. In this case, the administrators of the high school knew that they wanted to increase retention. Thus a student was considered to be retained in a class if the student was present during a randomly selected day of the last week of classes when the principal would
Table 4
Difference-in-Differences Analysis of Course Passing Percentages for Students 
Who Completed the Course

<table>
<thead>
<tr>
<th></th>
<th>School A (Merit Pay)</th>
<th>School B (Traditional Pay)</th>
<th>Difference (School A–School B)</th>
<th>Difference-in-Differences (Post–Pre or School A–School B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre (1994–95)</td>
<td>93.35</td>
<td>76.47</td>
<td>16.88</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.48)</td>
<td>(1.52)</td>
<td>(1.59)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2,707</td>
<td>782</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post (1998–99)</td>
<td>75.67</td>
<td>65.21</td>
<td>10.46</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.65)</td>
<td>(0.96)</td>
<td>(1.50)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>4,422</td>
<td>2,489</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (Post–Pre)</td>
<td>-17.68</td>
<td>-11.26</td>
<td>-6.42</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.80)</td>
<td>(1.79)</td>
<td>(1.96)</td>
<td></td>
</tr>
</tbody>
</table>

Source: Archive files from attendance and gradebook software used at School A and School B.
Note: Standard errors are calculated under the assumption that there is no covariance between the two districts. This assumption places an upper bound on the standard errors, since any positive covariance which would be expected, would lower the standard errors.

make an unannounced visit to the class and count how many students were present. If the count was greater than or equal to the minimum standard (80 percent of the starting enrollment of 16), then the teacher was awarded her bonus. The incentive "worked" by this measure of output as demonstrated in Table 1, which showed a significant increase in the percentage of course completers.

However, the second difficulty is that the output measure should be the organization's final product, or at least highly correlated with the final product. In this case, the definition of final product is ambiguous. If we assume that the school is concerned with maximizing revenue, then it should be concerned about student attendance during the period when the state measures enrollment upon which its aid is based.19 Unfortunately, the results are ambiguous. Table 2 shows that the merit pay system had a deleterious effect on student daily attendance rates, but Table 1 indicates an increase in course completion rates. At best, there may have been a small positive impact on average daily attendance.

If we assume that the school is interested in student learning, however, the results are more definitive. Indeed, the administrators expected that the incentives of their merit pay system were appropriately aligned to course completion, and consequently to improved achievement. The incentives apparently worked for the goal of course completion as show in Table 1, but average achievement seems to have deteriorated. Unfortunately, neither school conducted any sort of standardized testing of the entire student body, so there was no "testing" measure of learning achieved. In Table 3, we examined student grade point averages as a proxy for student achievement, and

19. In some states, aid is based on average daily attendance (ADA); in Michigan, where this case study is located, the aid is based on enrollment on a particular "count day" in fall and in winter.
in Table 4, we examined passing rates as another proxy. In both cases, the measure declined more in the school with merit pay (although the decline in grades was not significant).

Of course, changes in averages may be confounded by compositional changes, so the results do not necessarily imply that academic achievement declined for each student relatively more in that school. This effect can be illustrated by positing three kinds of students—high-achieving students, denoted by \( H \); low-achieving students who complete courses, denoted by \( LC \); and low-achieving students who drop out of courses, denoted by \( LD \). Our student achievement measures are based only on \( H \) and \( LC \) students. If the merit pay system caused \( LD \) students to become \( LC \) students, then overall grade point averages and passing rates will drop, but student learning may increase if the academic outcomes of \( LC \) students exceed the academic outcomes of \( LD \) students. The data are consistent with this sort of outcome in School A, so that the reductions in GPA and passing rates may not be all that bad.

On the other hand, the reductions in GPA and passing rates may be reasons for alarm. Administrators in School A provided anecdotes that suggested that teachers were altering their instructional style and course content in order to make their courses more interesting to and well liked by students. The teachers were trying to entice students who would otherwise have dropped out to stay in the course to ensure that they would earn their student retention bonus, and they were trying to get better student evaluations, which is the second component of the merit pay plan. Anecdotes included activities such as more field trips and in-class parties. If the instructional and content changes made by instructors as a result of the implementation of the merit pay system resulted in less rigorous curriculum, then the GPA results will underestimate the true decline in student learning. Furthermore, if there is a significant peer effect in learning, it may be the case that the presence of students who otherwise would not have completed the course reduced the learning of other students. Finally, if the students who otherwise would not have completed the course exert less effort and do not learn as much, then student achievement would not be increased by their retention.\(^{20} \)

A final note about the results concerns the difference in collective bargaining status at the two schools. Could the results be attributed to that difference? Because there was no change in collective bargaining status at either school over the period of analysis, the difference-in-differences approach should net out the influence of unionization on the outcomes. It is still conceivable that the nature of the collective bargaining process could have changed during the period coincidentally with the implementation of merit pay at the nonunionized high school. For example, the significant change in state financing of education that occurred at about that time might have strengthened or weakened the collective bargaining process. However, no evidence or argument has been put forward to suggest that the state financing change has had an effect on bargaining in the state.

---

20. Clearly, one should be cautious in generalizing our results. We examine the relative experiences of two (alternative) high schools and a two-year post-treatment period. However, the anecdotal evidence suggests that the two-year period of time (actually three years after implementation) was adequate to make substantial adjustments in curriculum and instruction. One teacher left the staff because s/he was unwilling to teach under the merit pay system.
IV. Conclusion

In summary, this case study of the implementation of a merit pay system in a specific high school suggests that incentives do “work.” The merit pay system is directly targeted at student retention, as defined by a measure understood and agreed upon by both teachers and administrators. The evidence is consistent with the implementation of the merit pay system resulting in higher student retention, as defined by attendance during the last week of classes. The administrators who implemented the system also had anticipated that other desired outcomes would follow. However, the study suggests that these outcomes were not achieved, and unintended consequences may have arisen as a direct result of the success of the merit pay system. Student grade-point averages, daily attendance rates, and course passing rates declined (although the grade point average change was not significant). There was also anecdotal evidence that suggested that course content was diluted.

Therefore, the results suggest that pay for performance incentives can motivate agents to produce outcomes that are directly rewarded. However, the study also suggests that incentive systems within complex organizations, such as schools, with multiple tasks and outcomes, team production, and multiple stakeholders, may produce results that are unintended and, at times, misdirected—unless carefully constructed and implemented.

References