

## Policy Retrospectives

*Douglas J. Besharov*  
*Editor*

Submissions to Policy Retrospectives should be sent to Douglas J. Besharov, School of Public Policy, University of Maryland/American Enterprise Institute, College Park, Maryland 20742; (301) 405-6341.

### TEACHER PERFORMANCE PAY: A REVIEW

Michael J. Podgursky, Matthew G. Springer

### INTRODUCTION

Salary schedules for teachers are a nearly universal feature of American K–12 public school districts. Data from national surveys show that close to 100 percent of traditional public school teachers are employed in school districts that make use of salary schedules in pay setting (Podgursky, 2007). Thus, roughly 3.1 million public school teachers from kindergarten through secondary level are paid largely on the basis of years of experience and education level—two variables weakly correlated, at best, with student outcomes (Hanushek, 2003).

The single salary schedule tradition contrasts with pay determination practices in the majority of professions, where performance-related pay programs are commonplace. In a survey of 1,681 firms, Hein (1996) found that 61 percent employed variable, performance-related compensation systems. A leading compensation textbook reports that over three-fourths of exempt (non-hourly) employees in large firms are covered by merit pay systems (Milkovich & Newman, 2005). Pay determination practices also vary between K–12 sectors. Examining early vintages of the Schools and Staffing Survey, Ballou and Podgursky (1997) and Ballou (2001) found that private school teachers were much more likely than their traditional public school counterparts to be rewarded for teaching performance, despite the fact that the majority of private schools reported relying on a salary schedule for teacher pay.

Pay determination practices in most professional fields are usually market driven, enabling organizations to match the offer of competitor firms for employees they

wish to retain or to create an attractive compensation package for professionals they wish to recruit. Even the federal General Schedule (GS) pay system is more flexible and market-based than those found in most traditional public schools. Civil servants advance through the GS not only in 15 grades, but also along 10 pay steps based on merit and experience (Ballou & Podgursky, 1997). Furthermore, the Department of Defense and Department of Homeland Security within the federal government recently began implementing additional performance-related pay programs to improve organizational performance.

NCLB-induced state accountability systems, coupled with the poor relative performance of U.S. students on international math and science tests, have stimulated interest in the design and implementation of performance-related pay policy. Many districts, and even entire states, are exploring performance-related pay to improve administrator and teacher productivity and recruit more qualified candidates. These performance-related pay plans come in many different forms, from compensation based on supervisor evaluations and portfolios created by teachers to payments awarded on the basis of student growth at the teacher, group (for example, subject or grade), and/or campus levels. By some journalistic (perhaps exaggerated) estimates, at least one-third of the nation's K–12 public school districts appear “poised” to participate in performance incentive policies initiated at the local, state, or federal level. Whether truly “poised” or not, it is clear that many states and districts are actively considering the option. Nor is this interest restricted to the United States. A number of European and developing nations have begun to experiment with performance pay (Sclafani & Tucker, 2006).

The purpose of this paper is to examine the economic case for performance-related pay in K–12 education system. While we focus on teachers, by far the largest group of employed professionals in K–12 public education, most of the arguments generalize to school administrators as well. Our review begins with a brief history of U.S. teacher compensation policy and then moves to general descriptions of six large-scale performance-related pay programs currently in operation or about to be launched in U.S. schools. We then review theoretical arguments involving performance-related pay policy, paying particular attention to issues such as performance monitoring, team production, the multitasking problem, and input-based versus output-based pay systems. We then review several strands of empirical research that have relevance for this debate, including teacher effect studies, direct evaluations of individual and group performance pay schemes, and studies of incentive pay in private schools and charter schools. While the direct evaluation literature is slender, it does provide some important results for policy. We conclude that while the empirical literature is not sufficiently robust to prescribe how systems should be designed—for example, optimal size of bonuses, mix of individual versus group incentives—it does make a persuasive case for further experiments by districts and states, combined with rigorous, independent evaluations.

## **A BRIEF HISTORY OF U.S. TEACHER PAY POLICY**

### **Room and Board Compensation Model**

The emerging transportation system of America in the early 19th century—by river and canal, and eventually rail—enabled communities situated in rural, agrarian-based locations to trade and prosper. Nearly 80 percent of all citizens living in rural areas

and half of all working citizens were farmers (Protsik, 1995). Out of this context emerged the one-room schoolhouse education systems of the late 18th and early 19th centuries, whose design was influenced by regional variation in the crop production schedules and the dependence of farm production on child labor.

In this environment, the *room and board* compensation model developed. In addition to a small stipend, teachers received room and board by rotating their residences weekly in different students' homes (Protsik, 1995). This facilitated not only attraction and retention of teachers in geographically isolated locations, but it also clearly solved several principal-agent problems, with each family in a community monitoring a teacher's ability to instill book-learning as well as to foster the appropriate "moral character" in their children (Tyack, 1975).

However, the one-room-schoolhouse education system lacked the capacity to deliver the level or variation in human capital demanded by an industrializing and urbanizing economy. Dramatic increases in the number of students seeking schooling caused a simultaneous increase in the demand for teachers per school. The combined effect of these trends spurred the move toward a grade-based system of education and dramatically altered the nature of teacher compensation.

### Grade-Based Compensation Model

With industrialization in the late 19th and early 20th century, the "new" economy involved a greater "use of science by industry, a proliferation of academic disciplines, a series of critical inventions and their diffusion" (Goldin, 2003). Given the intensified demand for greater skill from a better educated labor force, teacher compensation policy too was reconceptualized.

The *grade-based* compensation model was created in the late 1800s. Similar to the factory production model preoccupying most sectors of the American economy, the grade-based compensation model paid teachers for the level of skill needed to educate a child at their specified point of educational attainment. Because it was believed that elementary-age students were easier to educate, and less formal training was required to teach at that level, teachers who instructed children in their early years earned less than secondary level teachers (Guthrie, Springer, Rolle, & Houck, 2007).

While the design of the new grade-based system made pay uniform by grade level within the profession, the system fostered gender and racial inequities. Entry requirements to teach at the secondary level were more accessible to white males. Furthermore, subjective administrator evaluations of teacher merit were integrated into many grade-based compensation models, resulting in gender- and race-based inequities as well as nepotism.

### Position-Automatic or Single Salary Schedule

Around the turn of the 20th century, labor leaders like Samuel Gompers pushed management and factory owners for better working conditions and salaries for their employees. Strikes, boycotts, and negotiations carried out by the American Federation of Labor (1886), the Industrial Workers of the World (1905), and the Congress of Industrial Organizations (1938) were influential in promoting egalitarian pay policy (Kerchner, Koppich, & Weeres, 2003). While not a direct byproduct of collective bargaining, the *single salary schedule*, originally called the "position-automatic schedule," emerged in this tumultuous period of industrial relations. The single salary schedule is a system of uniform pay steps that ensures teachers with the same years of experience and education level receive the same salary

(Moelhman, 1927). In a typical schedule, rows indicate years of experience and columns indicate the levels of graduate coursework completed or degrees obtained. This system was implemented to create pay equity, professionalism, and employee satisfaction across grade levels, political wards, districts, and disciplines and to displace prior pay systems negotiated between individual teachers and local school boards (Kershaw & McKean, 1962).

Since its inception, the single salary schedule has been a nearly constant feature of the public school compensation scheme. By 1950, for example, 97 percent of all schools had adopted the single salary schedule (Sharpes, 1987). This figure is remarkably similar to contemporary estimates that 96 percent of public school districts use a uniform salary schedule to compensate teachers (Podgursky, 2007). While the single salary schedule has proved to be remarkably persistent, there have been attempts at change.

## 20th Century Compensation Experiments in Education

Since first implemented in 1921 in Denver, Colorado, and Des Moines, Iowa, the single salary schedule has attracted criticism. Most prominent among these critiques is that the schedule standardizes remuneration, depriving public school managers of authority to adjust an individual teacher's pay to reflect both performance and labor market realities. Numerous teacher compensation reform models have been proposed as alternatives, many under the banner of performance-related pay. The two most prominent types of reform programs have been (1) merit-based pay and (2) knowledge- and skill-based pay.

### *Merit-Based Pay*

Although merit-based pay programs date back to Great Britain in the early 1700s, and somewhat similar ideas formed around the notion of performance contracting in the late 1960s (Stucker & Hall, 1971), it was not until the release of the *A Nation at Risk* report in 1983 that a significant number of public school districts in the United States began considering merit-based pay as an alternative or supplement to the single salary schedule. Merit-based pay rewards individual teachers, groups of teachers, or schools on any number of factors, including student performance, classroom observations, and teacher portfolios. Merit-based pay is a reward system that hinges on student outcomes attributed to a particular teacher or group of teachers rather than on "inputs" such as skills or knowledge—a critical distinction that is emphasized later in this review. A report released by the Progressive Policy Institute in 2002 classified school-based performance awards as the most common type of merit-based pay programs operational in U.S. K–12 public schools, but noted as well that rewards can be distributed at, or targeted to, specific grade levels (grade-level teacher teams), departmental units, or combinations thereof (Hassel, 2002).

### *Knowledge- and Skill-Based Pay*

Since the 1990s, knowledge- and skill-based pay has garnered significant attention as an alternative strategy for compensating teachers (Odden & Kelley, 1996). This approach, which has some analogues in the private sector (Beer & Cannon, 2004; Heneman & Ledford, 1998), represents a policy compromise between proponents and opponents of performance-related compensation in education. Knowledge- and skill-based pay programs, such as those designed by the Consortium for Policy

Research in Education (CPRE) at the University of Wisconsin, reward teachers for acquisition of new skills and knowledge presumably related to better instruction. Salary increases are tied to external evaluators and assessments (for example, the Praxis III and National Board for Professional Teaching Standards) that gauge the degree to which an individual teacher has reached specified levels of “competency” (Odden & Kelly, 1996). Although proponents argue that these strategically focused rewards can broaden and deepen teachers’ content knowledge of core teaching areas and facilitate attainment of classroom management and curriculum development skills (Odden & Kelley, 1996), evidence to date suggests that the knowledge and skills being rewarded in these “input-based” pay systems may have a negligible impact on student outcomes (Ballou & Podgursky, 2001; Hanushek & Rivkin, 2004).

### Private Sector Compensation Practice

Before turning to current reforms, it is useful to briefly consider how some of these issues are framed in the private sector compensation literature and some sectors of government outside of K–12 public education. In compensation textbooks, a distinction is often made between *base pay* and *variable pay*. Base pay represents a foundation or floor on pay, which is guaranteed (or at low risk) and typically paid as a salary, hourly, or piece rate wage. A second component, variable pay, is riskier and is associated with the bonus or performance pay system. Variable pay is compensation that is contingent on discretion, performance, or results that comes in the form of an individual bonus, a group bonus, or some combination of the two. For the vast majority of teachers in the current public school system, base pay is set by the experience and education cells in the traditional single salary schedule, perhaps with additional pay for selected additional duties (for example, coaching, directing band). There is no variable pay component of the single salary schedule.

Some proponents of knowledge- and skill-based pay, and some of the reforms considered below, aim at reforming base pay for teachers. Although this would move base pay away from the single salary schedule, it still means pay is tied to certain teacher “inputs” such as credentials and that the variable, or risky, component would be negligible. For example, rather than rewarding teachers for an M.A., teachers might be rewarded for National Board certification, mentoring or coaching teachers early in their career, or assuming additional curricular, instructional, and school improvement responsibilities.

These types of base pay reforms are contrasted with pay plans that leave base pay alone but add a variable (and hence risky) component into teacher pay. In these instances, the risky component is typically tied to one or more “outputs” of interest. For example, rather than remunerating for certain teacher characteristics, a variable pay plan ties some part of teacher pay to individual, group (that is, grade level or subject teams of teachers), or school performance. If predetermined performance targets are met, then these teachers earn more; if they are not, they earn base pay. Most experiments or pilot programs, the evaluations of which we consider in a subsequent section, focus on variable pay schemes. With this distinction in mind, we now turn to a consideration of some current reform schemes.

### CURRENT PERFORMANCE-RELATED PAY PROGRAMS

There is growing national interest in performance-related pay in K–12 public education. While we are aware of no systematic compilation of these programs, groups

like the federally funded National Center on Performance Incentives (NCPI) at Vanderbilt University's Peabody College, Education Commission of the States (ECS), Mathematica Policy Research, Inc., and the Center for Educator Compensation Reform have begun tracking teacher and administrator compensation reforms, issues, and future research opportunities.<sup>1</sup> By all accounts, interest in performance-related pay programs is growing, as is the number of programs under development and being implemented. In this section, we consider briefly some current U.S. programs starting with district- and state-level programs and then national- and federal-level initiatives.

### **Denver Public Schools' Professional Compensation System for Teachers (ProComp)**

In 1999, the Denver Classroom Teachers Association and the Denver Public Schools reached agreement on an alternative teacher pay plan that linked pay to student achievement and professional evaluations. Following refinement of the pilot model by teachers, principals, administrators, and community members, the Professional Compensation Systems for Teachers (ProComp) was adopted in spring 2004 by the Board of Education and members of the Denver Classroom Teachers Association (Community Training and Assistance Center, 2004).

The ProComp approach is clearly weighted toward a knowledge- and skill-based pay model with variable pay supplements for student growth and market incentives. There are four components that enable teachers to build earnings through ten elements, or learning opportunities, including (1) knowledge and skills; (2) professional evaluation; (3) market incentives; and (4) student growth. As noted in Table 1, knowledge- and skill-based pay programs in the form of National Board for Professional Teaching Standard certification holds the greatest potential for pecuniary returns. Student achievement growth, which includes both teacher and school-wide growth awards, can generate a nice boost in pay (a maximum award of approximately \$2,000), while excellence in professional evaluations provides a salary increase of about \$1,000 for non-probationary teachers.

ProComp's position in Denver Public Schools' operational structure was recently strengthened. First, Denver voters approved a November 2005 ballot initiative to pay an additional \$25 million in taxes to fund a scale-up of ProComp. Furthermore, Denver Public Schools received a \$22.67 million, five-year Teacher Incentive Fund (TIF) award from the United States Department of Education (USDoe). TIF award funds will be used to expand ProComp to nearly 90 percent of Denver's 150 K–12 public schools. Now completing the first of nine voter approved years, ProComp has evolved from a four-year pilot program in 16 schools into one of the nation's most widely known performance-related pay programs.

### **Texas' Governor's Educator Excellence Award Programs**

In 2006, Governor Rick Perry and the 79th Texas Legislature crafted the Governor's Educator Excellence Award Programs (GEEAP), creating the single largest performance-related pay program in U.S. public education. GEEAP consists of three programs: (1) the Governor's Educator Excellence Grant (GEEG), (2) the Texas Educator Excellence Grants (TEEG), and (3) a district-level grant yet to be named. By 2008, GEEAP will provide approximately \$330 million per annum to high-performing, high-poverty public schools in Texas.

<sup>1</sup> See, for example, Azordegan, Byrnett, Campbell, Greenman, and Coulter (2005), NCPI Web site at [www.performanceincentives.org](http://www.performanceincentives.org), and Glazerman et al. (2006).

**Table 1.** Selected major performance-based pay programs.

Name of Plan	Target	Size of Bonus	Size of Program	Year of Inception	Funding Source
Denver's Professional Compensation System for Teachers (ProComp)	Teacher Award Program	<i>Knowledge and Skills:</i> \$1,000 tuition credit for professional development coursework; 2% salary index bonus for completing courses and demonstrating skills (\$659); 9% salary index bonus for NBPTS certification (\$2,967)  <i>Professional Evaluation:</i> Salary increase of 3% index for satisfactory evaluation of nonprobationary teacher (\$989)  <i>Student Growth:</i> 3% sustainable increase for CSAP goal completion (\$989); bonus of 2% index for "distinguished" schools (\$659); bonus of 1% for meeting one of two goals (\$330)  <i>Market Incentives:</i> 3% index bonus for (\$989) or (\$989) assignments  <i>Total Bonus Range:</i> \$330–\$7,582	Pilot program operated in 16 schools	Pilot program operated from 1999 through 2004	Scaled-up program locally-funded following a \$1 million levy approved by taxpayers
Florida's Merit Award Program (MAP)	Teacher and Administrator Award Program	At least 5%, no more than 10% of the average teacher salary for the district	\$147.5 million	Replaced Florida's Special Teachers Are Rewarded (STAR) program in 2005	State funded by the Florida Education Finance

(Continued)

**Table 1.** (Continued)

Name of Plan	Target	Size of Bonus	Size of Program	Year of Inception	Funding Source
<p>At least 60% of award must be based on student performance</p> <p>Up to 40% of funds may be used to award professional practices, as measured by principal assessment</p>	<p>Includes three major initiatives</p>	<p>School-based awards range from \$40,000 to \$290,000 per year based on student enrollment</p> <p>Texas Education Agency recommends Individual teacher awards range from \$3,000 to \$10,000</p>	<p>The three programs will include approximately 1300 schools and \$330 million per annum</p>	<p>in March 2007</p>	<p>Program (FEFP)</p>
<p>School Award Program</p>	<p>Schools must be in top third of schools in percentages of economically disadvantages students and have performance rating of either Exemplary or Recognized, or must in the top quartile of TEA's Comparable Improvement measure.</p>	<p>School-based awards range from \$60,000 to \$220,000 per year based on student enrollment.</p>	<p>Approximately 100 schools are eligible</p>	<p>Pilot program was implemented in 2006</p>	<p>Pilot program is funded through federal appropriations</p>

	25% to all school personnel, including principals, and/or professional development activities (Part II funds)	School-based awards range from \$40,000 to \$290,000 per year based on student enrollment	1,163 school are eligible during the 2006–2007 school year	Program was implemented during the 2006–2007 school year	Program is funded through state appropriations
<i>Texas Educator Excellence Grant</i>	Schools must be in top half of schools in percentages of economically disadvantaged students.	75% of award must be paid to full-time classroom teachers based on a variety of objective measures of student performance (Part I)	\$100 million annually through 2009		
	District Award Program	25% to all school personnel, including principals, and/or professional development activities (Part II funds)			
	District Award Program	District-based award that is contingent upon district and school size			
<i>District-Level Grants Program (To Be Named)</i>	All school districts are eligible	60% of funds to directly award classroom teachers 40% of funds go to other personnel stipends and/or TAP implementation	\$230 million annually through 2010	Program will be implemented in 2008 school year	State funded
Minnesota's O-Comp	Schools receive funds to award teachers for excellence in student achievement	Districts receive \$260 per student to implement program	\$86 million Currently in 22 districts with 134 additional districts expected by 2008 school year		State funded

(Continued)

**Table 1.** (Continued)

Name of Plan	Target	Size of Bonus	Size of Program	Year of Inception	Funding Source
Milken Family Foundation's Teacher Advancement Program (TAP)	Individual Teachers	<p><i>Master Teachers:</i> \$5,000 to \$11,000</p> <p><i>Mentor Teachers:</i> \$2,000 to \$5,000</p> <p>There is a range in the size of performance bonuses—TAP recommends school average bonus of \$2,500 per teacher</p>	<p>9 states totaling approximately 50 school districts</p> <p>10 additional states are actively pursuing implementation</p>	1999	Private Family Philanthropic Foundation

### ***Governor's Educator Excellence Grants (GEEG)***

This program is funded at \$10 million annually through the 2008 school year. Funds are distributed in the form of noncompetitive grants to approximately 100 schools that are in the top third of Texas schools in terms of percentage of economically disadvantaged students and either: (1) carry a performance rating of Exemplary or Recognized or (2) in the top quartile on TEA's Comparable Improvement measure.<sup>2</sup> Individual campus-award amounts vary according to student enrollment, ranging from \$60,000 to \$220,000 per year.

GEEG schools are required to use 75 percent of these funds, called Part I funds, for direct incentives to full-time classroom teachers. These incentives may be based both on improvement in student achievement and on teacher effectiveness in collaborating with colleagues to improve student achievement on the campus. Part II funds, representing 25 percent of the total award, may be spent on (1) direct incentives to other school employees (including principals) who contribute to improved student achievement, (2) professional development, (3) teacher mentoring and induction programs, (4) stipends for participation in after-school programs, (5) signing bonuses for teachers in hard-to-staff subjects, and/or (6) programs to recruit and retain effective teachers.

### ***Texas Educator Excellence Grants (TEEG)***

This program is state funded at \$100 million per year. Eligibility criteria and requirements are nearly identical to those of the GEEG program. However, schools must be in the top half of Texas schools in terms of percentage of economically disadvantaged students. Grant amounts range from \$40,000 to \$295,000 per year. For the 2006–07 school year, 1,163 campuses are eligible for grants. The TEEG program also separates funds into Part I and Part II funds, with the former based on objective measures of student performance and the latter on a variety of incentives and professional growth activities.

### ***District-Level Grant***

This program will be funded at approximately \$230 million annually with state funds provided through the Texas Educator Excellence Fund. All districts in the state will be eligible for funding. Districts may apply for funds for all campuses or for selected campuses. Districts are required to use at least 60 percent of funds to directly award classroom teachers based on improvements in student achievement. Remaining funds may be used (1) as stipends for mentors or teacher coaches, teachers certified in hard-to-staff subjects, or who hold post-baccalaureate degrees, (2) as awards to principals based on improvements in student achievement, or (3) to implement components of Milken Family Foundation's Teacher Advancement Program.

### ***Florida's Merit Award Program (MAP)***

The 2006–07 budget approved by the Florida State Legislature included a \$147.5 million appropriation within the Florida Education Finance Program (FEFP) for

<sup>2</sup> Comparable Improvement (CI) is a measure that calculates how student performance on the TAKS mathematics and reading/English language arts tests has changed (or grown) from one year to the next, and compares the change to that of the 40 schools that are demographically most similar to the target school.

the Special Teachers Are Rewarded (STAR) performance-related pay program. Suspending the 2001 State Board of Education Performance Pay Rule, known as E-Comp, STAR was intended to reward instructional personnel for student performance, at least 50 percent of which had to be measured by standardized tests. Bonus amounts were equal to a minimum of 5 percent of an individual's salary, and were awarded to at least 25 percent of instructional personnel.

In March 2007, the STAR program was replaced with the Merit Award Program (MAP). All MAP plans are determined at the district level, with bonuses representing no less than 5 percent and no more than 10 percent of the average district salary for teachers. The award is distributed as a flat amount to all qualifying professionals in the district, regardless of their position or job responsibilities. Awards can be distributed to individual teachers or teaching teams, though they may not be distributed at the campus level.

At least 60 percent of all MAP awards must be based on student performance. Performance can be measured as total gains (as determined by pre- and post-testing) or proficiency levels on exams, or a combination of both. Teachers whose students take the state standardized tests must use the results of those exams in determining proficiency levels. Non-tested subjects and grades may use any national, state, or district-determined testing instrument that is aligned with the Sunshine State Standards.

In addition to awards for student performance, up to 40 percent of funds may be used to award professional practices, as measured by principal assessment. Professional practices include (1) the ability to maintain appropriate discipline, (2) outstanding subject matter knowledge and instruction, including the use of technology, (3) appropriate use of data and assessment results, (4) collaboration, (5) the Florida Educator Accomplished Practices, and (6) any other factors as determined by the district school board. All MAP plans must be board approved and adopted each year. The terms of each district's MAP are subject to negotiation with any existing teacher's association.

### **Minnesota's Q-Comp**

In July 2005, the Minnesota State Legislature approved Q-Comp, a performance-related pay program for teachers. Q-Comp incorporates both traditional career ladders and professional development for teachers, while advancing existing state standards through integration of measures to compensate teachers according to state approved measures of student achievement. Under Q-Comp guidelines, 60 percent of any compensation increase must be based on district professional standards and on classroom-level student achievement gains. Q-Comp presently operates in only 22 of 348 regular school districts across the state; however, 134 school districts have indicated intent to submit a Q-Comp proposal to the state within the next two years. District plans that are approved by the state department of education can be awarded up to \$260 more per student to support implementation and sustenance of their merit-based compensation plan.

### **Milken Family Foundation's Teacher Advancement Program (TAP)**

The Teacher Advancement Program (TAP) was developed in 1999 by the Milken Family Foundation, a philanthropic organization based in Santa Monica, California, to increase the number of highly qualified teachers, improve instructional effectiveness, and enhance student achievement.<sup>3</sup> TAP consists of four major components:

<sup>3</sup> TAP was recently renamed the National Institute for Excellence in Teaching (NIET).

(1) multiple career paths, (2) ongoing applied professional growth, (3) instructionally focused accountability, and (4) performance-related compensation.

### ***Multiple Career Paths***

TAP's multiple career paths position high quality teachers to pursue a variety of positions, advance professionally, and earn higher salaries without having to abandon the classroom. If teachers demonstrate consistent success, they have the opportunity to become career, master, or mentor teachers. This option of multiple career paths is important considering career advancement in U.S. public schools typically removes teachers from the classroom.

### ***Ongoing Applied Professional Growth***

TAP allocates time during the instructional day for teachers to meet and collaborate on instructional and curricular issues. These meetings are either group- or individual-focused, and often scheduled with a TAP-identified mentor or master teacher. TAP's mission of ongoing applied professional growth provides a framework for teachers to (1) set learning goals based on analyses of students' performance, (2) identify proven research-based learning strategies, (3) develop new instructional practices, (4) integrate practices into the classroom, and (5) monitor how well these strategies help improve student learning.

### ***Instructionally-Focused Accountability***

Instructionally-focused accountability refers to TAP's mechanism for evaluating teachers. In an effort to assess teacher performance appropriately, TAP employs a grading rubric to measure systematically a teacher's content knowledge, instructional methods, and student learning gains. These evaluations are ultimately used to determine a teacher's career ladder advancement within the school.

### ***Performance-Related Compensation***

TAP's performance-related compensation scheme rewards teachers across three dimensions: (1) student performance, (2) increased roles and responsibilities, and (3) classroom teaching performance. In linking pay to these three dimensions, TAP's remuneration mechanism represents a substantial departure from more traditional practices in which teacher pay is based on years of experience and highest degree held.

TAP currently operates in more than 125 schools in 9 states and 50 districts. Another 10 states are presently pursuing program implementation in routinely low-performing schools. In the aggregate, there are approximately 3,500 teachers and 56,000 students in TAP schools across the country. These numbers are anticipated to grow, recognizing not only that TAP was a principal partner in three Teacher Incentive Fund awards totaling an approximate \$67 million in funding over five years, but also that Texas' district-level educator incentive program permits schools to use Part II funds to implement TAP.

### **United States Department of Education's Teacher Incentive Fund**

In 2006, Congress appropriated \$99 million per annum to school districts, charter schools, and states on a competitive basis to fund development and implementation of principal and teacher performance-related pay programs. As part of the United

States Department of Education's (USDoE) Appropriations Act (P.L. 109-149), the Teacher Incentive Fund (TIF) is a direct discretionary federal grant program. Although USDoE estimated TIF dollars would fund an approximate 10 to 12 performance-related compensation projects with a per-project award size of \$8 million per year, a total of 16 awards expending less than half of the \$99 million appropriation were granted in fall 2006.

As indicated in Table 2, round one TIF awards provided funding to a variety of programs and locations. The largest award, totaling an approximate \$33.96 million over five years, was given to a consortium of six school districts in South Carolina to implement TAP in 23 high-needs schools. The smallest award, totaling an approximate \$1.63 million over five years, was given to a home-grown compensation model designed by two charter schools in California. Performance measures used to identify high-performing educators were equally eclectic, ranging from teacher and principal value-added student achievement gains to acquiring new knowledge and skills or taking on additional duties. USDoE plans to distribute the remaining \$43 million of year one appropriations in summer 2007 through a second grant competition already underway; however, strong opposition from the National Education Association and American Federation of Teachers, coupled with a joint funding resolution in the House of Representatives asking for a reduction of TIF appropriations to \$200,000 per year, has some questioning whether TIF will be reauthorized in 2008.

#### **THEORETICAL ARGUMENTS FOR AND AGAINST PERFORMANCE-RELATED PAY PROGRAMS**

As noted earlier, following the influential *A Nation at Risk* report a number of school districts experimented with performance-related pay programs as a means to improve student outcomes and reform the single salary schedule. Research on these programs highlighted the difficulty inherent in creating a reliable process for identifying effective teachers, measuring a teacher's value-added contribution, eliminating unprofessional preferential treatment during evaluation processes, and standardizing assessment systems across schools (for example, Hatry, Greiner, & Ashford, 1994; Murnane & Cohen, 1986).<sup>4</sup> Criticisms stemming from these generally short-lived programs have since stigmatized more recent attempts to devise and implement performance-related pay programs claiming further that teachers do not support merit pay policy (Murnane & Cohen, 1986; Darling-Hammond & Barnett, 1988).

Murnane and Cohen (1986) offer one of the more influential critiques of this early wave of merit-based pay programs. Drawing on personnel economics literature, they argued that merit-based pay plans of recent decades failed because teaching is not a field that lends itself to performance-related compensation, a perspective that Goldhaber, Hyung, DeArmond, and Player (2005) recently termed the "nature of teaching" hypothesis. Given its influence, and the fact that subsequent critiques have often raised the same arguments, we devote some attention to this article.

#### **The "Nature of Teaching" Hypothesis**

*Performance Monitoring.* A major argument against merit-based pay programs concerns the difficulty in monitoring teacher performance. According to Murnane and

<sup>4</sup> Past programs have also failed due to insignificant financial incentives for successful teachers, teacher unions who were opposed to alternative compensation systems, and lack of an evaluation process that could assess outcomes and recalibrate programmatic components to bring the program to scale (see, for example, Ballou & Podgursky, 1997; Ballou, 2001).

**Table 2.** Summary of Teacher Incentive Fund Awardees.

Program	Award Size	Bonus Amount and/or Range	Scope of Program	District and/or Institutional Partners
<i>Alaska Teacher and Principal Incentive Program</i>	\$5,191,449	School-wide growth awards (\$2,500–\$5,500); One-time bonuses for math teachers (\$750–\$3,000); Individual math learning plans (\$125–\$500); teacher evaluation (\$125–\$500); additional responsibilities (\$500–\$2,000); highly qualified math teacher bonus (\$2,000–\$4,000); highly qualified math teacher signing bonus (\$1,000–\$2,000)	3 rural school districts, totaling 27 schools	Lake and Peninsula School District; Kuspuks School District; Chugach School District; Alaska Department of Education and Early Development; Re-Inventing Schools Coalition
<i>Recognizing Excellence in Academic Leadership (REAL) (Chicago Public Schools)</i>	\$27,336,693	Teachers (approx. \$4,000); Principals (\$5,000); Other staff (\$1,000)	10 schools per year, totaling 40 schools	National Institute for Excellence in Teaching; Teacher Advancement Program; Mathematica Policy Research, Inc.
<i>Dallas Independent School District (Texas)</i>	\$22,385,899	Principal bonuses (\$7,500–\$10,000); Teacher bonuses (approx. \$1,000)	District-wide initiative, approx. 214 of 220 schools	Dallas Independent School District; Dallas Achieves Commission
<i>Denver Public Schools (Colorado)</i>	\$22,674,393	Teachers knowledge and skills (\$666–\$2,997); Professional evaluation (\$333–\$999); Market incentive (\$999); Student growth (\$333–\$999); Principals (\$5,000); Assistant principals (\$3,750)	District-wide initiative to expand ProComp, approx. 132 of 150 schools	New Leaders for New Schools; Denver Public Schools; Denver Classroom Teachers Association

(Continued)

**Table 2.** (Continued)

Program	Award Size	Bonus Amount and/or Range	Scope of Program	District and/or Institutional Partners
<i>Eagle County School District Performance-Based Compensation Program (Colorado)</i>	\$6,779,204	Teacher skills and knowledge (max. \$1,300); Individual student achievement for teachers and school-wide for principals on non-high-stakes exam (max. \$650); School-wide achievement on state exam (max. \$650); Classroom bonuses for teachers may total \$10,000	District-wide initiative to expand Teacher Advancement Program; 13 schools are eligible	National Institute for Excellence in Teaching; Teacher Advancement Program
<i>Mission Possible (Guilford County Schools, North Carolina)</i>	\$8,000,005	Recruitment and retention bonuses (approx. \$2,500–\$10,000); Performance-based incentives for teachers and principals (approx. \$2,500–\$5,000)	District-wide initiative, 20 schools in the 2006-07 school year	Guilford County Schools; Value Added Research and Assessment Center at the University of Tennessee-Knoxville; SES Institute, Inc.; SERVE University of North Carolina
<i>Strategies for Motivating and Rewarding Teachers (Project SMART) (Houston Independent School District, Texas)</i>	\$11,781,323	Comparative improvement on TAKS (\$125–\$500); Student- and campus-level progress on Stanford/APrenda (\$250–\$1,000); Individual student progress on TAKS (\$500–\$1,000)	District-wide initiative, 109 of 306 schools	Houston Independent School District
<i>The New 3R's: Rigor, Results and Rewards (Mare Island Technical Academy, California)</i>	\$1,626,392	Professional development pay for teachers (\$500–\$5,000); Core Learning Support (\$3,000–\$4,000); Stretch Learning Support (\$1,000–\$2,400); Student Engagement Support (\$1,000–\$2,500); Personal Skill Development Support (\$2,500)	2 schools	MIT Academy

<p><i>Effective Practice Incentive Fund (EPIF) (Memphis City Schools, Tennessee)</i></p>	<p>\$13,836,434</p>	<p>Principals meeting Silver or Gold standard (\$10,000 or \$15,000); Assistant principals meeting Silver or Gold standard (\$7,500 or \$10,000); Teachers in Gold schools (\$7,500); Various other awards (\$750 or \$1,000 total; \$125 or \$175 per student)</p>	<p>17 schools</p>	<p>New Leaders for New Schools; Mathematics Policy Research Inc.; Teachscape; Standards &amp; Poors'</p>
<p><i>Effective Practice Incentive Fund (Consortium of Charter Schools)</i></p>	<p>\$20,752,420</p>	<p>Schools meeting Silver or Gold Standard (\$150 per pupil; \$250 per pupil); Principals meeting Silver or Gold Standard (\$15,000 or \$20,000); Teachers in Silver or Gold Schools (\$750; \$1,500); Exemplary teacher in Gold schools (up to an additional \$10,000)</p>	<p>47 charters schools in nine states and Washington, DC</p>	<p>New Leaders New Schools; Knowledge is Power Program; Achievement First; Uncommon Schools; Aspire Public Schools; Yes College Prep Schools; Mathematica Policy Research, Inc.; NewSchools Venture Fund</p>
<p><i>Northern New Mexico Performance-Based Compensation Program</i></p>	<p>\$7,647,796</p>	<p>Personnel in each of the participating districts will design a unique performance-based compensation plan during the 2007–08 planning year. Criteria used to identify high-performing teachers and principals will be based on specific tools and resources identified in the Baldrige in Education framework.</p>	<p>4 rural school districts</p>	<p>Northern New Mexico Network for Rural Education; Espanola Public School District; Springer School District; Cimarron School District; Des Moines School District; Wexford</p>
<p><i>Philadelphia Teacher and Principal Incentive Fund Project (Pennsylvania)</i></p>	<p>\$20,500,215</p>	<p>Cash bonuses will be provided to teachers and principals. Details about the incentives will be finalized during the 2007–08 planning year.</p>	<p>20 schools</p>	<p>Philadelphia Federation of Teachers; Commonwealth Association of School Administrators; Academy for Leadership in Philadelphia Schools; SAS Institute, Inc.</p>
<p><i>South Carolina Teacher Incentive Fund</i></p>	<p>\$33,959,740</p>	<p>Teachers (\$2,000–\$4,000); Mentor Teachers (\$5,000); Master Teachers (\$10,000); Principals (\$5,000–\$10,000)</p>	<p>6 school districts, totaling 23 schools</p>	<p>National Institute for Excellence in Teaching; Teacher Advancement Program; Anderson Research Group; SAS Institute, Inc.</p>

(Continued)

**Table 2.** (Continued)

Program	Award Size	Bonus Amount and/or Range	Scope of Program	District and/or Institutional Partners
<i>The Ohio Teacher Incentive Fund</i>	\$20,223,270	Teacher pay based on knowledge and skill (\$2,000); Teacher pay based on campus-level performance (\$2,000); Teacher pay based on student achievement (\$2,000); Master teacher (\$7,300); Mentor teacher (\$3,500); Principals (\$2,000)	4 urban school districts	Ohio Department of Education; Cincinnati City Schools, Cleveland Municipal School District; Columbus Public Schools; Toledo Public Schools; National Institute for Excellence in Teaching
<i>Effective Practice Incentive Fund (Washington, DC Public Schools)</i>	\$14,118,543	Gold schools receive \$250 per pupil; Silver schools receive \$150 per pupil; Principals meeting Silver or Gold Standard (\$15,000 or \$20,000); Exemplary teacher (up to an additional \$10,000); Other bonuses (\$750–\$2,250); Low-performing schools that show potential (per capita awards ranging from \$75,000–\$125,000)	25 schools	New Leaders for New Schools; District of Columbia Public Schools; Washington Teachers' Union; Mathematica Policy Research; Teachscape; Standards & Poor's
<i>Weld County School District (Colorado)</i>	\$3,670,133	Principals (sustainable 3% raise); Teachers with students scores at achievement level (sustainable 5% raise); Mentor teacher or additional duties (1% raise); School-wide achievement level (1% raise)	4 schools	META Associates

Cohen, teacher performance is more difficult to monitor than performance in many other professions because output is not readily measured in a reliable, valid, and fair manner.<sup>5</sup> Unlike, say, the sales of a salesman or the billable hours of a doctor or lawyer, the output of a teacher is not marketed. Thus, it is argued that the education sector cannot readily measure the value of the services provided by an individual teacher or group of teachers, since achievement is influenced by many factors beyond the instructor's control.

While this argument no doubt had merit at the time, its relevance may be waning given the major advances in data systems being put in place in states and districts. States and districts are rapidly developing massive longitudinal student-level databases that permit more precise estimation of value-added contributions at the building, grade, and, in a growing number of states, teacher level. Furthermore, the USDoE has also created a competitive grant program to encourage states to develop longitudinal data systems that support value-added measurement. As data and measurement systems grow in sophistication, the measurement of teacher and school performance will likely become considerably more reliable.

In spite of these technological advances, to the extent that these new performance-related pay programs rely on estimates of teacher value added, it is important to note that there are still concerns about the statistical reliability and robustness of these value-added estimates. Some researchers express caution in interpreting teacher effects purely as an attribute of the teacher without consideration of the school context and the stability of these measures over time (McCaffrey, Lockwood, Koretz, & Hamilton, 2003; McCaffrey et al., 2004; Ballou, Sanders, & Wright, 2004; Ballou, 2005; Koedel & Betts, 2005).

### *Team Production*

A second argument against merit-based pay programs concerns team production. To a considerable extent, teachers work as members of a team. Introducing performance-related rewards at the individual teacher level might reduce incentives for teachers to cooperate and, as a consequence, reduce rather than increase school performance. Some scholarship argues that the team dynamic can be destroyed between teachers as well as between teachers and administrators, especially if administrators are put in a position of rewarding individual teacher performance.<sup>6</sup> Of course, this is a criticism of individual performance-related pay programs. A performance bonus given to an entire team of teachers would not undermine team morale. This is especially germane considering most teachers work in relatively small teams, and economic literature suggests team incentives may work quite well in small teams because there is mutual monitoring coupled with an easy information flow among team members and options for subjects to reciprocate among each

<sup>5</sup> Similar points were made following implementation of the single salary schedule in 1921. In one of the first education finance textbooks published, Moehlman (1927) argued for development of a salary schedule that provides "as scientifically possible for the best returns to society for the increasing public investment" by approaching salaries from "its economic and social aspects and not in terms of sentimentality." However, he concluded that an objective and standardized system for determining merit did not exist, nor was there capacity to develop a school- or district-level system. Consequently, the most relied upon method for evaluation of merit was a single salary schedule called the "position-automatic schedule," which automatically advanced teachers by annual pay increments ranging between \$50 and \$200 after their first year of teaching until a predetermined maximum salary was reached somewhere between 10 and 15 years of service.

<sup>6</sup> A similar argument was developed concerning a performance-related pay scheme that was introduced in England and Wales during the 2000–2001 school year (Adnett, 2003).

other within the team (Kandel & Lazear, 1992; Vyrastekova, Onderstal, & Koning, 2006).<sup>7</sup>

There are other ways around the teacher production argument. A bonus scheme does not have to be a fixed-tournament. In a pay-for-performance experiment designed by the National Center on Performance Incentives at Vanderbilt University's Peabody College and implemented in the Metropolitan Nashville Public Schools, teachers are judged against a standard based on past performance of teachers in the district. This standard was determined at the beginning of the experiment and will remain fixed for the duration of the experiment. This means that all volunteer teachers in the treatment group have the opportunity to improve and, in principle, all could end up exceeding this standard and be awarded a bonus. Of course, this style of bonus scheme creates greater financial exposure for the district.

### *The Multitasking Problem*

Another theoretical criticism of performance-related pay programs in the literature concerns the issue of multitasking when relying on tests or other quantitative measures of teacher performance (for example, Holmstrom & Migrom, 1991; Hannaway, 1992; Dixit, 2002). This problem arises when the performance of a worker has multiple dimensions, only some of which are measured and incentivized. When there is structural misalignment between an organization's overall mission and the activity to which incentives are attached, not surprisingly, employees tend to shift work toward the metered, rewarded activity, and away from other important activities.

An important concern in this regard is "teaching to the test"—an education catchphrase used to describe narrowing of curriculum in an effort to elevate student test scores that was first used to critique performance contracting in education during the 1960s. Teachers' contributions to student learning are multifaceted; however, if an inordinate amount of weight is placed on student assessments, then other valuable activities might be slighted. In the general personnel literature, the solution to the multitasking problem is to diversify the measures used to evaluate performance, such as supervisor evaluations or other broad-based assessments to complement quantitative measures.

Incentive schemes that tie teacher pay to achievement gains by students—whether at the individual teacher or team level—may create more opportunities for cheating or other opportunistic behavior in the long run. For example, studies of high-stakes accountability systems have documented teachers focusing excessively on a single test and educators altering test scores and/or assisting students with test questions (Goodnough, 1999; Koretz, Barron, Mitchell, & Stecher, 1996; Jacob & Levitt, 2003). Related analyses have found evidence of schools' strategic classification of students as special education and limited English proficiency (Deere & Strayer, 2001; Figlio & Getzler, 2002; Cullen & Reback, 2006; Jacob, 2005), use of discipline procedures to ensure that low-performing students will be absent on test day (Figlio, 2003), manipulation of grade retention policies (Haney, 2000; Jacob,

<sup>7</sup> Of course, rewards to an entire team, rather than to each member, introduce the "free rider" problem. If a team member exerts effort and raises overall team output by  $X$ , he will only receive a return of  $X/N$ , where  $N$  is the size of the team. Clearly, as  $N$  grows, the performance incentive shrinks rapidly. For a discussion of this problem, see Prendergast (1999). Professional peer pressure may act as an offsetting effect. Gaynor, Rebitzer, and Taylor (2004) analyze an interesting case study of physician teams within an HMO. While the physicians, much like teachers, operated largely independently, a "high powered" group incentive seemed to offset free-rider problems. On the issue of peer pressure, see also Kandel and Lazear (1992).

2005), misreporting of administrative data (Peabody & Markley, 2003), and planning of nutrition enriched lunch menus prior to test day (Figlio & Winicki, 2005).

These findings suggest that performance pay is not a perfect substitute for monitoring; almost any incentive system can be gamed, and behavior will need to be monitored. Monitoring is not an impossible task. Jacob and Levitt (2003) have developed statistical algorithms to detect suspicious answer strings and test score fluctuations on standardized assessments, and such firms as Utah-based Caveon Test Security specialize in test security and fraud detection. Research findings also demonstrate that it is useful to have multiple indicators in pay-for-performance systems, if for no other reason than the fact that not all indicators will be equally susceptible to gaming.<sup>8</sup>

### **Payment for Input and Payment for Output**

Edward Lazear, a major contributor to the “new personnel economics” literature, provides a useful conceptualization of the performance-related pay problem in K–12 education and assesses the economics of alternative teacher compensation regimes which he terms payment for input and payment for output (Lazear, 2003). In the absence of externalities or information problems, payment for output always trumps payment for input in terms of raising overall productivity. Two principle reasons—hiring practices and labor market selection—are discussed below.

#### *Hiring Practices*

District and building administrators are restricted by informational deficiencies when hiring teachers and other instructional staff. This necessitates that principals use noisy signals of “true” teacher effectiveness (for example, years of experience, highest degree held, past-employer recommendations).<sup>9</sup> Informational deficiencies in the hiring process are ameliorated in most professions by subsequent employee performance assessments and as pay raises become more closely tied to actual productivity, thereby lessening dependence on input-based indicators for employees (Altonji & Pierret, 1996). Of course, the single salary schedule, along with teacher tenure, makes it difficult for pay and performance to align after hire. For example, if only effective teachers have their contracts renewed, then pay on the basis of seniority would tend to align pay and performance. While such a mechanism may work in the first probationary years of teacher employment, after teachers earn tenure, contract nonrenewal can only be triggered by severe malfeasance on the part of the employee.

#### *Labor Market Selection*

Lazear also discerned a more subtle, but important, factor in the gains from a performance-related, or output-oriented, pay system that arises from labor market selection. A performance-related pay program will tend to attract and retain individuals who are particularly good at the activity to which incentives are attached, and repel those who are not. He noted that this effect on the workforce can be very

<sup>8</sup> Dixit (2002, p. 719) provides a laconic assessment of the evaluation problem, “To sum up, the system of public school education is a multitask, multiprincipal, multiperiod, near-monopoly organization with vague and poorly observable goals.”

<sup>9</sup> Casual empiricism (we are aware of no survey data) suggests that most principals are also paid off of salary schedules similar to those for teachers (that is, with pay set by experience and education credits/degrees). Clearly, the information problem would be improved if principals were given stronger incentives to extract the productive “signal” from the “noise” concerning teacher productivity.

important in explaining productivity gains. For instance, in one of his own case studies outside of teaching, Lazear (2000) found that sorting effects were both substantial and roughly equal in magnitude to motivation effects. In other words, while the incentive system raised the productivity of the typical worker employed, it also raised the overall quality of the workforce.

Some researchers speculate that this selection effect will be a significant factor in teacher labor markets. Studies of teacher turnover, for example, consistently find that high-ability teachers are more likely to leave teaching than low-ability teachers, where ability is defined by a teacher's performance on the ACT (Podgursky, Monroe, & Watson, 2004) or National Teacher Exam (Murnane & Olsen, 1990). This trend may be due to constraints on wages rather than the attraction of other market opportunities.

A recent provocative study by Hoxby and Leigh (2004) found evidence that the migration of high-ability women out of teaching between 1960 and the present was primarily the result of the "push" of teacher pay compression—which took away relatively higher earnings opportunities for teachers—as opposed to the pull of greater nonteaching opportunities. Although the remunerative opportunities for teachers of high and low ability grew outside of teaching, it was pay compression within the education system that accelerated the exit of higher ability teachers. To the extent that these high ability teachers were more effective in the classroom, a performance-related pay program likely would have kept more of them in teaching.

Lazear's selection arguments also undermine one other critique of teacher merit pay by Murnane and Cohen (1986). These authors argue that in any effective merit pay system, employers should be able to tell workers what they need to do in order to become more effective. In other words, if ineffective teachers do not know what to do in order to raise their performance, and supervisors cannot provide such guidance, then the motivational effect of merit pay will be nil. However, if the underlying range of teacher effectiveness is great (and evidence considered below suggests that this is the case), then simply tying pay to performance may significantly raise performance even if no individual teacher's productivity rises, simply through differential recruitment and retention of high-performing, high-paid teachers.

## EMPIRICAL RESEARCH

Economic theory can take us only so far in hypothesizing about the effect of teacher performance pay. Ultimately, we must turn to the data. In this section, we review four strands of research relevant to the debate: (1) the teacher effects literature, (2) studies linking teacher effects to performance assessments, (3) direct evaluations of individual- and group-based performance pay programs, and (4) evidence from survey research on traditional public, public charter, and private schools compensation practices.

### Teacher Effects Studies

Over the last decade, researchers have begun to exploit massive longitudinal student achievement data files to undertake "value-added" studies of teacher effectiveness. Beginning with William Sanders' work in developing the Tennessee's Value Added and Assessment System (Wright, Horn, & Sanders, 1997; Ballou, Sanders, & Wright, 2004), teacher value-added studies have expanded to states such as Texas (Rivkin, Hanushek, & Kain, 2005), and to large school districts such as New York City (Kane, Rockoff, & Staiger, 2006; Boyd, Grossman, Lankford, & Loeb, 2006),

Chicago (Aaronson, Barrow, & Sanders, 2003), and San Diego (Koedel & Betts, 2005). These studies have consistently found evidence of large variation in achievement gain scores between classrooms and teachers, suggesting that teachers can have a substantial effect on student achievement growth, particularly if teacher effects are cumulated over a number of years. Other studies have found evidence of persistence in teacher effects over time.<sup>10</sup>

While researchers have found substantial variation in teacher effects within school districts, and even within schools, they also have consistently found that these effects are largely unrelated to measured teacher characteristics such as the type of teaching certificate held by the teacher, a teacher's level of education, licensing exam scores, and experience beyond the first two years of teaching. Indeed, nearly every researcher conducting rigorous teacher-effect studies has taken note of this fact (for example, Kane, Rockoff, & Staiger, 2006; Rivkin, Hanushek, & Kain, 2005; Aaronson, Barrow, & Sanders, 2003; Goldhaber & Brewer, 1997). For example, in a large-scale study of certification status and new teacher effectiveness in New York City Public Schools, Kane, Rockoff, and Staiger (2006, p. 40) write:

There is not much difference between certified, uncertified, and alternately certified teachers overall, but effectiveness varies substantially among each group of teachers. To put it simply, teachers vary considerably in the extent to which they promote student learning, but whether a teacher is certified or not is largely irrelevant to predicting their effectiveness.

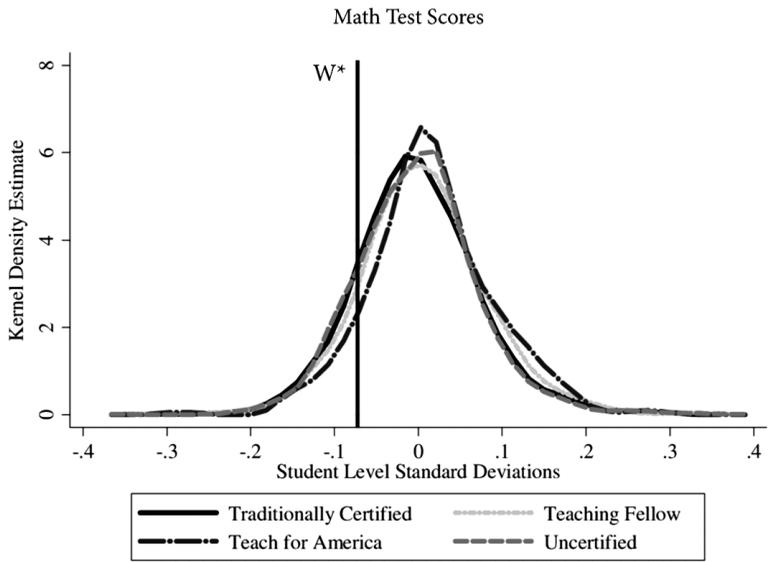
We have reproduced from their study a chart of estimated teacher effects that demonstrates clearly this point. Figure 1 reports variation in estimated teacher effects for new teachers by type of teaching certificate held.<sup>11</sup> It is readily apparent that the distributions overlap almost entirely, illustrating negligible differences between certified, uncertified, and alternately certified teachers. With respect to merit-based pay, note the very wide variation in teaching effectiveness within each certification group. Any policy that can retain and sustain the performance of teachers in the upper tail of the distribution, and enhance the performance of or counsel out teachers in the lower tail, possesses potential for substantial impact on student growth.

A study of Chicago Public School teachers by Aaronson, Barrow, and Sanders (2003) further illustrates this point. Like other such studies, this work was based on a very large longitudinal file of student achievement scores linked to teachers. What makes this study unique is that the authors had very extensive administrative data on teacher characteristics heretofore unavailable in other studies, including education, experience, types of teaching licenses, and selectivity of teachers' undergraduate college. Aaronson and colleagues found that over 90 percent of teacher effects are not explained by any of these measured teacher characteristics, thus highlighting inefficient resource allocation practices maintained by the single salary schedule.

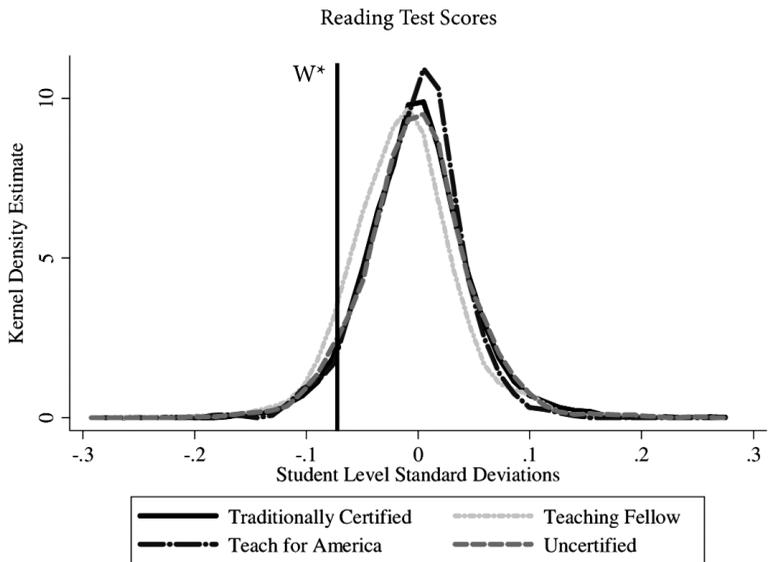
The fact that most studies to date conclude that teacher graduate degrees—the most common educational credential—have a marginal effect at best on student achievement (Hanushek, 2003), reiterates that there is little empirical support for

<sup>10</sup> Koedel (2007) explicitly considers the question as to whether the signal to noise ratio of his high school estimated teacher effects is sufficiently high to warrant their use in personnel decisions. Ultimately, his assessment is favorable.

<sup>11</sup> Another team examining New York City Public Schools reached a similar finding (Boyd et al., 2006).



Note: Shown are estimates of teachers' impacts on average student performance, controlling for teachers' experience levels and students' baseline scores, demographics and program participation; includes teachers of grades 4-8 hired since the 1999-2000 school year.



Note: Shown are estimates of teachers' impacts on average student performance, controlling for teachers' experience levels and students' baseline scores, demographics and program participation; includes teachers of grades 4-8 hired since the 1999-2000 school year.

Source: Kane, Rockoff, & Staiger (2006, fig.6)

**Figure 1.** Variation in teacher effectiveness by type of teacher certificate: New York City Public Schools, 1998-99—2004-05.

the current credential-based teacher compensation system. The teacher certification program developed by the National Board for Professional Teaching Standards has been promoted as an alternative to merit-based pay programs (National Commission on Teaching and America's Future, 1996). However, even here the evidence concerning performance is mixed (Goldhaber & Anthony, 2007; Sanders, Ashton, & Wright, 2005; Harris & Sass, 2007; Clotfelter, Ladd, & Vigdor, 2007).<sup>12</sup>

The widely dispersed and idiosyncratic nature of teacher effects has important consequences for the performance pay debate. On the one hand, it suggests that credential-based pay reforms are not likely to have substantial effects on student achievement. On the other hand, it points to substantial student achievement gains if the mix of low- and high-performing teachers can be altered. A policy that ties pay to performance over time would likely recruit or retain more teachers in the upper tail of ability into the teaching workforce, and encourage low-productivity teachers to either improve or leave for nonteaching positions.

Suppose, for example, that a pay system is monotonically related to the teacher effectiveness measured on the horizontal axis in Figure 1. Further assume that all teachers have identical nonteaching earnings, illustrated by  $W^*$ , indicating the teaching productivity equivalent of the alternative wage. Ignoring, for the moment, nonpecuniary preferences for teaching versus other jobs, teachers with productivity to the right of the vertical bar  $W^*$  would move to, or stay in, teaching, and those to the left would exit. Teacher turnover would thus become part of a virtuous cycle of quality improvement, rather than a problem to be minimized.

### Teacher Effects, Performance-Related Awards, and Principal Evaluations

The multitasking problem identified in previous critiques of performance-related pay programs makes the case for subjective assessment by supervisors being part of a multi-factor assessment system of teachers. The assumption is that the supervisor evaluation picks up important teacher behaviors that student achievement gains do not. Glewwe, Ilias, and Kremer (2004), discussed in more detail below, raise this issue in the context of their assessment of a short-lived teacher incentive scheme in Kenya. However, independent of the multitasking issue, it is useful to know the strength of association between supervisor evaluations and student achievement gains, where this relationship can be measured. This at least increases our confidence in supervisor measures in contexts in which they cannot be validated by test score gains (for example, music or social studies teachers).

A small number of studies have examined the relationship between these subjective assessments and teacher performance as determined by student test score gains. As early as the mid-1970s, a number of educational researchers concluded that principal evaluations are a reliable guide to identifying high- and low-performing teachers as measured by student test score gains (Armor et al., 1976; Murnane, 1975). More recently, Sanders and Horn (1994) demonstrated that there is a strong correlation between teacher effects as measured by the Tennessee Valued-Added Assessment System and subjective evaluations by supervisors.

In a particularly rigorous study focused entirely on the predictive validity of supervisor evaluations, Jacob and Lefgren (2005) assessed the relationship between teacher performance ratings, as identified on a detailed principal evaluation,

<sup>12</sup> Angrist and Guryan (2005) estimate the effect of state teacher testing requirements on teacher wages and teacher quality as measured by educational background and conclude that state-mandated teacher testing increases teacher wages with no corresponding increase in quality.

and teacher effects, as measured by student achievement gains. In estimating teacher effectiveness measures for 202 teachers in grades 2 through 6 in math and reading, Jacob and Lefgren found a statistically significant and positive relationship between value-added measures of teacher productivity and principals' evaluations of teacher performance.

Another interesting dimension of this study was an "out of sample" prediction of 2003 student achievement scores based on principal ratings and teacher value-added estimates from 1998 through 2002. Students had higher average scores in math and science if they had teachers with not only higher measured teacher effectiveness in prior years but also higher principal ratings. Jacob and Lefgren demonstrated further that the principal evaluation remained a statistically significant predictor of current student achievement even when teacher value-added (in the previous year) was added to the model. This finding suggests that principal evaluations provide an important independent source of information on teacher productivity.

Although these studies tend to indicate that principals are relatively adept at identifying above- and below-average teachers, it is important to question whether effective assessment practices persist in a performance pay regime. The fact that a principal identifies a teacher as "inadequate" on an anonymous survey does not mean necessarily that she will do so in a high-stakes environment. Indeed, a primary reason the single salary schedule replaced the grade-based compensation system was that subjective measures used to reward teachers were highly susceptible to gender and racial discrimination as well as nepotism.<sup>13</sup>

Two studies shed some light on whether "old style" merit plans that were based in part on supervisor evaluations are positively associated with measures of teacher productivity. Cooper and Cohn (1997) found that classroom gain score measures were higher for teachers who received merit pay awards in South Carolina. However, in the individual pay component of the plan, teachers who applied for the award were evaluated on four criteria, one of which was a performance evaluation and another was evidence of superior student achievement gains. Accordingly, these findings are not considered a strong test of the hypothesis.

A more recent study, by Dee and Keyes (2004), examined the relationship between career ladder bonuses and student achievement gains in Tennessee's Project STAR data. What makes this study unique is that students were randomly assigned to teachers in the experiment. While the focus of the STAR experiment, and subsequent research studies, has been the effect of class size, these researchers take advantage of the fact that students were also randomly assigned to Tennessee career ladder teachers. Teachers advanced on the career ladder rungs primarily on the basis of subjective evaluations typically conducted by a local principal. Dee and Keyes found that teachers with career ladder status (that is, those who have passed one or more evaluations) were more effective than teachers who had not obtained career ladder status.<sup>14</sup>

While no single study is definitive in this area, a small literature has developed showing that principal evaluations and performance-related promotions and/or

<sup>13</sup> Marsden and Belfield (2006) report results from a panel survey of classroom and head teachers' views of a performance-related pay system in England and Wales. They found the number of teachers who believed managers would use subjective evaluations to reward favorites dropped from over 50 percent when the program was first introduced to less than 20 percent four years later.

<sup>14</sup> Studies of student achievement gains in Cincinnati, a Nevada school district, and a large Los Angeles charter school provided support for the validity of a widely used teacher assessment framework (Milanowski, 2004; Kimball, White, Milanowski, & Borman, 2004; Gallager, 2004).

awards based in whole or in part on principal evaluations are associated with higher classroom teacher effectiveness as measured by student achievement gains. This finding does not address the multitasking problem that recognizes the presence of many valuable attributes to teacher performance not adequately measured by state assessments. However, to the extent that principals' subjective assessments capture these attributes, it is useful to know that these evaluations are also correlated with teacher productivity when measured by student gain scores.

### **Assessments of Performance-Related Pay Programs**

While there have been numerous experiments in individual and group incentive pay for teachers over the years, the evaluation literature is very slender. In Table 3 we list all of the studies we found in the literature that employ a conventional treatment and control evaluation design, with pretreatment benchmark data on student performance for both groups. For each study included in our assessment, we summarize key characteristics of each study or program being evaluated, including whether it was a school-wide or an individual incentive bonus, and the size of the bonus. The last column represents our assessment of the outcome of the study.

We have not attempted a more sophisticated "meta-analysis" or analytical synthesis. Nor have we attempted to compute "effect sizes." There are several reasons for this. First, unlike education inputs such as class size or teacher education, the "treatments" in these studies vary considerably from study to study. Ideally, one would want a set of studies that could yield estimates of student achievement gains (if any) per thousand dollars of bonus. Unfortunately, these programs are sufficiently diverse that such calculations are not possible.<sup>15</sup> Second, the outcome variables analyzed also vary considerably, sufficiently so that we do not feel it is useful to convert them to a common metric.

It is interesting that, in spite of these limitations, the overall findings in Table 3 stand in rather sharp contrast to the mixed but generally negative findings of production function studies of the effect of teacher characteristics such as teacher certification, education, or class size (Hanushek, 2003). In most of these studies, the incentive regime was found to yield positive student achievement effects. Moreover, in every study, the effect of incentives was to raise the level of the variable being incentivized. However, as we shall see, some of these incentive systems were not well designed and not always focused on student achievement.

We broadly divide the studies by what we judge to be the rigor of the evaluation, ranging from two randomized field trials to conventional matched comparison group designs that rely on nonexperimental data to identify treatment and control groups, or otherwise estimate program effects.<sup>16</sup> It should be recognized that because there is significant variation in the character of the programs being evaluated as well as the process determining participation, none of which is under control of the researcher, it follows that the data and methods available for rigorous estimation of program effects varies widely as well. The four most rigorous evaluations to date come from abroad.

<sup>15</sup> There are two exceptions to this statement (Lavy, 2002, 2004).

<sup>16</sup> Our ordering in no way orders the skills of the researchers. There was significant variation in the programs being evaluated (for example, targeted versus broad-based), data, and data available for estimating program effects. Since only one of these was a true experiment, researchers had to make the best use of available nonexperimental data.

**Table 3.** Quantitative studies of the causal effect of teacher incentive programs on measures of student achievement.

Study	Sample	Time Span of Study	Type of Teacher Incentive	Size of Incentive (per teacher)	Outcome Variable	Results
Muralidaran and Sundararaman (2006)	500 Rural Indian Primary Schools, randomly assigned 100 individual incentive 200 extra resource 100 control	2004–2005	Individual and School-wide	Average 4% group, 5% individual	Math and language, various primary grades	Positive
Glewwe, et.al. (2004)	100 Primary schools, rural Kenya, 50 randomly chosen for program	1997–1999	School-wide	Up to 43 percent of monthly salary	Grade 4, 8 test scores	Mixed
Lavy (2002)	Israel, high schools	1993–1995 to 1996–1997	School-wide (tournament)	\$200–\$715	Test scores, pass rates, dropout rates, course-taking	Positive
Lavy (2004)	Israel, high schools	1999–2001	Individual (tournament)	\$1,750–\$7,500+ <sup>a</sup>	Pass rates and test scores	Positive
Figlio and Kenny (2006)	NELS-88 matched to FK survey or 1993-94 SASS, 12th grade public and private schools	1993	Individual	Varied within sample	12th grade, composite reading, math, science, and history score	Positive
Winters, Ritter, Barnett, and Greene (2006)	2 treatment and 3 control elementary schools in Little Rock, Arkansas	2002–2003 to 2005–2006	Individual	\$1,800–\$8,600	Grade 4, 5 math test scores	Positive
Atkinson, et al. (2004)	UK High Schools	1997–2002	Individual	>9% in salary base	English, science, math assessments	Positive
Ladd (1999)	Dallas grade 7 schools relative to other Texas urban districts <sup>b</sup>	1991–1995	School-wide (tournament)	\$1000	Math and reading test scores, dropout rates	Positive
Clotfelter and Ladd (1996)	2 MI alternative high schools (1 treatment, 1 control)	1994–1995 to 1998–1999	Individual	Up to 20 percent of base pay	Course completion rates, pass rates, daily attendance, GPA	Mixed

<sup>a</sup>These are winnings per class. However, a teacher could enter multiple classes.

<sup>b</sup>Incentive applied to all schools but data limitations only permitted examination of grade 7 effects.

We begin our discussion with the two random assignment studies. Muralidharan and Sundararaman (2006) report first-year results from a World Bank–sponsored experiment on performance pay in rural Indian schools. This is a first-year report on a project that is slated to run until 2011. The researchers randomly sampled 500 rural schools in a large Indian state (Andhra Pradesh) and assigned them to one of four treatment groups or a control group, with each group comprising 100 schools. One of the treatment groups had an individual teacher pay bonus system tied to student test score gains, and another had a school-wide bonus tied to test score gains. The average bonus payments in either incentive scheme were small relative to base pay (4–5 percent), but the maximum possible payment amounted to a substantial share of pay (roughly 14 and 29 percent of pay for group and individual, respectively). The two other treatment groups were provided additional resources (teacher aides or an extra block grant), and a control group received no additional resources.

Muralidharan and Sundararaman estimated the incentive program effects to be 0.19 and 0.12 in math and languages, respectively, relative to the control group. They found no evidence of adverse effects of the program on other test scores or teacher morale, and no significant difference in program effects between the group and individual incentive schools.<sup>17</sup> Since the researchers attempted *ex ante* to hold incremental spending in the different treatment groups the same, another interesting finding is that the point estimates of the incentive schemes yielded test score gains exceeding those of the added-resource treatments. Thus, the incentive schemes were not only found to be effective, but cost-efficient, relative to added resource schemes. (This finding is replicated in Lavy's Israel studies discussed below.)

Glewwe, Ilias, and Kremer's work (2004) is a second random assignment study of incentive pay in rural schools. Fifty schools were chosen at random for participation from among 100 relatively low-performing rural Kenyan primary schools. The teacher bonuses were school-wide and tied to student pass rates on district exams in a variety of subject areas.<sup>18</sup> The bonuses were substantial, ranging from 21 to 43 percent of monthly pay. However, the limited duration of the program (originally announced as one year, later extended to two) did not permit Glewwe and colleagues to examine the long-run effects.

Glewwe and colleagues found increased pass rates on the district exams during the two years of the program, but those gains did not persist into a third year after the end of the program, which they took as evidence of gaming/opportunistic behavior on the part of teachers. While targeted teachers provided more after-school test preparation, the researchers found no evidence of differences in homework assignment or pedagogy. Of particular concern was that the program seemed to have no effect on a significant teacher absenteeism problem, which runs at least 20 percent in these rural Kenyan schools.

The researchers' strongest evidence of multitasking seems to be that student learning gains in treatment condition schools did not persist. While Glewwe and coworkers link this finding with the hypothesis that incentive programs will lead to manipulating short-run scores, it is also important to note that, like many first

<sup>17</sup> The authors also examined the effectiveness of the incentive schemes for students at different points of the achievement distribution or by socioeconomic characteristics of the students. They found no evidence of heterogeneous effects.

<sup>18</sup> Subjects tested included English, math, Swahili, GHCR (geography, history, and Christian religion), arts-crafts-music, and home science/business education.

generation experiments, the program was short-lived. It is possible that a more sustained program would have produced more sustained student learning effects. Furthermore, it is interesting that the teachers obtained these effects without coming to work with greater frequency. If teacher absenteeism is seen as a problem, clearly, a better designed system would incentivize attendance as well as students' test scores.<sup>19</sup> We score this study as "mixed."

Lavy (2002, 2004) has undertaken two careful studies of performance "tournaments" in Israel.<sup>20</sup> In both of these studies, the program was designed to raise pass rates on high school exit exams in low socioeconomic high schools in Israel. Although schools were not randomly assigned to a control or treatment condition, both programs were implemented using three formal assignment rules (that is, grade range, past performance, and matriculation rate) permitting for a more rigorous regression-discontinuity evaluation design. The Israeli Teacher-Incentive Experiment was also carefully designed to minimize gaming or other opportunistic behavior on the part of teachers and school administrators (for example, performance measures based on the size of the graduating cohort in order to discourage schools from encouraging transfer or dropout of poor students, or by placing poor students in nonmatriculation tracks).

Lavy's (2002) first study considered a tournament in which a selected group of low-performing high schools competed on the basis of school-wide performance. The top third of schools as determined by their year-to-year improvement in test scores were given awards ranging in size from \$13,250 to \$105,000. Teacher bonuses ranged from about \$250 to \$1,000, and were distributed equally to all teachers in the "winning" schools. Lavy found a positive effect on participating schools relative to a nonparticipating comparison group of low-performing schools. He also concluded that endowing schools with additional resources (that is, 25 percent of school awards had to go to capital improvements) contributed to increased student performance.

The second study examined an individual teacher bonus program, also run as a tournament (Lavy, 2004). Essentially, teacher participants were ranked on the basis of value-added contributions to student achievement on a variety of exit exams, and bonuses were given to top performing teachers. The program included 629 teachers, of whom 302 won awards. The bonuses were substantial, as large as \$7,500 per class on an average base pay of \$25,000. Results indicated a positive effect in that the performance of participating teachers (that is, both bonus recipients and non-recipients) rose, relative to a comparison group of teachers who did not participate in the incentive program.

Lavy (2004) also investigated whether the program exhibited the type of negative spillover consequences often discussed in the "contracting" literature. First, apropos the multitasking problem, test scores in other non-tournament subjects did not fall. In addition, and consistent with the teacher value-added literature discussed above, teacher characteristics such as experience or certification could not predict the winners. Another interesting feature of this study is that Lavy compared the cost effectiveness of the individual bonus scheme with that of group bonuses or another program providing additional educational resources, aside from pay, to

<sup>19</sup> The researchers note that malaria and AIDS are serious problems in these villages, which tend to raise employee absenteeism in the workforce as a whole. The authors present anecdotal evidence suggesting that it may be higher for teachers than other professional workers, however.

<sup>20</sup> Tournaments award prizes not on the basis of an absolute standard but on the basis of relative performance.

traditionally low achieving schools. He found that the cost per unit gain in the individual teacher incentive program dominated that in the group incentive or added resource programs.

The studies considered thus far evaluated specific incentive intervention. Figlio and Kenny (2007) take a different tack and analyze data from a national sample of U.S. K–12 schools in an attempt to estimate the effect of merit pay by comparing the academic performance of schools with various types of incentive programs to those without. Merging data from the National Educational Longitudinal Survey of 1988, their own survey on merit pay, and the 1993–94 Schools and Staffing Surveys, they examine the natural variation in the use of incentive-based pay among both public and private schools.<sup>21</sup> Variation in incentive programs enabled construction of a school-level measure of the strength of the teacher incentive “dosage,” reflecting not only the existence of a merit-based pay scheme but also its pecuniary consequences. Figlio and Kenny concluded that the effects of even modest doses of incentive pay are statistically significant in both public and private schools, as well as the effect of a high level of implementation of incentives relative to no incentive program. In substantive terms, a vigorous merit pay program’s impact is comparable to a one standard deviation decrease in days absent for the average student, and an increase in maternal education of three years.

While the authors creatively linked multiple national data systems with their Survey of School Teacher Personnel Practice, there are methodological concerns that warrant mention. First, there was an eight-year lag between student test scores reported in NELS and the Figlio and Kenny survey, thus making sample attrition a significant concern. If differential sample attrition took place, this makes it difficult to interpret the reason for differences in test scores between the treatment and comparison conditions. Second, while the authors were able to increase the number of schools satisfactorily responding to their survey by matching within district responses across two or more schools, the response rate was still very low (approximately 40 percent). Finally, there are challenges in assuring that the merit pay programs were in place at the time of the NELS testing. In spite of these measurement problems, which might be expected to bias their estimates of the treatment effect toward zero (errors in measurement of the treatment variable), Figlio and Kenny add crucial insight into the relationship between individual teacher performance incentives and student achievement.

Winters, Ritter, Barnett, and Green (2007) is a small-scale, but rigorous, evaluation of the first two schools participating in Little Rock, Arkansas’ Achievement Challenge Pilot Project (ACPP). Their evaluation examines the effect of ACPP on student proficiency in math compared to three other elementary schools with similar demographic and baseline achievement characteristics. ACPP ties performance bonuses to individual student fall-to-spring gains on a standardized student achievement test, ranging from \$50 per student (0–4 percent gain) up to \$400 per student (15 percent gain). In practice this yielded bonus payouts ranging from \$1,200 up to \$9,200 per teacher per year.

An attractive feature of the study is that the student gain score outcomes are estimated with a different assessment from that used to determine the bonuses (that is, the students took two different standardized spring assessments). Use of

<sup>21</sup> Clearly, using natural variation has its cost, since the variation may not arise exogenously. However, one benefit of a study using natural variation is that many of the schools using the incentive plans may have had them in place for a sufficient length of time to pick up both motivation and selection effects.

an alternative test reduces the potential bias caused by teachers narrowly “teaching to the test” used for the bonus payout. The preferred student fixed-effect estimates of Winters et al. find a statistically significant 4.6 Normal Curve Equivalent (NCE) math gain for every year a student spent in an ACPD school. The ACPD bonus system, unlike many of the studies considered in this review, remains in place and has since expanded to five elementary schools during the 2006–07 school year.

Some U.S. and foreign pay-for-performance experiments have been implemented in a way as to not permit rigorous program evaluation. Ladd (1999) and Clotfelter and Ladd (1996) examined the effect of a school-wide incentive scheme implemented in the Dallas Independent School District (DISD) in the mid-1990s. The Dallas Accountability and Incentive Program provided a modest pay boost to all teachers in high-performing schools. Since the program was intended to raise the performance of all schools in the district, the district was the treatment unit in Ladd’s and in Clotfelter and Ladd’s analyses. The authors found that achievement in DISD rose relative to other Texas public school districts. This is suggestive evidence, and probably the best one can do in an assessment of district-wide programs. However, it must be recognized that other factors may have been changing over time in both the Dallas and comparison districts in ways that confound the true effect of the Dallas Accountability and Incentive Program.

Atkinson et al. (2004) evaluated the effect of an ongoing teacher bonus pay scheme in the United Kingdom. Prior to introduction of the U.K.’s performance-based management plan, all teachers in the United Kingdom were compensated on their unified wage scale, which was predicated upon qualifications and experience. The performance-based management scheme changed pay practices such that teachers who applied and were considered performance-award eligible were given an immediate pay increase of up to £2,000 and access to a higher pay level on the traditional unified wage scale. Ultimately, the pay system facilitated teachers to earn up to £30,000 without taking on additional management responsibilities or having to leave the classroom.

While it turned out *ex post* that the bonus was provided to nearly 97 percent of teachers who submitted an application to the United Kingdom’s Education Ministry, Atkinson and colleagues find that introduction of the payment scheme did improve test score gains, on average, by about half a grade per pupil relative to ineligible teachers. Equal to 73 percent of a standard deviation, and given that the program’s high-stakes assessments taken at age 16 are the key qualifications for entry into higher education, these researchers conclude findings are “not trivial.” It is important to note, however, that the authors had difficulty in developing a representative national sample with pre- and post-program gain-score data and, as a result, were forced to rely on a small sample of schools for which linked student-teacher data were available.

Finally, Eberts, Hollenbeck, and Stone (2002) studied the effect of an incentive scheme in a single alternative high school in Michigan. In response to a growing dropout rate problem, the school introduced a bonus system that paid teachers to raise their students’ course completion rates. The researchers compared the single “treatment” school to one other alternative high school considered comparable. The bonus program significantly raised course completion relative to this control school but, not surprisingly, relative values of nontargeted variables such as student pass rates or grade point average dropped because academically marginal students were induced to stay in school. Clearly, a better performance pay plan would have incorporated a larger set of performance indicators, including

student achievement and remediation mechanisms for students who traditionally opted out of formal education. However, the results of the Eberts et al. study show that teachers responded to a short-term incentive plan, and raised the course completion rate.

In conclusion, the evaluation literature on teacher incentive pay programs is small. Clearly, more studies are needed. However, the studies that have been conducted to date are generally positive and provide a strong case for further policy experimentation in this area by states and districts (combined with rigorous evaluation). In addition, even the “mixed” studies suggest that incentive programs change teacher behavior—“you get what you pay for.” Thus, education policymakers need to be careful in designing such programs, and must expect to continually refine the programs as they learn about behavioral responses. A review of the principal-agent multitasking literature outside of education by Courty and Marschke (2003) highlights the dynamic learning context of these incentive systems. Their model illuminates the importance of experimentation and trial and error in schools’ efforts to develop a reliable performance measure. In this respect, these two “mixed” studies, and Courty and Marschke’s review, point to the need for experimentation and careful evaluation and the willingness for successive iterations of improvement.

### Incentive Pay in Private and Charter Schools

If contracting problems in K–12 public education, such as performance monitoring, multitasking, and team production, are inherent in the production process, and sufficiently severe so as to preclude group or individual performance-related pay programs, then we would expect to see similar pay structures in charter schools and private schools when compared to traditional public schools. Or, we would at least expect to see negative attitudes to various incentive pay programs that may be present in public charter or private schools. Several studies have examined these questions and found significant differences between the two sectors as well as self-report teacher data refuting the age-old supposition that educators do not support pay tied to performance.

Hoxby (2002) hypothesizes that parental freedom to choose schools leads to greater use of merit pay and performance-related pay in public charter schools and private schools. Data from the 1990–91 and 1993–94 Schools and Staffing Survey’s samples of public and private school teachers and administrators are matched with scores from the SAT reasoning test, competitive college ranking definitions in *Barron’s Profile of American Colleges*, and the Common Core of Data. Hoxby estimates an earnings equation to uncover how choice affects a school’s willingness to pay for each unit of a teacher characteristic (for example, credentials, math skills, etc.), demonstrating that choice creates an incentive environment within the profession requiring teachers to have higher levels of human capital and effort in return for higher marginal wages.

Ballou (2001) examines data from four national surveys of teachers and school administrators to compare pay for performance in public and private schools. He finds that there is a higher incidence of merit pay in “other religious” and “nonsecular” schools when compared to traditional public schools and Catholic schools. Ballou then estimates a teacher earnings equation to investigate the size of merit bonuses as a percent of base pay across sectors. His estimates indicate significant differences between sectors. Public school teacher merit pay increases averaged

about 2 percent of base pay, whereas private sector teacher merit pay bonuses were almost 10 percent of base pay. Ballou concludes that failure of merit pay is not necessarily due to the complexity of teachers' jobs and need for teamwork and cooperation in schools; rather, influential stakeholders such as teacher unions have played a key role in obstructing merit pay policy.

A more recent study by Podgursky (2007) examines data from the 1999–2000 Schools and Staffing Surveys. This paper examines reasons why personnel policy and wage setting differ between traditional public, private, and charter schools and the effects of these policies on academic measures of teacher quality. Survey and administrative data suggest that the regulatory freedom, small size of wage-setting units, and a competitive market environment make pay and personnel practices more market- and performance-based in private and charter schools as compared to traditional public schools (see Table 4). These practices, in turn, permit charter and private schools to recruit teachers with better academic credentials as compared to traditional public schools.

Finally, Ballou and Podgursky (1993) examine survey data on teacher attitudes from the 1987–88 Schools and Staffing Survey to investigate determinants of teacher attitudes toward merit pay. While conventional wisdom suggests that the majority of teachers oppose merit pay, these researchers find strong evidence that teachers in districts that use merit pay do not seem demoralized by the system or hostile toward it, and teachers of disadvantaged and low-achieving students are generally supportive of merit pay. Moreover, teachers' first-hand experiences with merit pay are not negative, even when the respondent did not receive a merit bonus. Ballou and Podgursky rightfully caution that findings are based on a very limited battery of questions, possibly making these findings sensitive to the wording

**Table 4.** Teacher salary schedules and teacher incentive pay in traditional public, charter, and private schools. (standard error in parentheses)

	Traditional Public (%)	Charter (%)	Private (%)	Nonreligious Regular School (%)
Is there a salary schedule for teachers in this school?	96.3 (0.29)	62.2 (0.72)	65.9 (1.24)	45.1 (5.60)
Does this school currently use pay incentives such as cash bonuses, salary increases, or different steps on the salary schedule to reward:				
NBPTS certification?	8.3 (0.37)	11.0 (0.43)	9.6 (0.88)	14.8 (5.5)
Excellence in teaching?	5.5 (0.35)	35.7 (0.65)	21.5 (0.93)	42.9 (5.5)
Completion of in-service professional development?	26.4 (0.70)	20.5 (0.56)	18.7 (0.88)	26.0 (5.67)
Recruit or retain teachers in fields of shortage?	10.4 (0.464)	14.9 (0.54)	7.9 (0.61)	15.0 (3.40)

Source: 1999–2000 Schools and Staffing Surveys, reported in Podgursky (2007).

or context of the questions. Regardless, it is clear that private school teachers are much more supportive of (and benefit from) performance-related pay than public school teachers, which is consistent with the sorting effect hypothesized by Lazear (2003).

## CONCLUSION

In this paper we examine the economic case for performance-related pay in K–12 education. Our focus is on teachers, by far the largest group of employed professionals. However, many of the arguments generalize to school administrators as well. We began with a historical overview of teacher compensation policy from the 18th century to present and then moved to a description of notable district, state, national, and federal performance-related pay initiatives currently operating in the American K–12 public education system.

We also reviewed several ideas from the growing personnel economics literature that have particular relevance for teacher performance pay. There are some well-known problems in the use of performance-related pay programs in any organizational context; however, we are not persuaded that these are any more severe in K–12 education. One important theme, often ignored in education studies, is motivation versus selection effects in an incentive system. In the long run, a pay scheme tends to attract employees who prefer or prosper under it. The wide dispersion of teacher effectiveness found in large-scale value-added studies certainly suggests that substantial gains may be possible through sorting. A second important theme that emerges is the role of credentials versus performance in pay determination, which finds a private sector analogue in the design of base versus variable pay.

The evaluation literature on performance-related compensation schemes in education is very diverse in terms of incentive design, population, type of incentive (group versus individual), strength of study design, and duration of the incentive program. While the literature is not sufficiently robust to prescribe how systems should be designed—for example, optimal size of bonuses, mix of individual versus group incentives—it is sufficiently positive to suggest that further experiments and pilot programs by districts and states are very much in order. It is critical that these programs be introduced in a manner amenable to effective evaluation. Moreover, as noted by Courty and Marschke (2003), an overarching lesson seems to be that trial and error is likely required to formulate the right set of performance incentives. Development of massive student longitudinal achievement databases, such as those sponsored by USDoe, further opens prospects for rigorous value-added assessment over time.

We see two possible impediments to productive pilot studies or experiments. In our survey, we noted that the strongest findings to date arose from two experimental merit pay systems implemented in high schools in Israel (Lavy, 2002, 2004). Both of these systems were rank-order tournaments and involved substantial rewards for teachers. States and districts have been reluctant to implement rank-order tournaments in part due to strong union opposition. By their very nature, these are zero-sum games and many assert that such incentive schemes discourage teacher collaboration and cooperation, to the detriment of overall school performance. However, tournaments have the important benefit that the pool of “winnings” is capped. School districts have also been reluctant to implement incentive plans involving large bonuses for many of the same reasons.

This suggests an important role that foundations can play in advancing policy research in this area. Foundations routinely award prizes to teachers, often in substantial amounts. In fact, some of the more interesting U.S. experiments such as Little Rock, and a randomized experiment now under way in Metropolitan Nashville Public Schools system—both involving substantial teacher bonuses—have relied on private foundations to fund them.

In the last decade, public school districts have absorbed large numbers of new school teachers. As these teachers age and move down (experience) and across (education) salary schedules, school districts will find themselves devoting ever larger expenditures to schedule-driven pay increases that are unlikely to have any significant effect on student achievement. Private sector employers understand that strategic pay policies are a very important lever in raising firm performance and are thus continually refining and/or revamping their compensation systems. Even the Office of Personnel Management has taken major steps in implementation of performance-based pay in the federal system. School administrators need to channel some of these funds toward more strategic pay experiments designed to raise student achievement. Education policy makers should nurture, expand, and evaluate these local experiments.

*MICHAEL J. PODGURSKY is Professor of Economics at the Harry S. Truman School of Public Affairs and Economics Department, University of Missouri–Columbia.*

*MATTHEW G. SPRINGER is Research Assistant Professor of Public Policy and Education at Peabody College of Vanderbilt University and Director of the National Center on Performance Incentives.*

## ACKNOWLEDGMENTS

This research was supported by the National Center on Performance Incentives, which is funded by the United States Department of Education's Institute of Education Sciences (R305A06034). The authors wish to thank Samantha Dalton, Kelly Fork, and Brian McInnis for research assistance, and Robert Costrell, Jay Greene, James Guthrie, Jessica Lewis, Jeffrey Springer, and several anonymous referees for comments. The views expressed in this paper do not necessarily reflect those of sponsoring agencies or individuals acknowledged.

## REFERENCES

- Aaronson, D., Barrow, L., & Sanders, W. (2003). Teachers and student achievement in Chicago public high schools. Chicago: Federal Reserve Bank of Chicago.
- Adnett, N. (2003). Reforming teachers' pay: Incentive payments, collegiate ethos and UK policy. *Cambridge Journal of Economics*, 27, 145–157.
- Altonji, J. G., & Pierret, C. R. (1996). Employer learning and the signaling value of education. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.
- Angrist, J. D., & Guryan, J. (2005). Does teacher testing raise teacher quality? Evidence from state certification requirements. National Bureau for Economic Research, Working Paper.
- Armor, D., Conry-Oseguera, P., Fox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). Analysis of the school preferred reading program in selected Los Angeles minority schools. Santa Monica, CA: RAND Corporation.
- Atkinson, A., Burgess, S., Croxon, B., Gregg, P., Propper, C., Slater, H., & Wilson, D. (2004). Evaluating the impact of performance-related pay for teachers in England. University of Bristol: Centre for Market and Public Organization.

- Azordegan, J., Byrnett, P., Campbell, K., Greenman, J., & Coulter, T. (2005). Diversifying teacher compensation. Denver: Education Commission of the States
- Ballou, D. (2001). Pay for performance in public and private schools. *Economics of Education Review*, 20, 51–61.
- Ballou, D. (2005). Value-added assessment: Lessons from Tennessee. In Robert W. Lissitz (Ed.), *Value-added models in education: Theory and applications* (pp. 1–26). Maple Grove, MN: JAM Press.
- Ballou, D., & Podgursky, M. (1993). Teachers' attitudes toward merit pay: Examining conventional wisdom. *Industrial and Labor Relations Review*, 47, 50–61.
- Ballou, D., & Podgursky, M. (1997). Teacher pay and teacher quality. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Ballou, D., & Podgursky, M. (2001). Let the market decide. *Education Next*, 1, 1–7.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29, 37–66.
- Beer, M., & Cannon, M. D. (2004). Promise and peril in implementing pay-for-performance. *Human Resource Management*, 43, 3–48.
- Boyd, D., Grossman, P., Lankford, H., & Loeb, S. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, 1, 176–216.
- Clotfelter, C., & Ladd, H. (1996). Recognizing and rewarding success in public schools. In Ladd, H. (Ed.), *Holding schools accountable: performance-related reform in education* (pp. 23–64). Washington D.C.: The Brookings Institution,
- Clotfelter, C., Ladd, H. F., & Vigdor, J. (2007). How and why do teacher credentials matter for student achievement? Center for Analysis of Longitudinal Data in Education Research working paper.
- Community Training and Assistance Center (2004). Catalyst for change: Pay for performance in Denver final report. Boston, MA: Author.
- Cooper, S. T., & Cohn, E. (1997). Estimation of a frontier production function for the South Carolina educational process. *Economics of Education Review*, 16, 313–327.
- Courty, P., & Marschke, G. (2003). Dynamics of performance-measurement systems. *Oxford Review of Economic Policy*, 19, 268–284.
- Cullen, J. B., & Reback, R. (2006). Tinkering toward accolades: School gaming under a performance accountability system. NBER Working Paper #12286. Cambridge, MA: National Bureau for Economic Research.
- Darling-Hammond, L., & Barnett, B. (1988). The evolution of teacher policy. Santa Monica, CA: The RAND Corporation.
- Dee, T., & Keys, B. J. (2004). Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management*, 23, 471–488.
- Deere, D., & Strayer, W. (2001). Putting schools to the test: School accountability, incentives, and behavior. Texas A&M University, unpublished.
- Dixit, A. (2002). Incentives and organizations in the public sector. *Journal of Human Resources*, 37, 696–727.
- Eberts, R., Hollenbeck, K., & Stone, J. (2002). Teacher performance incentives and student outcomes. *Journal of Human Resources*, 37, 913–927.
- Figlio, D. (2003). Testing, crime and punishment. University of Florida, unpublished.
- Figlio, D., & Getzler, L. (2002). Accountability, ability and disability: Gaming the system? National Bureau for Economic Research Working Paper 9307. Cambridge: NBER.
- Figlio, D., & Kenny, L. (in press). Individual teacher incentives and student performance. *Journal of Public Economics*.

- Figlio, D., & Winicki, J. (2005). Food for thought? The effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89, 381–394.
- Gallagher, Alix H. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79, 79–107.
- Gaynor, M., Rebitzer, J., & Taylor, L. (2004). Physician incentives in health maintenance organizations. *Journal of Political Economy*, 112, 915–931.
- Glazerman, S., Silva, T., Addy, N., Avellar, S., Max, J., McKie, A., Natzke, B., Puma, M., Wolfe, P., Gresler, R. U. (2006). Options for studying teacher pay reform using natural experiments. Washington, DC: Mathematica Policy Research, Inc.
- Glewwe, P., Ilias, N., & Kremer, M. (2004). Teacher incentives. Mimeograph. Cambridge, MA: Harvard University.
- Goldhaber, D., & Anthony, E. (in press). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *Review of Economics and Statistics*.
- Goldhaber, D., & Brewer, D. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on education production. 32, 505–523.
- Goldhaber, D., Hyung, J., DeArmond, M., & Player, D. (2005). Why do so few public school districts use merit pay? University of Washington Working Paper.
- Goldin, C. (2003). The human capital century: Has U.S. leadership come to an end? *Education Next*, 73–78.
- Goodnough, A. (1999, December 8). Answers allegedly supplied in effort to raise test scores. *New York Times*.
- Guthrie, J. W., Springer, M. G., Rolle, A. R., & Houck, E. A. (2007). Modern education finance and policy. Mahwah, NJ: Allyn & Bacon.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Analysis Policy Archives*, 8. <http://epaa.asu.edu/epaa/v8n41/>.
- Hannaway, J. (1992). Higher order skills, job design, and incentives: An analysis and proposal. *American Educational Research Journal*, 29, 3–21.
- Hanushek, E. A. (2003). The failure of input-based resource policies. *Economic Journal*, 113, F64–F68.
- Hanushek, E. A., & Rivkin, S. G. (2004). How to improve the supply of high-quality teachers. In D. Ravitch (Ed.), *Brookings papers on education policy: 2004* (pp. 7–25). Washington, DC: Brookings Institution Press.
- Harris, D. N., & Sass, T. R. (2007). The effects of NBPTS-certified teachers on student achievement. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. Urban Institute. Working Paper No. 4.
- Hassel, B. C. (2002). Better pay for better teaching: Making teacher pay pay off in the age of accountability. PPI Policy Report. Washington, DC: Progressive Policy Institute.
- Hatry, H., Greiner, J., & Ashford, B. (1994). Issues and case studies in teacher incentive plans. Washington, DC: Urban Institute Press.
- Hein, K. (1996). Raises fail, but incentives save the day. *Incentive*, 170, 11.
- Heneman, R., & Ledford, G. E. (1998). Competency pay for professionals and managers in business: A review and implications for teachers. *Journal of Personnel Evaluation in Education*, 12, 103–21.
- Holmstrom, B., & Milgrom, P. (1991). Multitask principal agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organizations*, 7, 24–52.
- Hoxby, C. M. (2002). Would school choice change the teaching profession? *Journal of Human Resources*, 37, 846–891.

- Hoxby, C. M., & Leigh, A. (2004). Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States. *American Economic Review*, 93, 236–240.
- Jacob, B. (2005). Testing, accountability, and incentives: The impact of high-stakes testing in Chicago Public Schools. *Journal of Public Economics*, 89, 5–6.
- Jacob, B., & Lefgren, L. (2005). Principals as agents: Subjective performance measurement in education. National Bureau for Economic Research Working Paper No. 11463. Cambridge, MA: NBER.
- Jacob, B., & Levitt, S. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118, 843–877
- Kandel, E., & Lazear, E. (1992). Peer pressure and partnerships. *Journal of Political Economy*, 100, 801–817.
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006). What does certification tell us about teacher effectiveness? Evidence from New York City. National Bureau for Economic Research Working Paper 12155. Cambridge, MA: NBER.
- Kerchner, C. T., Koppich, J. E., & Weeres, J. G. (2003). *United mind workers: Unions and teaching in the knowledge society*. San Francisco, CA: Jossey-Bass.
- Kershaw, J. A., & McKean, R. N. (1962). Teacher shortages and salary schedules. A RAND Research Memorandum. Santa Monica, CA: RAND Corporation.
- Kimball, S. M., White, B., Milanowski, A., & Borman, G. (2004) Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79, 54–78.
- Koedel, C. (2007). teacher quality and educational production in secondary school. Department of Economics, University of Missouri–Columbia, Working Paper.
- Koedel, C., & Betts, J. (2005). Re-examining the role of teacher quality in the education production function. University of California–San Diego, Working Paper.
- Koretz, D., Barron, S., Mitchell, K., & Stecher, B.M. (1996). Perceived Effects of the Kentucky Instructional Results Information System (KIRIS). Santa Monica, CA: RAND Corporation.
- Ladd, H. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review*, 18, 1–16.
- Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy*, 110, 1286–1317.
- Lavy, V. (2004). Performance pay and teachers' effort, productivity and grading ethics. National Bureau for Economic Research, Working Paper 10622. Cambridge, MA: NBER.
- Lazear, E. (2000). Performance pay and productivity. *American Economic Review*, 90, 1346–1361.
- Lazear, E. (2003). Teacher incentives. *Swedish Economic Policy Review*, 10, 197–213.
- Marsden, D., & Belfield, R. (2006). Pay for performance where output is hard to measure: The case of performance pay for school teachers. In B. E. Kauffman & D. Lewin (Eds.), *Advances in industrial and labor relations*. London: JAI Press.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). Evaluating value added models for teacher accountability, MG-158-EDU. Santa Monica, CA: RAND Corporation.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. S. (2004). Models for value-added modeling of teacher effects. *Journal of Education and Behavioral Statistics*, 29, 67–101.
- Milanowski, A. (2004) The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*. 79, 33–53.
- Milken Family Foundation (2005). *Understanding the teacher advancement program*. Santa Monica, CA: Milken Family Foundation.

- Milkovich, G. T., & Newman, J. M. (2005). *Compensation*. 8/E. New Delhi: Tata McGraw-Hill Publishing, Ltd.
- Moehlman, A. B. (1927). *Public school finance*. New York: Rand McNally & Company.
- Muralidaran, K., & Sundararaman, V. (2006). *Teacher incentives in developing countries: Experimental evidence from India*. Department of Economics, Harvard University. (November).
- Murnane, R. J. (1975). *The impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger Press.
- Murnane, R. J., & Cohen, D. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and few survive. *Harvard Education Review*, 56, 1–17.
- Murnane, R. J., & Olsen, R. J. (1990). The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina. *Journal of Human Resources*, 25, 106–124.
- National Commission on Excellence in Education. (1983). *A Nation at risk*. Washington, DC: U.S. Department of Education. <http://www.ed.gov/pubs/NatAtRisk/index.html>.
- Odden, A., & Kelley, C. (1996). *Paying teachers for what they know and do: New and smarter compensation strategies to improve schools*. Thousand Oaks, CA: Corwin Press.
- Peabody, Z., & Markley, M. (2003, June 14). State may lower HISD rating; almost 3,000 dropouts miscounted, report says. *Houston Chronicle*, p. A1.
- Podgursky, M. (2007). Teams versus bureaucracies: Personnel policy, wage-setting, and teacher quality in traditional public, charter, and private schools. In M. Berends, M. G. Springer, & H. Walberg (Eds.), *Charter school outcomes*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Podgursky, M., Monroe, R., & Watson, D. (2004). The academic quality of public school teachers: An analysis of entry and exit behavior. *Economics of Education Review*, 23, 507–518.
- Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37, 7–63.
- Protsik, J. (1995). *History of teacher pay and incentive reform*. Washington: Educational Resources Information Center.
- Rivkin, S., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73, 417–458.
- Sanders, W. J., Ashton, J. J., & Wright, S. P. (2005). *Comparison of the effects of NBPTS-certified teacher with other teachers on the rate of student achievement progress*. Cary, NC: SAS Institute Inc.
- Sanders, W. J., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System: Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299–311.
- Sclafani, S., & Tucker, M. S. (2006). *Teacher and principal compensation: An international review*. Washington, DC: Center for American Progress.
- Sharpes, D. K. (1987). Incentive pay and the promotion of teaching proficiencies. *The Clearinghouse*, 60, 407–410.
- Stucker, J.P., & Hall, G.R. (1971). *The performance contracting concept in education*. Santa Monica, CA: RAND Corporation.
- Tyack, D. B. (1975). *The one best system: A history of American urban education*. Cambridge, MA: Harvard University Press.

- Vyrastekova, J., Onderstal, S., & Konig, P. (2006). Team incentives in public organizations: An experimental study. CPB Discussion Paper No. 60. The Hague: Netherlands Bureau for Economic Policy Analysis.
- Winters, M., Ritter, G., Barnett, J., & Greene, J. (2006). An evaluation of teacher performance pay in Arkansas. University of Arkansas. Department of Education Reform.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement. *Journal of Personnel Evaluation in Education*, 11, 57–67.