D R A F T

Not for Quotation

TEACHER PERFORMANCE INCENTIVES AND STUDENT OUTCOMES

Randall Eberts, Executive Director
W.E. Upjohn Institute for Employment Research

Kevin Hollenbeck, Senior Economist
W.E. Upjohn Institute for Employment Research

Joe Stone, W.E. Miner Professor
Department of Economics
University of Oregon

Version 1.5

August 2000

The paper was prepared for a National Academy of Sciences Conference entitled "Devising Incentives to Promote Human Capital," held December 17-18, 1999, in Irvine, California. The authors appreciate the helpful comments of Derek Neal, Richard Murnane, and other participants of the conference. The research assistance of Kristine Kracker and Noyna DebBurman is gratefully acknowledged, as is the excellent clerical assistance of Phyllis Molhoek and Claire Black. The views expressed do not necessarily represent the views of The W.E. Upjohn Institute for Employment Research or the University of Oregon. The usual disclaimer applies.

Abstract

This paper reviews the evidence on the effectiveness of individual merit pay systems for teachers on student achievement, and it presents new empirical results based on a system established within a collective bargaining environment. While many merit pay systems have been established in school districts across the U.S., very little empirical evidence concerning their influence on student achievement exists. A natural experiment arose in a county in which one high school piloted a merit pay system that rewarded student retention and student evaluations of teachers while another comparable high school maintained a traditional compensation system. A difference-in-differences analysis implies that merit pay had no effect on grade point averages, reduced the percentage of students who dropped out of courses, reduced average daily attendance, and increased the percentage of students who failed.

The outcomes illustrate the difficulty of instituting individual merit pay in schools. The goal was to increase student retention. 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. The system "worked" by this measure because the school experienced a significant reduction in course noncompleters. However it is not clear that this measure was correlated with student achievement or even average attendance, and indeed, neither of these outcomes were improved.

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 among 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, the efforts and attitudes of fellow teachers, and the classroom environments. The

---

[1]For example, Ballou and Podgursky (1997), pp. 107-08, present this argument.

[2]See Eberts 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 correlated with student achievement.

complexity of the process may tend to mitigate the 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 on the subject. 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. Our review finds very little empirical evidence concerning the influence of merit pay on student achievement. To begin to fill this gap in our knowledge about merit pay, we have acquired student achievement data from a high school that is operating a merit pay system within a unionized district 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

[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).

approach to examine its effects on student outcomes. While 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 evaluate the effects of such a system.

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.[5] 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."

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 (e.g., design, or arts and entertainment) or whose output is a dimension of

[5]Much of the argument presented here was also presented in Murnane and Cohen (1986).

quality of life (e.g., 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 (theft); 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[6] (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, 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

[6]In their case studies, Murnane and Cohen (1986) found that school principals reported causing morale problems by subjectively rating very good teachers as "Excellent," when there was a higher level category of "Outstanding." These problems would be avoided by giving fewer or many more "Outstanding" ratings.

supervisors may conflict with each 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—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 that ultimately evaluations are inherently subjective.

Schools (at all levels of the K-12 system) typically have literally 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 (e.g., 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 comprise 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. However, even apart from explicit team teaching, 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 problemsolving, 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). But since school administrators have little revenue to share, they cannot offer sizeable increases in compensation were teachers willing to accept the risk inherent in a merit-pay system.[7] Furthermore, increased teacher performance does

[7]We do note that collective bargaining seems to impose added costs (inefficiency) to a district. Perhaps these costs could be identified, captured, and re-allocated as merit pay. Also market-based alternatives to traditional public

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.

In short, while empirical evidence and common sense show clearly that economic actors respond to incentives, we argue that there are several wedges between performance measures and the actions of teachers that tend to militate against individual level, incentive-based compensation schemes. In developing this paper, we found little empirical evidence on the effects of merit pay on student achievement. Most of the literature on merit pay systems documents the institutional experiences 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%) merit-pay programs that had been in existence in 1983 and had been studied by the researchers were no longer operational in 1993.[8] An interesting self-described limitation of the study is that they did not examine student achievement. They noted,

> 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.

---

schools may be less reliant on state or local government funding, and therefore may be in more control of their revenue.

[8]Murnane and Cohen (1986) also emphasized the short-lived, ineffective nature of merit pay systems.

III. A Case Study of a Merit Pay System within a Collective Bargaining Environment

To add to the empirical literature on merit pay systems, we have acquired data from a 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 school in its local public school system that is nonunionized. The Michigan district where this high school is located has an enrollment of about 9,000 students and has 15 school buildings, three of which are high schools. The high school that 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. There are approximately two dozen faculty members at the high school.

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 had taken 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), this 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 the 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 who usually experience attendance problems and intermittent drop-out and re-enrollment episodes. Consequently, the performance-based incentives are targeted on student retention.

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.[9] The bonus is the same for all teachers, no matter what their educational background,

[9]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, and 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

and it is approximately 12.5 percent of the base for teachers with Bachelor's degrees (12 percent for Masters 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 was 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.

Teachers who receive an average rating of 4.65 or higher[10] on the 5-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 *and* increases their retention bonus by 10 percent.[11] To give the reader a sense for the size of these bonuses, during school year 1998-99, the base pay for a teacher with a Bachelor's degree was $816 per class ($22,848 for nine months; four quarters with seven classes).[12] With the performance bonus

---

it was warranted by scheduling and overall enrollment concerns.

[10]This was the average rating in the previous school year.

[11]Hatry et al. (1994) found a range of merit pay awards in their study from at most 25 percent of salary to five percent or less. See also Lawler (1983).

[12]The high school is on an eight-period per day schedule, and the average teaching load is seven classes.

and retention bonuses in all classes, the per class pay would be $979 ($27,412 for nine months; four quarters with seven classes).[13]

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, so they draw from the same population of students. In particular, we have obtained data for a five-year period (1994/1995 - 1998/1999) 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 grade point average, class attendance, course completion, and passing rates conditional on course completion. 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.

Table 1 shows the analysis of GPA data. The pre- and postimplementation data are for the 1994/1995 and 1998/1999 school years, respectively, which are two years prior to and two years after implementation of the merit pay system in School 1.[14] The (student) average GPA in both schools declined over the five-year period, but the decline in School 1 of 0.53 points was

[13]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.

[14]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.

greater in magnitude than the decline of 0.37 points in School 2.[15] 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 1 was greater than the decline in School 2 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.

Table 1
Difference-in-Differences Analysis of Grade Point Average

	School 1 (Merit pay)	School 2 (Traditional pay)	Difference	Diff.-in-Diff.
Pre (1994/95)	2.71	2.19	0.52	
Std. error	(0.04)	(0.08)	(0.09)	
No. of obs.	392	324		
Post (1998/99)	2.18	1.82	0.36	
Std. error	(0.04)	(0.05)	(0.07)	
No. of obs.	578	304		
Difference (Post - Pre)	-0.53	-0.37		-0.16
Std. error	(0.06)	(0.09)		(0.11)

What about the drop-out rate itself, which was the outcome targeted by the merit-pay system? Table 2 presents the analysis of this outcome. The entries in table 2 are the percentage of students who did not complete a class after enrolling in it and attending it for at least one day. The percentages do include students who may have attended a class for a short period of time and then transferred to another class. There was no way to trace transfers in the student data base system. However, interviews with administrators suggest that such a phenomenon was quite small at both high schools.

[15]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.

Table 2
Difference-in-Differences Analysis of Course Noncompletion Percentages

	School 1 (Merit Pay)	School 2 (Traditional pay)	Difference	Diff.-in-Diff.
Pre (1994/95)	49.08	60.98	-11.90	
Std. error	(0.67)	(1.09)	(1.29)	
No. of obs.	5,316	2,004		
Post (1998/99)	28.34	45.53	-17.19	
Std. error	(0.58)	(0.76)	(0.95)	
No. of obs.	6,171	4,318		
Difference (Post	-20.74	-12.74		-5.29
Std. error	(0.89)	(1.33)		(1.60)

While the noncompletion percentages decreased in both schools over the five-year period, the decrease was quite dramatic in School 1, 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 students in School 1 completed their courses (note the courses in both schools are on the quarter system). After the implementation of the merit pay system, the fraction increased by almost 50 percent to three-quarters.

An interesting aspect of the merit pay system is that it rewarded teachers for the number of students still enrolled at the end of the course. Attendance was not rewarded (except that a student had to be present during the last week of classes to be considered a completer). Table 3 presents our analysis of the daily attendance of students. The entries in the table are the percentage of students attending class averaged over each class and each reporting period. Again, interviews with administrators confirmed that the very low rates of attendance are correct. The merit pay system appears to have little effect (and, in fact, the sign is negative) on daily attendance. School 1's attendance rate stayed approximately the same in the two years, and School 2's rate actually went up slightly, which is the opposite of what one would expect if teachers were to respond to economic incentives by finding ways to increase overall attendance and not simply during the week the actual class count was taken.

Table 3
Difference-in-Differences Analysis of Course Daily Attendance

	School 1 (Merit pay)	School 2 (Traditional pay)	Difference	Diff.-in-Diff.
Pre (1994/95)	59.02	34.24	24.78	
Std. error	(0.41)	(0.69)	(0.86)	
N	5,316	2,004		
Post (1998/99)	58.62	36.33	22.29	
Std. error	(0.35)	(0.47)	(0.58)	
N	6,171	4,318		
Difference(Post	-0.40	2.09		-2.49
Std. error	(0.54)	(0.83)		(1.04)

Finally, we looked at the percentage of students who passed the course given that they completed the course (i.e., did not receive 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 1, 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 over 6 percentage points greater than School 2's. Again, this is consistent with the hypothesis that School 1 is retaining, on average, more low-achieving students.

Discussion of the results

The merit pay system, established in School 1, 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

Table 4
Difference-in-Differences Analysis of Course Passing Percentages for Students Who Completed the Course

	School 1 (Merit pay)	School 2 (Traditional pay)	Difference	Diff.-in-Diff.
Pre (1994/95)	93.35	76.47	16.88	
Std. error	(0.48)	(1.52)	(1.59)	
N	2,707	782		
Post (1998/99)	75.67	65.21	10.46	
Std. error	(0.65)	(0.96)	(1.50)	
N	4,422	2,489		
Difference(Post	-17.68	-11.26		-6.42
Std. error	(0.80)	(1.79)		(1.96)

classes when the principal would 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 the bonus. The incentive "worked" by this measure of output as demonstrated in table 2, which showed a dramatic reduction in the percentage of course noncompleters.

However, the second difficulty is that the output measure has to be the organization's final product that will be sold, 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 in business to maximize revenue, then it should be concerned about student attendance during the period when the state measures enrollment upon which its aid is based.[16] Table 3 shows that the merit pay system perhaps had a deleterious effect on student daily attendance.

[16]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.

If we assume that the school is in business to maximize student learning, then it may be the case that the merit pay system that was implemented had the appropriate incentives. 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 1, we examined student grade point averages as a proxy for student achievement, and in table 4, we examined passing rates as another proxy. In both cases, the measure declined more in the school with merit pay. However, because of a compositional change, this does 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. The student achievement measures that we use 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 1, 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 great alarm. Administrators in School 1 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. Presumably 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.

Since the change in collective bargaining status in School 1 is coincident with the adoption of the merit pay system, it is possible that our results may reflect not only the effect of the merit-pay system but also the effect of collective bargaining on student achievement and other outcomes. However, closer inspection suggests that collective bargaining may not be a large mitigating factor, since our results are not consistent with the effects one would expect from collective bargaining. For instance, Eberts and Stone (1984) found that low-achieving students have lower student test score gains in districts covered by collective bargaining than do low-achieving students in districts not covered.[17] While the measure of student outcomes may not be exactly comparable between our study and Eberts and Stone's, it is instructive to make the comparison. If collective bargaining did have an effect on our results, we would expect student GPA to decline more in School 2, the traditional pay system with collective bargaining, than in School 1, because School 1 is no longer covered by collective bargaining in the post period but

[17]Hoxby (1996) reported similar results when one considers those student likely to drop out as low academic achievers. Milkman (1989) followed a similar approach for twelfth-grade high school students. Argys and Rees (1995) also found small but significantly positive union effects on student achievement.

School 2 still is covered. Similarly, we would expect the difference in GPA between the two districts to be larger in the post period after merit pay is implemented and collective bargaining eliminated in School 1 as compared to the period before merit pay when both schools are covered by collective bargaining. However, neither of these expectations is supported by the results, suggesting that collective bargaining may not play a major role in explaining our results. Furthermore, the effects of collective bargaining should move in the same direction as the effects of merit pay and both should yield a positive difference-in-differences result. Yet, even with the supposedly reinforcing effects of the two factors, the difference-in-differences result is not positive and statistically significant.

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 had also anticipated that other desired outcomes would follow. However, the case study suggests that these outcomes were not achieved and, in fact, unintended consequences may have arisen as a direct result of the success of the merit-pay system. Student grade-point averages and daily attendance rates were virtually unchanged, and course passing rates declined. 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, which have multiple tasks and outcomes, team production, and multiple stakeholders, may produce results that are unintended and at times misdirected.

References

Argys, Laura M., and Daniel I. Rees (1995). "Unionization and School Productivity: A Reexamination." In *Research in Labor Economics* 14: 49–68.

Ballou, Dale, and Michael Podgursky. (1997). *Teacher Pay and Teacher Quality*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Eberts, Randall W., and Joe A. Stone. (1991). "Unionization and Cost of Production: Compensation, Productivity, and Factor-Use Effects." *Journal of Labor Economics* 9(2): 171–185.

———. (1987) "Teachers' Unions and the Productivity of Public Schools." *Industrial and Labor Relations Review* 40: 355–63.

———. (1986) "Teacher Unions and the Cost of Public Education." *Economic Inquiry* 24: 631–44.

———. (1984) *Unions and Public Schools: The Effect of Collective Bargaining on American Education*. Lexington, MA: Lexington Books.

Hatry, Harry P., John M. Greiner, and Brenda G. Ashford. (1994). *Issues and Case Studies in Teacher Incentive Plans*. Second edition. Washington, DC: The Urban Institute Press.

Hoxby, Caroline Minter (1996). "How Teachers' Unions Affect Education Production." *Quarterly Journal of Economics* 111: 671–718.

Lawler, E.E., III. (1983). *Pay and Organization Development*. Reading, MA: Addison-Wesley.

Milkman, Martin I. (1989). "Teacher Unions and High School Productivity." Ph.D. dissertation, University of Oregon.

Murnane, Richard J. and David K. Cohen. (1986). "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." *Harvard Educational Review* 56(1): 1–17.

Prendergast, Canice. (1999). "The Provision of Incentives in Firms." *Journal of Economic Literature* 37: 7–63.

Tulli, Dennis J. (1991). *An Assessment of Student Achievement before and during a Merit Pay Program for Teachers of the Penn Manor School District*. Ed.D. dissertation, Temple University.