

# TEACHER INCENTIVES AND MERIT PAY

JAY P. GREENE AND GREG FORSTER

CENTER ON  
INNOVATION & IMPROVEMENT

Twin paths to better schools

---

## CENTER ON INNOVATION & IMPROVEMENT

121 N. Kickapoo Street  
Lincoln, IL 62656 USA  
Phone: 217-732-6462  
Fax: 217-732-3696

[www.centerii.org](http://www.centerii.org)

Information • Tools • Training

Positive results for students will come from changes in the knowledge, skill, and behavior of their teachers and parents. State policies and programs must provide the opportunity, support, incentive, and expectation for adults close to the lives of children to make wise decisions.

The Center on Innovation & Improvement helps Regional Comprehensive Centers in their work with states to provide districts, schools, and families with the opportunity, information, and skills to make wise decisions on behalf of students.

A national content center supported by the U. S. Department of Education's Office of Elementary and Secondary Education.

Award #S283B050057

***The opinions expressed herein do not necessarily reflect the position of the supporting agencies, and no official endorsement should be inferred.***

Series Editor: Herbert J. Walberg  
Copy Editors: Lori Thomas, Pam Sheley  
Design: Pam Sheley  
Executive Director: Sam Redding

## Summary

This paper finds that little empirical evidence exists on “merit pay” policies, but the evidence we do have is promising, so the logical next step is to enact more rigorous pilot programs than have yet been tried.

## Abstract

Teacher pay is currently structured in a way that pushes high performers out of the profession and pulls low performers into it. “Merit pay” policies seek to reestablish a link between pay and performance. Few pilot programs have been tried, and those that have been tried have been subject to significant compromises that can be expected to undermine their effectiveness (such as the use of subjective rather than objective measurements of teacher merit) or have not been designed to allow for high-quality scientific evaluation. Nonetheless, the small amount of empirical evidence we do have is promising. This paper recommends the enactment of pilot programs for merit pay that are free from political compromises and that use random assignment methods, to allow for a more fair empirical test of this policy.

## Introduction

The most prominent strategies for education reform share a basic idea: the importance of incentives. These reforms seek to improve education by changing the incentives that are brought to bear on schools, teachers, and students. This paper focuses on teachers, reviewing the empirical research on merit pay policies. Most other reforms, such as school-level accountability programs and school choice, bring incentives to bear on teachers, but only indirectly, because they create incentives for whole schools to improve. Merit pay seeks to create direct teacher incentives by providing greater financial rewards for more effective teachers. This is expected to improve incentives for classroom performance in two related ways: an increased incentive for teachers to become more effective, and an increased incentive for people who make more effective teachers to enter and remain in the teaching profession. (The other major reform that provides direct teacher incentives is exams for student promotion or graduation, which motivate teachers to avoid the stigma of having large numbers of their students fail.)

Reforms that provide direct teacher incentives may be the only way that incentive-based reforms can strongly influence teachers’ effectiveness. Due to the strict workplace rules and teachers’ rights imposed by union contracts, school administrators are highly constrained in their ability to influence teachers’ performance. They can make policy decisions for the school, such as making more use of basic skills tests and shifting curricular emphases and so on, but they have little ability to get teachers to take on

more students, work more hours, adopt different teaching styles, or otherwise change teacher behavior through policy. As a result, reforms that bring incentives to bear upon schools may have a stronger effect on school policies and administrators than on teachers. These incentives may improve schools, but if they do so, it is likely that they do so by bringing about better school policies and administration rather than better teaching as such. A recent analysis of federal survey data by Christiana Stoddard of Montana State University and Peter Kuhn of the University of California at Santa Barbara found that education reform policies such as school accountability and competition from charter schools were associated with higher test scores (confirming the findings of other studies that these reforms improve academic outcomes), but that they had no effect on the number of hours teachers worked per day. This suggests that teacher behavior is not being affected.<sup>1</sup> Seeking an explanation, they write: “We conjecture that the lack of an effect of education reforms on teacher hours may be due to the weak connection between individual teacher effort and rewards that characterize most of the reforms that have been implemented.”<sup>2</sup>

Unfortunately, the majority of the material described as “research” on any given education policy does not follow the empirical methods necessary for sound social science. Instead, most education policy “research” consists of theoretical discussion, large generalizations drawn from observations of a very small number of cases, the misuse of statistics that are merely descriptive to make causal claims, and so on. The research on merit pay is no exception to this rule. It is essential that we rely only on sound science when evaluating the evidence on these and other policies. This may leave us with fewer studies to look at, but it is the only way we can have confidence that we are getting a clear picture of what the evidence actually indicates.

We were able to identify only two sound empirical studies on merit pay, so any conclusions must be considered preliminary. That having been said, to the extent that good research is available on this reform, it provides some grounds for moderate optimism about its prospects for success. However, it also suggests serious concerns about the possibility that merit pay programs can be undermined by political compromises. The presence or absence of a program called “merit pay” may be less important than the actual rigor of the program. Compromises in the design of a program that are accepted as the political cost of getting the program enacted – such as using peer evaluation rather than test scores to determine which teachers receive merit pay – may dilute or even remove the incentives that these reforms are intended to create.

## Key Principles

### Confront the Problem: Teacher Pay and Incentives

Few people appreciate how highly abnormal the pay structure for teaching is, as compared to other professions requiring a college degree. Public school teachers are paid much more like factory workers than like professionals. In virtually all other professional occupations, salaries are determined by separate negotiations between each employer and employee. Each worker makes as much as he or she can convince an employer to pay. Teachers, by contrast, are compensated according to uniform pay scales negotiated by labor unions. These scales are based almost exclusively on two factors: the possession of credentials, such as a teaching certificate or a graduate degree in education, and years of experience.

This unique pay system for teachers drastically alters the incentives faced by teachers, compared to the incentives faced by the rest of the professional workforce. The pay system used by the non-teaching professions – individual salary negotiation – efficiently rewards more effective workers, because employers are willing to pay more for better work. This efficiency creates a powerful incentive for professionals to be as effective in their jobs as possible, since they have a reasonable expectation that improved effectiveness on the job will be rewarded by improved compensation. It also creates a powerful incentive for professionals to sort themselves into the professions where they are most effective. If a particular person performs somewhat well as a mechanical engineer but performs even better as a civil engineer, that person will be better paid in the latter profession than in the former, providing a strong incentive for him or her to enter that profession, and also to remain in it longer (as opposed to switching jobs or retiring).

The teacher pay system undermines this incentive structure. Teachers cannot reasonably expect that improved performance will be rewarded with improved compensation, so they have no incentive to exert extra effort to improve their performance. Similarly, there is no particular incentive for people who make good teachers to enter and remain in the teaching profession, as there is in all other professions. Obviously people who enjoy teaching have an incentive to enter and remain in the profession, but this is not the same thing. In other professions, the powerful incentive of work enjoyment for people who like a given occupation is coupled with a similarly powerful monetary incentive for people who are good at that occupation. The lack of efficient incentives in the teaching profession is made much more costly by the existence of an efficient incentive structure in other professions. A college graduate with a high academic aptitude will make an effective worker in a wide variety of occupations. This aptitude will be well rewarded in every profession other than teaching;

highly capable college graduates can expect to make an above-average salary in other professions, but in teaching they know they will only make the same salary as everyone else. By contrast, a low-aptitude college graduate can expect to make a below-average salary in other professions, but is guaranteed to make the same salary as everyone else in teaching. Thus, the absence of efficient incentives in the teaching profession creates a powerful incentive for the lowest-performing college graduates to enter and remain in the profession.

This is the most important explanation for the long-term decline in the academic aptitude of college graduates entering the teaching profession, a well-established phenomenon that has been the subject of much debate. Some assert that lower-performing graduates go into teaching because teaching is poorly compensated relative to other professions, but this is difficult to square with the facts. Federal data show that teacher compensation is comparable to compensation in similar professions.<sup>3</sup> Others assert that improved opportunities for women in non-teaching professions have drawn high-performing female workers out of the teaching “ghetto” to which they were previously confined. While this is more plausible on its face, sound scientific evidence supporting it has been lacking, and some has been brought against it. Caroline Hoxby of Harvard University and Andrew Leigh of the National Bureau of Economic Research have shown that, according to federal labor market survey data, changes in teacher aptitude over time have been much more strongly related to the introduction of unionization in the teacher workforce (with its factory-like pay scales) than to changes in male/female pay ratios in the non-teaching workforce.<sup>4</sup>

One might think that the teacher pay system rewards better performance because it provides increased salaries to teachers who have educational credentials and years of experience. However, this would only be the case if credentials and experience were indicators of better teacher performance. This appears not to be the case. While it may seem counter-intuitive, a surprisingly strong body of empirical evidence shows that there is no relationship between teachers’ effectiveness in the classroom and the possession of traditional educational credentials, and makes it difficult to confirm the existence of any relationship with teaching experience after the first few years. Eric Hanushek of Stanford University reviewed every available empirical study of teachers’ credentials and performance, 171 studies in all, and found only nine that identified a significant positive relationship; he also found five studies identifying a negative relationship. When he looked only at the very best-quality studies, he identified 33 studies, none of which found a significant relationship.<sup>5</sup> Several more recent studies confirm this finding.<sup>6</sup> Studies of teacher performance and years of experience are more subtle to interpret. Hanushek identified 207 studies, of which only 29 identified a positive relationship. However, among top-quality studies, 14 of 36 found a positive relationship.<sup>7</sup> This can be best described as inconclusive.

More recent research suggests a reason why a firm conclusion has not been reached: the relationship appears to be nonlinear across a teacher's career. Specifically, several studies have found that teacher performance rises somewhat in the first two or three years, but does not rise thereafter.<sup>8</sup> In light of this evidence of a nonlinear effect that is confined to the first few years, it is difficult to take Hanushek's review as providing evidence of a robust relationship between teacher experience and performance.

We are left with the conclusion that teacher pay is, on the whole, unrelated to teacher performance. Teachers are compensated based on their ability to acquire educational credentials and accumulate years of experience, neither of which seems to have a reliable relationship to the ability to teach well. And where pay is unrelated to performance, workers have no incentive to improve their performance, high performers have a strong incentive to stay away from the profession or leave it, and weak performers have a strong incentive to enter it and remain.

---

### Give Merit Pay a Fair Test

Given the importance of teacher quality and the unusual challenges posed by the teacher pay system, it isn't surprising that merit pay has emerged as a reform strategy. At its heart, merit pay is an attempt to restore the teaching profession to something like the standard compensation system for professional occupations. Obviously merit pay is not the same as individual negotiation between employer and employee – the unionization of teachers prevents the restoration of this system as such. Instead, merit pay attempts to approximate the benefits of the standard professional pay system by establishing at least some relationship between teacher pay and performance.

Unfortunately, merit pay has not yet been allowed a fair test. To begin with, some programs called "merit pay" are based on the aggregate performance of schools or districts rather than the individual performance of teachers. If a school or district performs well, its teachers get bonus pay. We do not consider these programs to be true "merit pay" programs, since they provide school-level or district-level incentives rather than individual teacher incentives. These programs provide only a very weak connection between performance and reward for each individual teacher. For this reason, we have not included such programs in this review.

A second problem in giving merit pay a fair test is to set up a test that can be evaluated using high-quality scientific methods. To evaluate merit pay policies, we need to find out whether students who are taught by teachers receiving merit pay have higher academic achievement. One challenge when studying merit pay is drawing appropriate comparisons. For the results of a study to be valid, we must find comparable sets of students who are and are not assigned to teachers receiving merit pay. Unfortunately, the assignment of students to teachers receiving merit

pay can be determined by factors that also affect student outcomes. This will interfere with our ability to draw valid comparisons between the academic achievement of students who are and are not being taught by merit-pay teachers.

We are aware of only one empirical study that overcomes this problem in a fully satisfactory way. It is well known in the education research world that during the 1980s Tennessee conducted a random-assignment experiment in smaller class sizes, known as the STAR program. What is less well known is that Tennessee also had a merit pay program at the time, and that some of the teachers in the STAR experiment were part of this program, while others were not. Since students in the STAR program were assigned to their teachers randomly, the program allows for a valid comparison among students who were and were not assigned to merit-pay teachers. (It is true that other studies of STAR have produced evidence that the program's random assignment method was compromised in practice by non-random reassignment of students to different classes.<sup>9</sup> However, while this has cast significant doubt on the program's positive findings for smaller class sizes, the authors of the merit-pay study we examine here ran several statistical analyses confirming that it is very unlikely their finding is affected by this problem.)

Thomas Dee of Swarthmore College and Benjamin Keys of the University of Michigan examined the relationship between merit pay assignments and test scores of students in the STAR program. They found that students assigned to merit pay teachers had math scores that were 3 percentile points higher than students assigned to other teachers. They did not find any statistically significant difference in reading scores. They found inconsistent results on whether higher-rated teachers (that is, teachers in the higher pay ranks based on their official performance evaluations) produced better results than lower-rated teachers. Teachers in the upper ranks of the program produced better reading scores than teachers who were not in the program at all, while teachers in the lower ranks of the program did not. On the other hand, teachers in the higher ranks did not produce better math scores than teachers who were not in the program at all, while teachers in the lower ranks of the program did.<sup>10</sup>

On the one hand, it appears that the presence of a merit pay program produced better teacher performance. Teachers who were promised a reward for better performance appear to have performed better. Three percentage points is a moderate effect, but it should not be underestimated – consider the impact of this effect when compounded over twelve years in the school system. The absence of a statistically significant finding for reading scores does not cast doubt on the positive finding for math scores; it is not uncommon for studies of education policies to find significant effects in math but not in reading. This is almost certainly because schools have a greater impact on math skills, as opposed to the greater influence

of parents on reading skills. It is much more common for parents to read with their children than it is for them to teach their children math.

On the other hand, it is not clear that the program reliably identified better teachers and promoted them to the higher performance ranks. Advancement within the program was not based on test scores, but on subjective judgments based on classroom visits. In other words, merit pay was not being provided based on an objective standard of merit. In addition to the introduction of subjective judgment, the evaluators were particularly likely to be biased in their judgment. Teachers in the lower ranks were judged by their principals, with whom they work on a daily basis. A principal's evaluation of each teacher's work cannot help but be influenced by his or her working relationship with that teacher. This poses an especially strong problem, because it is evaluations of teachers in the lower ranks that determine which teachers advance to the upper ranks. Teachers in the upper ranks were judged by peer evaluators (teachers from other school districts), which is a little better. Even here, however, there is a strong potential for bias. Members of the teaching profession, while they may know the most about teaching on a practical level, are also, for the same reason, the least objective observers of it. Like all practitioners of any kind, they are subject to biases arising from the particular perspective of the practitioner.

This distortion of the evaluation system represents a third major obstacle to giving merit pay a fair test. Such distortion cannot help but undermine the incentives that merit pay is supposed to create. The system depends on a connection between compensation and more effective teaching. If higher pay is connected to teaching that only looks better during classroom visits rather than teaching that actually is better, or (even worse) to having a good relationship with the principal, then we should not expect merit pay policies to succeed in the long term.

One possible interpretation of this study's results is that teachers in the merit pay program were more effective because the program attracted higher-performing teachers. The merit pay system was voluntary for most teachers from its inception, and it was made fully voluntary after its first two years. Thus, teachers in the program may produce higher test scores because better teachers chose to enter the program. This would help explain why the merit pay program produced better test scores even though it is not clear that it succeeded in promoting better teachers to the higher pay ranks. It would also be consistent with strong theoretical expectations – it makes sense that more effective teachers would want to be paid based on performance, while less effective teachers would not. This interpretation of the evidence, if true, would not undermine the case for merit pay, but rather confirm it. If merit pay attracts better teachers, implementing merit pay throughout the school system would attract better teachers into the system while motivating worse teachers to leave.

Another empirical study is worth considering, although it has a more limited method for overcoming the problem of bias resulting from the differential assignment of students to teachers receiving merit pay. This study, conducted by John Schacter of The Teaching Doctors and Yeow Meng Thum of the University of California at Los Angeles, examines the Teacher Advancement Program (TAP), a privately funded pilot merit pay program. Its method is to match participating schools with other schools that are similar in their size, locale, initial student achievement, percentage of minority students, percentage of students eligible for free lunch, and percentage of limited English proficient students. This matching provides a control group that is somewhat comparable, but not perfectly comparable. In particular, the TAP program is voluntary for schools – the school principal must take the initiative to sign up. It is very likely that students in schools whose principals volunteer for TAP – that is, students in schools with highly motivated and reform-minded principals – will be dissimilar in important respects from students whose principals did not volunteer for TAP. This must be borne in mind when interpreting the study's results.

The study found that test scores in TAP schools rose 10% to 21% more than in the somewhat-comparable control group. When math and reading scores were disaggregated, it found that math scores rose by 14% more in TAP schools, while reading scores rose by 4% to 6% more.<sup>11</sup> These would be very strong findings if there were no problems with the comparison group. As it is, the best we can say is that the findings are large enough that it is plausible – to say no more – that they encompass a positive effect from merit pay as well as a positive effect from the selection bias discussed above.

## Conclusion

More research is obviously needed in order to provide a large enough body of empirical evidence to form a basis for solid conclusions about the performance of merit pay programs. This poses a significant challenge to the research community, given the difficulty of identifying appropriate comparison groups. The evidence that is available, however, provides some grounds for moderate optimism about merit pay. But it also suggests that when teachers are evaluated based on subjective judgment rather than on objective test scores, this invites systematic problems of bias that we could expect to undermine the incentives that merit pay is supposed to provide. To our knowledge, a merit pay program based on objective measurement of students' academic outcomes has never been attempted. Such a program would provide a much fairer test of merit pay than has been permitted so far. If it were carried out with random assignment methods, as in the merit pay program that coincided with the Tennessee STAR experiment, it would allow for the first ever truly fair test of this promising policy alternative.

## References

- Betts, J. R., Rueben, K. S., & Danenberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California*. San Francisco: Public Policy Institute of California.
- Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). *The effects of Teach for America on students: Findings from a national evaluation*. Mathematica Policy Research.
- Dee, T. S., & Keys, B. J. (2004). Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management*, 23(3), 471-488.
- Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, 2(1), 50-55.
- Greene, J. P., Forster, G., & Winters, M. A. (2005). *Education myths*. New York: Rowman & Littlefield.
- Grissmer, D. W., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us*. Santa Monica, CA: RAND Corporation.
- Hanushek, E. A. (1996). School resources and student performance. In G. Burtless (Ed.), *Does money matter?* Washington DC: Brookings Institution.
- Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR Experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis*, 21(2), 143-163.
- Hoxby, C. M., & Leigh, A. (2005). Wage distortion. *Education Next*, 5(2), 50-56.
- Podgursky, M. (2003). Fringe benefits. *Education Next*, 3(3), 71-76.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2000). *Teachers, schools and academic achievement*. Unpublished Paper.
- Schacter, J., & Thum, Y. M. (2005). The Teacher Advancement Program: 2004 evaluation. Unpublished Paper.
- Stoddard, C., & Kuhn, P. (2006). *Incentives and effort in the public sector: Have U.S. education reforms increased teachers' work hours?* National Bureau Of Economic Research. (Working Paper 11970).
- Teacher Certification Reconsidered: Stumbling for Quality. (2001). Abell Foundation.
- Vedder, R. (2003). Comparable worth. *Education Next*, 3(3), 14-19.
- Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Educational Testing Service.

## Endnotes

<sup>1</sup>If teachers were in fact responding to incentives from reforms, some might respond by working more efficiently in the same number of hours, but we would expect at least some to respond by working more hours. This would lead to a rise in the average hours worked by teachers. Thus, if average teacher work hours remain constant, this provides evidence that reforms are not affecting teacher behavior.

<sup>2</sup>Stoddard and Kuhn, 2006. One of the reforms included in this study is programs labeled “merit pay.” However, the authors note that most of the “merit pay” programs they examine are district-level or school-level programs, rather than merit pay for individual teachers. For reasons explained below, we do not consider these programs to be truly “merit pay” in the relevant sense, so we do not include Stoddard and Kuhn’s findings on “merit pay” in our analysis.

<sup>3</sup>See Vedder, 2003; Podgursky, 2003; and Greene, Forster, and Winters, 2005, pp. 71-84.

<sup>4</sup>Hoxby and Leigh, 2005.

<sup>5</sup>Hanushek, 1996.

<sup>6</sup>See Betts, Rueben, and Danenberg, 2000, esp. p. 195; Rivkin, Hanushek, and Kain, 2000, p. 32; Wenglinsky, 2000; Grissmer, Flanagan, Kawata, and Williamson, 2000; Teacher Certification Reconsidered, 2001; Goldhaber, 2002; Decker, Mayer, and Glazerman, 2004; and Greene, Forster, and Winters, 2005, pp. 59-70.

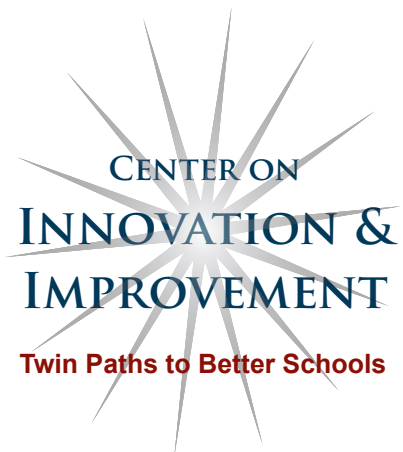
<sup>7</sup>Hanushek, 1996.

<sup>8</sup>See Rivkin, Hanushek, and Kain, 2000; Grissmer, Flanagan, Kawata, and Williamson, 2000; and Greene, Forster, and Winters, 2005, pp. 59-70.

<sup>9</sup>See Hanushek, 1999.

<sup>10</sup>Dee and Keys, 2004.

<sup>11</sup>Schacter and Thum, 2005.



**Center on Innovation & Improvement**  
121 N. Kickapoo Street  
Lincoln, IL 62656 USA  
Phone: 217-732-6462  
Fax: 217-732-3696

[www.centerii.org](http://www.centerii.org)