Using Performance-Based Pay to Improve the Quality of Teachers

Victor Lavy

Summary

Tying teachers' pay to their classroom performance should, says Victor Lavy, improve the current educational system both by clarifying teaching goals and by attracting and retaining the most productive teachers. But implementing pay for performance poses many practical challenges, because measuring individual teachers' performance is difficult.

Lavy reviews evidence on individual and school-based incentive programs implemented in recent years both in the United States and abroad. Lavy himself evaluated two carefully designed programs in Israel and found significant gains in student and teacher performance. He observes that research evidence suggests, although not conclusively, that pay-for-performance incentives can improve teachers' performance, although they can also lead to unintended and undesired consequences, such as teachers' directing their efforts exclusively to rewarded activities.

Lavy also offers general guidelines for designing effective programs. He emphasizes that the system must measure true performance in a way that minimizes random variation as well as undesired and unintended consequences. It must align performance with ultimate outcomes and must be monitored closely to discourage gaming if not outright fraud in measured output. Goals should be attainable. Incentives should balance individual rewards with school incentives, fostering a cooperative culture but not at the expense of free riding. All teachers should be eligible for the incentive offered, but only a subset of teachers should be rewarded in practice. If too many teachers are rewarded, teachers may not need to exert much extra effort to benefit.

Many of the practical challenges faced by performance-related pay, Lavy says, can be addressed through careful design of the system. He emphasizes that setting up a performance-related pay system that works is not a one-time task. Even with the best preparation, initial implementation is likely to be problematic. But if the effort is seen as ongoing, it should be possible to make progress gradually in developing incentives that motivate the desired teaching behaviors and that will be perceived by teachers as fair and accurate.

www.futureofchildren.org

Victor Lavy is professor of economics at the Hebrew University of Jerusalem and a research associate at the National Bureau of Economc Research and the Centre for Economic Policy Research.

Interest in improving public education is growing not only in the United States but worldwide. One reason for the heightened public attention is the key role played by education in determining both individual earnings and broader economic growth. Another is widespread dissatisfaction with the education sector's performance of late: substantial increases in spending on public schools have failed to bring corresponding increases in student achievement.[1]

The quest to improve public education has led policymakers and researchers to focus on how to increase teachers' effectiveness. One obvious means is compensation. According to many observers, the traditional basis for teacher pay—years of service and education—provides little incentive for excellence. To make teachers more effective, these critics argue, pay should be tied to performance. And some school districts, here and abroad, are undertaking reforms to test those ideas. In November 2005, for example, Denver voters approved a $25 million tax increase to fund a form of "merit pay" to reward elementary and secondary school teachers along a variety of dimensions, including their own demonstrated knowledge and skills and student academic growth. Whether Denver's new merit pay system will improve student achievement remains uncertain; an earlier pilot study in Denver found mixed results.[2]

In this article I examine academic and policy analysis of performance-based reward programs for primary and secondary school teachers. I stress, in particular, several questions. What are the pros and cons of implementing teachers' pay incentives in schools? What criteria are to be applied in designing optimal teacher incentives? How much is performance affected by incentives offered to practicing teachers? How will incentives affect the composition of applicants to teacher-training institutions and to teaching positions in the schools? What policy measures can remedy existing distortions in teachers' compensation? My intent is not to review exhaustively what is known about performance-based pay in education, but rather to summarize selected key findings, highlight some guidelines for designing effective teacher incentives schemes, and identify areas requiring additional evidence. My objectives are to present the theoretical benefits of performance-based pay as well as some of the practical obstacles to its effective implementation, to review critically the empirical evidence, and to draw policy conclusions.

Teachers' Compensation

Pay for performance is meant to solve the twofold problem of motivating high teacher performance while attracting and retaining good teachers under conditions where their effort or ability is not readily measured or observed. In the teaching profession, earnings are based primarily on input (that is, skills and time worked), rather than on output. Such a basis, critics say, is not "results-oriented." Moving to an earnings structure that ties pay—at least partially—to some performance indicators should thus improve the current system. In theory, the idea makes good sense. But implementing pay for performance poses many practical challenges. In the teaching profession effort and output are difficult to define and measure because the work is generally complex, unique, and often results from team efforts, with any one teacher's effort difficult to disentangle from that of the others on the team. One key goal of education is to give students the skills needed to ensure a productive career and sustain their economic well-being. Yet, because it takes years for the adult earnings of a

student to materialize, it is impossible to tie a teacher's wages to his students' earnings. Recent studies do in fact suggest that students' test scores are strongly correlected with their future earnings, but using test scores to measure a teacher's performance presents practical problems.[3] Peer or principal evaluations are yet another way to measure performance, given the drawbacks of testing and the many teachers for whom testing would not apply. But these approaches also have drawbacks.

Performance-Based Pay

Pay based on performance usually involves some objective assessment of schools' or teachers' efforts or success or some measure of their students' performance. Performance-based pay schemes have many variable features. They can compensate teachers only for their own performance or they can be structured as a team incentive program, with group performance determining the total incentive payment, which is then divided among team members regardless of individual performance. The group can include all of the school's teachers or a subgroup, such as the teachers of a given grade or a specific subject. Performance-based pay schemes can, but need not, involve sanctions for below-threshold performance. Although monetary rewards are the most common incentive in performance-related pay, other incentives can include reduced teaching load, promotion, and public recognition of outstanding teachers. The reward can be just a one-time event or it can be ongoing, leading to a permanent salary increase. It can be based on a relative criterion (for example, the average test score gain of a teacher's class relative to the classes of other teachers) or on an absolute criterion (such as the class average test score being higher than a predetermined threshold). The reward may be a fixed sum that is equal for all winners, or it can vary and increase with the winner's level of achievement. The total amount of awards may be predetermined (for example, only a certain number of teachers can win an award) or it may be open. The performance criteria can include outcomes for the teachers themselves, such as measures of absenteeism or performance on a test. They can also include measures of the teacher's students' performance, such as attendance, grade retention, dropout rates, or performance on tests. These criteria are not mutually exclusive.

Pay based on performance usually involves some objective assessment of schools' or teachers' efforts or success or some measure of their students' performance.

The target set for determining award winners is critically important both for efficiency and for equity. For example, if schools are ranked according to how many students attain a certain level of literacy as determined by an examination, gains near the cutoff are most rewarded. But if schools or teachers are rewarded on the basis of average test scores or changes in those averages, then credit is given for gains at all parts of the achievement distribution, not just those close to the cutoff point. For example, students in the bottom 5 percent of the achievement distribution might be too far away from the literacy standard to pass the test after one or two years, but raising their test scores might be worthwhile nonetheless. A hybrid measure could

target both some cutoff as well as some average. It is also possible to design a scheme that differentially weights improved test scores at different points of the achievement distribution, with the bottom weighted more than the top, the middle more than either extreme, and so forth. Finally, an important, but potentially underappreciated, part of a pay-for-performance plan is the identity of those who evaluate the teachers. The evaluators can be external to the school or can be peers, principals, or district supervisors.

Despite the almost innumerable combinations offered by these options, three prototypes of performance-based reward programs are most often implemented in education systems and are commonly examined by researchers. The first model, *merit pay*, generally involves individual incentives based on student performance. The second, *knowledge- and skill-based compensation*, generally involves individual incentives based on teacher skills. Knowledge- and skill-based pay differs from merit pay because it provides clear guidelines on what is being evaluated. The third model, *school-based compensation*, generally involves schoolwide incentives, typically based on student performance.

Potential Benefits of Performance-Based Pay

Performance-based pay in education brings with it many potential benefits but also many challenges. This section and the next present the main issues.

Productivity and Efficiency Considerations. In theory, pay based on output has two advantages over input-based pay in terms of efficiency (that is, producing "more" education for the same cost). The first, most frequently noted efficiency advantage has to do with incentives. Rewarding teachers or schools on the basis of an agreed metric aligns incentives directed at teachers or schools with those directed at students and potentially the entire society. If wages are based on student performance, for example, they provide teachers or schools with powerful signals about what is valued and what is not. Absent such signals, even well-meaning teachers may emphasize material that is obsolete or generally no longer valued by parents or the labor market. Similarly, if wages are based not only on the individual benefits of schooling to students (social scientists call these "private returns") but on the benefits to society as a whole ("social returns"), teachers or schools would take into account the social returns to education when making choices about their work. A student, for example, may want to drop out before completing high school because he feels that the costs of staying in school outweigh his individual benefits. A teacher considering only those individual benefits may not work as hard to discourage him from dropping out as would a teacher considering the costs and benefits to society as a whole.

Individual performance-based pay schemes improve efficiency by helping correct distortions in a teacher's effort that might result from gaps between her preferences and those of her students. For example, a teacher might fail to assign homework even though she knows its value for her students because correcting and grading assignments involves more work for herself. Individual performance-based pay provides some incentive for the teacher to do the "right thing."

The second efficiency advantage of output-based pay, mainly relevant for the merit pay model, involves sorting and selection. Assuming that the compensation system accurately identifies productivity, basing pay on per-

formance will attract and retain the most productive teachers. Even if teachers are unable to alter their own behavior to enhance performance, as measured, say, by students' test scores, some people are still inherently better than others at affecting test scores. Basing pay on output also tends to discourage teachers who cannot enhance their students' performance from remaining in the profession. A related point is that output-based pay will create a market for teaching quality that will help teachers move to schools where their talent is most highly valued. Equalization between productivity and wages will result, with poorly performing teachers receiving reduced wages and lower probabilities of promotion, and more capable teachers commanding better options. Finally, if teachers are able to improve their classroom performance, linking compensation to performance will provide all teachers incentives to improve through professional development, which will therefore induce still further productivity gains.[4]

Performance-related pay based on individual or schoolwide schemes could also improve school productivity by inducing better governance. For one thing, it requires school principals to monitor closely the quality of their teachers' work.[5] It is also assumed to bring about more coherent and common teacher-management goals in addition to an improved flow of information and feedback among all school agents. This result is assumed to flow from a common interest in improved outcomes.[6]

Other Potential Benefits. Critics of traditional pay schemes that reward experience and formal qualifications instead of performance argue that these schemes are unfair to highly motivated, effective, and efficient teachers whose extra efforts are not rewarded.[7] Performance-based pay can thus make compensation systems more equitable.

Finally, performance-based pay may increase support for public education from politicians and members of the public who are convinced that the reform will reverse the education sector's poor reputation and perceived inefficient use of resources.[8]

Potential Drawbacks to Performance-Based Pay

Despite its theoretical benefits, performance-based pay offers many practical challenges.

Measurement Problems. Performance measurement poses two separate problems for performance-based pay. Incentive systems assume that everyone can agree on goals; they also assume that it is possible to measure accurately progress toward these goals. Agreeing on goals is particularly difficult in education because competition between public schools is rare. In the private sector, market mechanisms discipline firms into providing products that consumers value, but public schools lack market discipline. Schooling is compulsory and public, and students are simply assigned to attend their neighborhood school. Parents and students who are unhappy with what their schools offer generally have no alternative except to attend a private school or move to another neighborhood or city—alternatives that are too costly for many.[9]

The other measurement issue represents the most common claim made against performance-based pay: evaluating progress toward the goal fairly and accurately is problematic. This is especially so when evaluation is based on proxies (as it often is), such as self-reported effort and motivation.[10] Identifying precisely what one teacher contributes to a student's performance and separating his

contribution from those of other teachers, the school, the principal, and the family is extremely difficult.[11] Compounding the problem is the fact that students are often deliberately assigned to specific teachers—that is, the assignment of teachers to classes is not random. A still further complication is how to identify the contributions of previous teachers, who may have been superior or inferior.

Pay based on reading and math test scores, for example, might encourage teachers to favor those subjects at the expense of, say, music and art or values and civic responsibility.

Negative Effects on Motivation and Collegiality. Another concern is that implementing individual-based incentives may create unfair competition between teachers, especially in the absence of transparent criteria, thus undermining collaboration. Even if evaluation is accurate and fair, teachers may still feel aggrieved if their competence is questioned. Evaluation may also create new hierarchies by giving administrators an additional source of power over teachers and the curriculum. Individual incentives could also undermine principal-teacher relationships because of the asymmetry in how each party views teacher evaluation: teachers use it to determine how they are performing and how they can improve, while principals use it to measure teachers' contribution to the school.[12]

Unintended Consequences. Some analysts caution that performance-based pay may have unintended consequences. Teachers, for example, may focus on the easiest way to increase the rewarded measure while ignoring measures that schools and parents ultimately want to improve.[13] Similarly, when one dimension of output is easily measured but another is not, teachers may dedicate their efforts to maximizing the measurable at the expense of the unmeasured dimension. Collectively, such efforts could even begin to constrict a school's curriculum to measurable subjects.[14] A further risk is that because test scores measure only certain skills, linking compensation to test scores might cause teachers to sacrifice the nurturing of curiosity and creative thinking to teaching the skills tested on standardized exams—a practice known as teaching to the test.[15] Pay based on reading and math test scores, for example, might encourage teachers to favor those subjects at the expense of, say, music and art or values and civic responsibility.[16] A teaching-to-the-test mentality is thus assumed to support the creation of a system where a narrow curriculum necessarily restricts student achievement in domains not tested.

Unintended consequences may also arise if teachers "game play" and develop responses that generate rewards contradicting the profession's spirit.[17] In other words, measuring student output may stimulate teachers to participate in inappropriate or deviant behavior, such as cheating. Using data from Chicago's public schools, Brian Jacob and Steve Levitt detected cheating in approximately 4 to 5 percent of the classes in their sample.[18] They also found that cheating responds swiftly to changes in teacher incentives. After standardized tests took on increased salience in Chicago's public schools in 1996, the prevalence of cheating rose sharply in low-achieving classrooms, but not in classes with average or above-average students. The

prevalence of cheating also appeared to be systematically lower where the costs of cheating were higher or the benefits of cheating lower, as in classrooms where a large number of students' test scores were excluded from official calculations because they were bilingual. Other studies of unintended consequences include altering school lunch menus during testing periods in an apparent attempt to artificially increase student test scores and manipulating who takes the test.[19]

Providing financial incentives to improve performance may be counterproductive in other ways as well. First, it may demoralize teachers, resulting in reduced effort. In laboratory experiments, one study found that workers in high-powered incentive systems may become unmotivated and thus work less than they would under a flat wage regime.[20] Second, financial incentives may undermine intrinsic motivation, that is, the sense of duty or satisfaction that motivates coming to work.[21] This threat is particularly real for teachers, who, as a group, exhibit strong intrinsic motivation flowing from the value they place on interacting with children and seeing them succeed.[22]

Another potential distortion is that teachers may focus disproportionate attention on those students who are most likely to improve their test scores or to cross a designated threshold.[23] The highest- and lowest-performing students may consequently be neglected because they do not promise adequate returns on investments of teachers' quality time.

Risks Posed to Teachers Could Increase Costs. The risks posed to teachers by performance-based pay could lead them to demand high compensation, which could in turn raise the cost of education.[24] Unlike relatively risk-free input-based payment, performance-based pay exposes employees to earnings variability beyond their control. If teachers, like other workers, are risk averse, inducing them to accept a risky compensation package will entail higher average pay overall.

Teachers Are Motivated by Nonfinancial Incentives. A frequent criticism of performance-based pay is that teachers, as professionals relatively immune to motivation by pecuniary rewards, will not respond to financial incentives. Monetary rewards could thus simply inflame resentment toward management and decrease employee loyalty, both of which could reduce productivity. One study suggests that nonmonetary rewards, such as additional holidays, may be better motivators.[25]

Union and Teachers' Opposition. Teacher unions worldwide strongly oppose performance-based pay.[26] Unions view wage differentiation on the basis of subject taught, as well as any sort of subjective evaluation of teachers, as threats to their collective bargaining strategies and therefore reject them outright. And union views weigh heavily: lobbying by unions has often halted efforts to legislate performance-based rewards.[27] Union objections appear to reflect opposition voiced by teachers directly.[28] Teachers see performance-based pay, supported by unfair evaluation, as a threat to their autonomy. Sanctions against poorly performing schools, which are included in some performance-based schemes, are another major source of union and teacher opposition.[29]

Disappointing Experience with Past Merit Pay Programs. The repeated failures of poorly designed and implemented merit pay programs over the past two decades have undermined the credibility of new and better-

designed initiatives. A key weakness in past programs has been opaque goals, which make it hard for teachers to understand the program and undermine their support for it. Opaque goals also make it difficult for administrators to explain why some staff members receive a bonus and others do not. One study finds that even in established programs such as those implemented in Kentucky and North Carolina, many participants remain skeptical that bonuses go to qualified teachers.[30]

The High Cost of Performance-Based Pay Schemes. Finally, implementing performance-based pay is easier in small organizations, such as private schools, than in large public school systems with sizable teaching staffs. System size therefore impinges on the observed high cost of performance-related pay, making the program infeasible. One study argues that adequate evaluation of every teacher, expensive in itself, would require considerable resources if performed regularly.[31] The time alone required to administer a pay-for-performance system would have severe budgetary implications.[32] Moreover, as a research study points out, improved productivity in the private sector can generate added income to help mitigate budget problems, but enhanced productivity in public schools has no such effect.[33]

Overcoming Some of the Obstacles

Several of the many potential obstacles to implementing an effective performance-related pay system can be addressed. For example, one solution to the measurement problem is to compensate teachers on the basis of principal evaluations. Brian Jacob and Lars Lefgren compared subjective principal assessments with measures of teacher effectiveness based on gains in student achievement on standardized tests—measures often known as teacher value added—and found that principals are quite good at identifying those teachers who produce the largest and the smallest standardized achievement gains in their schools, but far less able to distinguish between teachers in the middle of the distribution.[34] They also found that principals systematically discriminate against male and untenured faculty.[35]

A principal-based assessment system would likely result in higher student achievement than today's input-based compensation system. But Jacob and Lefgren cite an important limitation of their research. First, the principals whom they examined were not themselves evaluated explicitly on the basis of their ability to identify effective teachers. Moving to a system where principals have more authority and responsibility for monitoring teacher effectiveness might enhance a principal's capacity to identify the required characteristics. But principals may be less willing to assess teachers honestly under such a system, perhaps in response to social or political pressures. Further, the inability of principals to distinguish between teachers in a broad middle range of quality suggests caution in relying on principals for the finely tuned performance determinations that might be required under certain merit pay policies.

In response to the concern that merit pay models may hamper collaboration, one could structure the system to reward teacher cooperation, especially through group-based pay.[36] This strategy can foster both teacher interdependence and acknowledgement of that interdependence. That said, team-based incentive systems raise the risk of "free riding."[37] If each teacher's share of the team reward is small relative to the cost of effort and if effort is difficult to observe, every teacher in the team will have an incentive to shirk and free ride on the efforts of others.[38] One

way to avoid this problem is to encourage peer pressure and mutual monitoring within the team.[39]

Fears that teachers would need to be highly compensated for the increased risk in a performance-based compensation system are probably overstated. If teachers are paid on the basis of student performance, and if the number of students whom teachers teach each year is high, the year-to-year variation in average class test scores is likely to be small. Furthermore, even under the most ambitious schemes only a fraction of teachers' wages would be tied to performance, thus making compensation based on incentives only a marginal component of pay. It is thus unlikely that earnings will fluctuate by more than a few percent annually around some basic trend.

The idea that teachers themselves—as reflected in the positions of their unions—oppose performance-based rewards may also be overstated. One study found that most teachers favor additional pay for additional duties per se and as a component of a career ladder where performance dictates the speed of advancement.[40] Unsurprisingly, performance-based rewards are more popular when they supplement, rather than replace, other forms of salary.

The same study found that the pay level in a school district appeared to have no influence on teachers' attitudes toward merit pay, although teachers who were paid low salaries and who belonged to ethnic minorities were more likely than others to support the program. Attitudes toward merit pay were independent of the number of students who were eligible for free lunches, suggesting that students' socioeconomic status did not affect teachers' views on merit pay. Interestingly, private school teachers viewed performance-related pay more favorably than their public school counterparts.[41] Teachers' attitudes thus appear relatively malleable and to depend on program design.

Teachers' and union objections were overcome in some cases when specific interest groups and legislators supported perform-

Fears that teachers would need to be highly compensated for the increased risk in a performance-based compensation system are probably overstated.

ance-based pay. But political turnover makes such support fragile, particularly in times of economic recession, because the cost of performance-based pay is more visible than are the benefits of improved student achievement. [42]

Many of the practical challenges faced by performance-related pay, then, can be addressed through careful design of the system. And despite the opposition of teachers unions to performance-based compensation, it is not clear that the objections to such systems come from the teachers themselves.

Evidence on School-Based Performance Systems

In this section I review evidence on several school-based incentive programs implemented in recent years both in the United States and in other countries. The programs vary in their basic structure and details, with

some targeted at teams of teachers and others at individuals.[43]

Evidence from the United States

Although school-based performance pay theoretically has many attractive features, researchers have been able to find little causal evidence that it is effective in U.S. programs. For example, three researchers studied school-based incentive pay systems in Kentucky, North Carolina (Charlotte-Mecklenburg), and Maryland.[44] They concluded that in the Charlotte-Mecklenburg and Kentucky programs, but not in the Maryland program, both teacher motivation and student outcomes improved. But because all three studies lacked a control group, they could not establish definitively that the program itself—and not some other factor—was the cause of the improvements.[45]

Similarly Helen Ladd studied a school-based bonus program in Dallas.[46] The program, which began in the 1991–92 school year and continued through 1995, ranked schools by how well their students' test scores compared with state average scores, adjusting for students' socioeconomic status. To avoid teaching to the test or other gaming behavior, the program relied on multiple measures of student outcomes, including two tests given each year. Ladd compared gains in school-level test scores in Dallas with gains in other cities (adjusting for many school characteristics, such as racial mix and relative deprivation) to evaluate the impact of this bonus scheme. She found that pass rates appeared to increase more quickly in Dallas than in other cities. Effects were most positive for Hispanics and whites and insignificant for blacks. Although the study suggests that a school-based program can be effective, it was not conclusive. It had, for example, only a limited number of student and school characteristics to adjust to make the participating schools comparable to other schools in the state. In addition, the test score gains in Dallas may have been part of a trend that started before the program was implemented.

The Dallas study also highlights some unintended consequences. In an earlier study, Charles Clotfelter and Ladd had reported that in the Dallas program, schools of low socioeconomic status rarely won awards.[47] In response, the state divided schools into five groups based on socioeconomic characteristics and rewarded the top performers in each group. But some of the lower-performing schools in the upper socioeconomic bands felt that they had been treated unfairly. Dividing the schools into socioeconomic groups also encouraged an undesired strategic response from principals who realized that their ability to gain an award was based on the socioeconomic category into which they were placed.

Finally, two studies of a South Carolina performance-based program that included both school-based and individual-based rewards found that student performance improved.[48] The studies, however, may overstate the incentive effects because teachers could choose whether to apply for an award. If, as would be expected, only the most productive teachers chose to apply, then part of the student gains may be attributable not to the incentives but to the fact that participants were better teachers in the first place.

International Evidence

One of the stronger examples of a school-based incentive program comes from Israel. In February 1995, Israel announced a competition for a monetary bonus for secondary schools and teachers based on their students' performance.[49] The objectives were to re-

duce dropout rates and improve scholastic achievement. The three performance measures were average number of credits per student, share of students receiving a matriculation diploma, and school dropout rate.

Sixty-two schools were initially selected for the program, with several schools added later. In 1996, participating schools competed for about $1.5 million in awards. Schools were ranked according to their annual improvement, adjusting for the socioeconomic background of the students. Only the top third of performers won awards. The distribution of cash incentives among the award-winning schools was determined solely by their ranking in terms of relative improvement (in 1996, the highest-scoring winner won $105,000; the lowest-scoring, $13,250). Teachers received 75 percent of the award as a salary bonus (proportional to gross income); the remainder was used to improve faculty facilities, such as teachers' common rooms. In 1996, the bonuses ranged from 1 to 3 percent of average teacher salary. The combined performance of a team determined the total incentive payment, which was split among individuals regardless of performance.[50]

The student outcomes rewarded included most of those that can be affected by teachers, thereby reducing the dilemma teachers are assumed to face regarding how to allocate their time between rewarded and nonrewarded activities. School averages of all three performance measures were based on the size of the graduating cohort while in ninth grade rather than in twelfth grade. This procedure was adopted to discourage schools from gaming the incentive system—by encouraging weak students to transfer or drop out or by placing them in the nonmatriculation track. To encourage schools to direct more effort toward weak students, only the first 22 credit units taken by each student were counted in computing the school's mean to determine its rank in the bonus program.

Two years after the program was implemented, I compared the program schools with a control group and found significant gains in student performance in the former.[51] Average credits were 0.7 unit higher, the share of students sitting for matriculation examinations increased by 2.1 percent, and average scores and passing rates in these examinations improved as well. Of particular importance was the decline in the dropout rate in students' transition from middle to high school. The programs also appeared mainly to affect weaker students.

Another analysis of a school-based teachers' incentive program, this one in Kenya, examined effects on both teacher behavior and test scores.[52] The program randomly assigned fifty Kenyan primary schools to a treatment group eligible for monetary incentives (21–43 percent of monthly salary). The winning schools were determined by their average test score performance relative to other treatment schools in districtwide examinations; all teachers in the winning schools received awards. The program penalized schools for dropouts by assigning low scores to students who did not take the examination. Data were collected on many types of teacher effort—teacher attendance, homework assignments, pedagogical techniques, and holding extra test preparation sessions—and on student scores obtained after the program's conclusion.

During the two years the program was in place, student scores increased significantly in treatment schools (0.14 standard deviation above the control group). But the gain in scores was not attributable to the expected incentive-induced changes in teacher behav-

ior. In fact, teacher attendance did not improve, and no changes were found in either homework assignment or pedagogy. Instead, teachers were more likely to conduct test preparation sessions outside regular class hours. Data collected the year after the program ended showed no lasting test score gains, suggesting that the teachers focused on improving short-term rather than long-term learning. Consistent with this hypothesis, the program had no effect on dropout rates even though examination participation rose (presumably because teachers wanted to avoid penalties for no-shows). The test score effect was also strongest in geography, history, and Christian religion, arguably subjects involving the most memorization.

Summary

Although group-based pay, either alone or combined with individual-based incentives, has the promise of overcoming some of the difficulties inherent in implementing individual-based systems, the little causal evidence of its effectiveness is mixed. The strongest evidence comes from the Israeli experience; whether it could be replicated either in the United States or abroad is unknown.

Evidence on Individual-Based Performance Systems

In this section I review evidence on several individual-based incentive programs, again both in the United States and abroad.

Evidence from the United States

Studies of individual-based incentive schemes in the United States have had some success in isolating the programs' causal effects on student outcomes. But their findings have also been quite mixed. For example, one study assessed the effect on student achievement of a merit pay scheme in Michigan that rewarded individual teachers according to student retention rates and evaluation questionnaires completed by their students.[53] The scheme, which did not directly target student achievement, did improve student retention. But pass rates fell, while attendance rates and grade point averages remained unchanged. The authors concluded that "incentive systems within complex organizations such as schools may produce results that are unintended and at times misdirected."

In contrast, another study combined panel data from the U.S. National Education Longitudinal Survey of 1988 (NELS88) to estimate the effects of teacher incentives on student outcomes.[54] The authors defined incentive schemes as any merit raise or bonus awarded to any proportion of teachers in a school, although the variables did not identify whether schemes stipulated that rewards were to be tied directly to student achievement. The wealth of data in the survey enabled the authors to control for many student, teacher, school, and family characteristics to make it easier to compare students taught by treatment teachers (those in the incentive scheme) with students taught by teachers in a control group. The results were positive, particularly in public and poor (low economic status) schools. Test scores were higher when awards were higher and when awards were given only to a few teachers within a school.

Finally, a third study analyzed incentive effects on student SAT scores in the Tennessee STAR (Student Teacher Achievement Ratio) and Career Ladder Evaluation programs.[55] It controlled for student and teacher characteristics as well as for class attributes that do not change over time (by including class fixed effects based on panel data). It found that SAT scores improved, with gains varying across subjects and with teacher seniority.

One reason for the mixed evidence may be that the studies combine students at all grade levels. One analysis of merit pay reforms in South Carolina in the 1980s and 1990s suggested that merit pay might be more effective in earlier grades than in later grades. More generally, the study cautioned that the effects of performance-based pay may vary across countries, schools, population groups, or time.

International Evidence

The first international example of an individual-based program is an experiment, begun in fifty high schools in Israel in December 2000, that offered teachers a bonus based on student achievement. The experiment included all English, Hebrew, Arabic, and mathematics teachers who taught tenth- to twelfth-grade classes in preparation for matriculation examinations in these subjects in June 2001. Each teacher was ranked separately on the basis of the mean performance of each class she taught. The ranking was based on the difference between actual class performance and performance predicted on the basis of students' socioeconomic characteristics, their level of proficiency in each subject, and a fixed school-level effect. Each teacher was ranked twice, once for the students' passing rate and once for average score.

Each school submitted student enrollment lists, itemized by grade, subject, and teacher, on the program's starting date. All students on these lists were included in the class mean outcomes. Students who dropped out or did not take the exams, regardless of the reason, were imputed a score of zero to neutralize any incentive for teachers to keep poorly performing students out of the tests.

All teachers who performed better than predicted in both passing rate and average score were ranked from first to fourth place and awarded points according to ranking. The awards, based on total points, ranged from 6 to 25 percent of the average annual income of high school teachers. A teacher could win several awards if she prepared more than one class for a matriculation examination.[56] Of the 629 teachers in the program, 302 won awards.

My analysis of the program found that it significantly improved matriculation examination participation rates as well as the passing rate and average test scores among those who took the test.[57] These gains accounted for about half of the improved outcomes among all students. They appear to have resulted from changes in teaching methods, afterschool teaching, and increased responsiveness to students' needs, not from artificial inflation or manipulation of test scores. The evidence that the incentive program improved teacher effort and pedagogy is important in the context of concerns about the programs' unintended effects, such as teaching to the test or cheating and the manipulation of test scores, and fears that such programs do not produce real learning.

Although group-based pay . . . has the promise of overcoming some of the difficulties inherent in implementing individual-based systems, the little causal evidence of its effectiveness is mixed.

As a second example, in 1999 the United Kingdom introduced a systemwide perform-

ance-related pay policy for teachers using student progress (value added) as one key criterion.[58] Using long-term teacher data and a before-and-after comparison research design, one study evaluated the policy's effect on test scores on the important GCSE (General Certificate of Secondary Education) exams, taken by students at age sixteen at the end of compulsory education.[59] Because the incentive scheme was explicitly teacher- rather than school-based, the study followed teachers over two complete two-year teaching cycles before and after the policy was introduced. Students were linked to teachers who taught them specific subjects, making it possible to compare treatment and control group teachers. This panel data structure also made it possible to control for student and teacher unchanged characteristics (by including respective fixed effects) and to measure the scheme's target: student progress.

The study reported statistically and economically significant student progress. For instance, relative to control teachers, treatment teachers increased their value added by almost half a GCSE grade per student, equal to 0.73 standard deviation. Also significant were differences between school subjects, with treatment math teachers showing no effect from participating in the program. Although promising, again, the study is not definitive. One concern is that it applied only to teachers who had been in the profession for about eight years, so that treatment and control teachers differed systematically in experience. If teachers improve in their capacity to generate value added as they gain more experience but at a decreasing rate, then, all else being equal, one can expect to see greater improvements in progress between the two teaching cycles for the less experienced (control) group. Taking teacher experience into account in the analysis may not solve the problem if the relationship between teachers' experience and productivity is nonlinear.

Summary

In general, the evidence suggests that well-designed individual-based incentives can significantly improve student outcomes. But the research base is small, and implementing purely individual-based programs presents many challenges.

Policy Implications of the Evidence on Teachers' Incentives

Research evidence suggests, although not conclusively, that pay-for-performance incentives can improve teachers' performance. Yet these incentives can also lead to unintended and undesired consequences growing out of their inherent structural challenges, such as measurement issues or the possibility of teachers' directing their efforts exclusively to rewarded activities. Although there are no magical cures for any of these problems, it is possible to draw lessons from the experience and evaluation of the better performance-related pay programs. In this section I offer some general guidelines for designing effective programs.

Which Outcomes to Reward

Teachers should be evaluated on the basis of their true performance, not random variation in performance. Therefore, the performance measures chosen should exhibit relatively little random variation. Because teachers are expected to direct their efforts according to the incentives provided, the performance measures should cover all outcomes of interest, including quality and quantity (relevant examples are the two outcome measures used in the Israeli pay programs—passing rate and average score). Outcome measures should be as close as possible to the ultimate

outcomes, difficult to game, and easy to monitor regularly.

Monitoring

Close supervision and monitoring are needed to minimize gaming and shading the truth, if not outright fraud, in measured output. Because monitoring can be costly, an appropriate strategy should balance the trade-off between penalties and probability of detection. The matriculation examination system in Israel minimizes gaming: students are tested twice—in a school and a state examination—in each subject, and their final score is a weighted average of the two. Large gaps between the scores invite penalties. Nonrandom gaps between the two scores are audited at the class and school levels. Obviously, even the best-designed measurement and monitoring system may lead to inappropriate reallocation of effort because no single system is likely to capture all important behavior. One way to avoid this constriction in effort is to include two outcomes, for instance, test scores and processes (for example, attention directed explicitly to disadvantaged students) and teacher practices (for example, teacher absenteeism).

Reward Criteria

Teachers can be compared according to student performance levels or to some form of gain in performance. Value-added measures, in particular, are more appropriate for measuring individual teacher effectiveness and ensuring fair ranking. Such methods can substantially affect school or teacher rankings, in addition to the behaviors engaged in to game the system. Performance levels create incentives to pay greatest attention to highest-performing students; value-added criteria may increase attention to lowest-performing students. It is thus important to use performance measures that can be adjusted for contributing factors, such as student socioeconomic characteristics, but not at the cost of complex value-added measures that lack transparency and are difficult to understand and accept. In addition, goals should be attainable and not be perceived as too ambitious.

Rewarding Both Short- and Long-Term Measures

Performance-related pay programs typically focus on key short-term schooling objectives—such as increasing the number of students reading at a given grade level—rather than on long-term objectives related to postsecondary education and labor market outcomes. This emphasis is due in large part to the difficulty of measuring long-term objectives, as well as to an almost universal belief that achieving short-term goals is a necessary condition for promoting long-term goals. But medium-term outcomes, such as dropout rates, can be affected by performance-related pay. The team-based program developed in Israel demonstrates how a mix of short- and medium-term goals can be achieved simultaneously in an effective performance-related pay program.

Subjective and Objective Evaluation

Subjective evaluation invites the problem of performance measure inflation (namely, that teachers under the pressure of incentives tend to give higher than warranted scores), while objective metrics invite the problems of measurement cost and the inability to encompass all school targets and goals. Given these limitations, it may be appropriate to use both—each imperfect but still informative. For example, selective emphasis can be assigned to each subjective and objective evaluation criterion, depending on subject taught or specific task, and associated with individual, team, and school performance. The evaluation process and its

results should be transparent to all parties involved.

School and Individual Incentives

Because some degree of teamwork characterizes all schools, incentives should balance individual rewards with school incentives. Design of these incentives should foster a cooperative culture, but not at the cost of an aggravated free-riding problem, a condition likely to arise when only group incentives are used.

Monetary and Other Incentives

Money is not alone in motivating performance and promotions. Improved working conditions and increased decisionmaking authority may also be attractive to teachers, especially when combined with bonuses or salary increases. The Israeli school incentive program included modest monetary bonuses but large rewards in the form of media attention and enhanced reputation for the winning schools. Recognition and prestige are, it appears, highly effective motivators.

Eligibility and Size of Awards

All teachers should be eligible for the incentive offered. Only when the majority of teachers are eligible to enjoy the benefits of hard work and improved outcomes will the incentive scheme be effective. However, only a subset of teachers should be rewarded in practice. If too many teachers are rewarded, teachers may not need to exert much effort to benefit.

Flexibility

A performance-related pay program can motivate teachers and schools even more effectively if it also allows teachers to make decisions regarding instruction, curriculum, and other aspects of schooling that contribute to attaining the desired outcomes and goals. Creativity flourishes in the absence of bureaucratic constraints. The search for ways to improve teaching technology arises both because teachers and schools have a direct stake in the matter and because merit pay plans reduce the need for regulation of inputs and thus give teachers and schools more freedom. It is therefore essential that performance-related pay plans be accompanied by the increased flexibility required for their success.

Conclusions

Recent reforms in public education have in common a heightened stress on effectiveness—setting standards, measuring progress toward those standards, and imposing rewards or penalties for meeting or failing to meet them. Performance-related pay systems use monetary rewards to meet these goals. In most cases these systems involve marginal changes rather than a complete revamping of the educational system. Research evidence suggests, though not conclusively, that incentive-based compensation can generate gains in student performance and teacher effectiveness. Teachers and schools appear to respond to monetary incentives by exerting more effort, applying more creativity, and modifying their pedagogical practices. Measuring and rewarding performance are thus potentially important elements in modern public school compensation systems. They can also, theoretically, discourage poor performers and attract better performers in the medium and longer terms.

It is telling that private schools typically have more flexible pay structures than public schools and that they are more likely to use variations of performance-related pay. Dale Ballou reports that in 1993, 12 percent of U.S. public school districts had merit pay plans, as compared with 35 percent of nonsectarian private schools.[60] Merit pay can

clearly be integrated and implemented in the private schools. Political or technical constraints may be impeding its success in public schools.

Implementing performance-based pay requires meeting the daunting challenge of devising a system for measuring performance. The system must measure true performance in a way that minimizes random variation, as well as undesired and unintended consequences. It must align performance with ultimate outcomes and monitor performance to discourage cheating. Clearly, setting up an effective performance-related pay system is not a one-time task, but an ongoing effort. Even with the best preparation, initial implementation is likely to be problematic. Measurements will have a random component. Teachers will find ways to game the system. Any initial system will almost certainly be flawed. But if the effort is seen as an ongoing one, it should be possible to make progress gradually in addressing each of the challenges. For example, better measurement instruments should shrink the random component in the measures, measurements can be adjusted to prevent gaming, careful monitoring can detect and therefore deter cheating. Eventually incentives will be developed that motivate the desired teaching behaviors and will be perceived by teachers as fair and accurate.

Although the suggestions above are consistent with most of the limited empirical evidence available on the effectiveness of performance-related pay strategies, researchers have much work to do. Their efforts are impeded both by the reluctance of school systems to conduct careful evaluations and by a lack of appropriate outcome measures and comparison groups. Starting now, researchers should follow closely all efforts to implement performance-based pay in public schools, tracking carefully the behavioral changes induced and the ultimate outcomes of monetary rewards given to school staff. Meeting these research challenges is critical to the design of effective performance-related pay systems in the future.

Notes

1. For example, in describing the need for fundamental reform of the system, the U.S. Department of Education noted that while real spending on K–12 education has grown substantially over the last decade, academic achievement as measured by the National Assessment of Educational Progress (NAEP) has barely budged for most student categories. Such comparisons are often used in policy debates to buttress the imposition of new forms of accountability on educators instead of continuing to "throw money at schools."

2. William J. Slotnick and Maribeth D. Smith, *Catalyst for Change: Pay for Performance in Denver Final Report* (Boston: Community Training and Assistance Center, 2004). The study found the largest gains for high school students.

3. High school test scores are significantly correlated with adult (mid-20s through mid-30s) earnings, even after controlling for background variables. See R. J. Murnane, J. B. Willett, and F. Levy, "The Growing Importance of Cognitive Skills in Wage Determination," *Review of Economics and Statistics* 77 (1995): 251–66; Derek A. Neal and William R. Johnson, "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy* 104, no. 5 (1996): 869–95; J. S. Zax and D. I. Rees, "IQ, Academic Performance, Environment, and Earnings," *Review of Economics and Statistics* 84, no. 4 (2002): 600–16. Janet Currie and Duncan Thomas, "Early Test Scores, Socioeconomic Status and Future Outcomes," Working Paper 6943 (Cambridge, Mass.: National Bureau of Economic Research, 1999) present estimates based on test scores of much younger children (aged seven). They find a correlation with adult earnings, after controlling for factors such as father's socioeconomic status (SES), and father's and mother's and maternal grandfather's education. These correlations lend additional support to the assumption that increased learning—as proxied by test scores—in elementary school or high school will lead to better labor market outcomes in adulthood.

4. See, for example, Edward P. Lazear, "Paying Teachers for Performance: Incentives and Selection," Stanford University, mimeo, 2001; Lewis Solomon and Michael Podgursky, "The Pros and Cons of Performance-Based Compensation" (Pasadena, Calif.: Milken Family Foundation, 2001); and H. Tomlinson, "Proposals for Performance Related Pay in English Schools," *School Leadership and Management* 20, no. 3 (2000): 281–98.

5. There is some evidence that in performance-based systems, principals tend to evaluate teachers more critically than they would otherwise; see R. Murnane and D. Cohen, "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive," *Harvard Educational Review* 56, no. 1 (1986): 1–17. As a precaution, Solomon and Podgursky suggest that principals themselves become subject to schoolwide performance-based evaluation to ensure the continued objectivity of their evaluations; see Solomon and Podgursky, "The Pros and Cons" (see note 4).

6. C. Kelley, "The Motivational Impact of School-Based Performance Awards," *Journal of Personnel Evaluation in Education* 12, no. 4 (1999): 309–26.

7. See, for example, T. Hoerr, "A Case for Merit Pay," *Phi Delta Kappan* 80, no. 4 (1998): 326–27.

8. See, for example, Solomon and Podgursky, "The Pros and Cons" (see note 4); and Tomlinson, "Proposals for Performance Related Pay" (see note 4).

9. However, in the United States not only are there many options for choice within the public schools (such as citywide choice systems and magnet and charter schools), but also individuals can effectively choose where

to live based on the quality of the schools. And there is compelling evidence that school quality affects where households choose to live; see, for example, Lisa Barrow, "School Choice through Relocation: Evidence from the Washington, D.C., Area," *Journal of Public Economics* 86, no. 2 (2002): 155–89; Lisa Barrow and Cecilia E. Rouse, "Using Market Valuation to Assess Public School Spending," *Journal of Public Economics* 88, nos. 9–10 (2004): 1747–69.

10. See R. Richardson, "Performance Related Pay in Schools: An Assessment of the Green Papers," Report prepared for the National Union of Teachers (London School of Economics and Political Science, 1999).

11. See D. Evans, "No Merit in Merit Pay, "*American School Board Journal* 188, no. 1 (2001): 48–50; and M. Holt, "Performance Pay for Teachers: The Standards Movement's Last Stand?" *Phi Delta Kappan* 83, no. 4 (2001): 321–28.

12. American Federation of Teachers, "Merit Pay," Working Paper (2001), www.aft.org/issues/meritpay/meritpay.html (January 15, 2003); L. Barber and K. Klein, "Merit Pay and Teacher Evaluation," *Phi Delta Kappan* 65, no. 4 (1983): 247–51; T. Cutler and B. Waine, "Mutual Benefits or Managerial Control? The Role of Appraisal in Performance Related Pay for Teachers," *British Journal of Educational Studies* 48, no. 2 (2000): 170–82; M. Holt, "Performance Pay for Teachers" (see note 11); and A. Ramirez, "How Merit Pay Undermines Education," *Educational Leadership* 58, no. 5 (2001): 16–20.

13. See Bengt Holmstrom and P. Milgrom, "Multi-Task Principal-Agent Problems: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization* 7 (1991).

14. R. Chamberlin and others, "Performance-Related Pay and the Teaching Profession: A Review of the Literature," *Research Papers in Education* 17, no. 1 (2002): 31–49.

15. See, for example, Paul Glewwe, Nauman Ilias, and Michael Kremer, "Teacher Incentives," Working Paper 9671 (Cambridge, Mass.: National Bureau of Economic Research, 2003).

16. The reallocation of effort and resources toward subjects that are measured and rewarded and away from subjects that are not measured and thus not rewarded has been documented empirically. For example, teachers have reported spending more time on tested topics and less on untested topics as a result of high-stakes testing programs; see L. A. Shepard and K. C. Dougherty, "Effects of High-Stakes Testing on Instruction," paper presented at the annual meeting of the American Educational Research Association and National Council on Measurement in Education, Chicago, 1991; M. L. Smith and others, *The Role of Testing in Elementary Schools,* CSE Technical Report 321 (Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, 1991); B. M. Stecher and others, *The Effects of the Washington State Education Reform on Schools and Classrooms,* CSE Technical Report 525 (Los Angeles, Calif.: Center for Research on Evaluation, Standards, and Student Testing, 2000). Teachers have also reported stressing certain formats or styles used in test items in their instruction; see D. M. Koretz and L. S. Hamilton, *Teachers' Responses to High-Stakes Testing and the Validity of Gains: A Pilot Study,* CSE Technical Report 610 (Los Angeles, Calif.: Center for Research on Evaluation, Standards, and Student Testing, 2003); J. J. Pedulla and others, *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers* (Boston, Mass.: National Board on Educational Testing and Public Policy, 2003); T. A. Romberg, E. A. Zarinia, and S. R. Williams, *The Influence of Mandated Testing on Mathematics Instruction: Grade 8 Teachers' Perceptions* (National Center for Research in Mathematical Science Education, University of Wisconsin-Madison, 1989). Shepard and Dougherty, for instance, found that in two districts with high-stakes writing tests, writing teachers admitted that as a result of the format of the writ-

ing test used in those districts, they emphasized student searches for mistakes in written work rather than the production of students' own writing.

17. See B. Malen, "On Rewards, Punishments, and Possibilities: Teacher Compensation as an Instrument for Education Reform," *Journal of Personnel Evaluation in Education* 12, no. 4 (1999): 387–94.

18. Brian Jacob and Steve Levitt, "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics* 3 (2003): 843–77.

19. David N. Figlio and J. Winicki, "Food for Thought: The Effects of School Accountability Plans on School Nutrition," Working Paper 9319 (Cambridge, Mass.: National Bureau of Economic Research, 2002); and David N. Figlio and Lawrence S. Getzler, "Accountability, Ability and Disability: Gaming the System," Working Paper 9307 (Cambridge, Mass.: National Bureau of Economic Research, 2002).

20. Ernst Fehr and Klaus M. Schmidt, "Fairness and Incentives in a Multi-Task Principal-Agent Model," *Scandinavian Journal of Economics* 106, no. 3 (2004): 453–74.

21. David Kreps, "Intrinsic Motivation and Extrinsic Incentives," *American Economic Review* 87, no. 2 (1997): 359–64.

22. This concern could be more meaningful in developing countries, where incentives are employed to increase teachers' work attendance. For example, if the incentives are based solely or primarily on presence in the classroom, teachers may come to believe that class attendance per se is more important than their performance in the classroom. Equally worrisome is the possibility that teachers who previously believed that they were required to work every day in the month might increase their absenteeism once they reached their target monthly income; Ernst Fehr and Lorenz Gotte, "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment," Working Paper 125 (University of Zurich, 2002).

23. Murnane and Cohen, "Merit Pay and the Evaluation Problem" (see note 5).

24. See Lazear, "Paying Teachers for Performance" (see note 4).

25. W. Firestone and J. Pennell, "Teacher Commitment, Working Conditions, and Differential Incentive Policies,'" *Review of Educational Research* 63, no. 4 (1993): 489–525.

26. See D. Ballou and M. Podgursky, "Teachers' Attitudes towards Merit Pay: Examining Conventional Wisdom," *Industrial and Labor Relations Review* 47, no. 1 (1993): 50–61; S. McCollum, "How Merit Pay Improves Education," *Educational Leadership* 58, no. 5 (2001): 21–24; H. Tomlinson, "Proposals for Performance Related Pay in English Schools," *School Leadership and Management* 20, no. 3 (2000): 281–98.

27. Recently, though, a dent has been observed in this formerly united front in the United States with the appearance of a group of teachers unions that supports the efforts of the Consortium for Research and Policy in Education (CRPE) to introduce knowledge- and skills-based pay; see A. Odden, "Paying Teachers for Performance," *School Business Affairs* (June 2000): 28–31.

28. Ballou and Podgursky, "Teachers' Attitudes towards Merit Pay" (see note 26).

29. C. Kelley, H. Heneman, and A. Milanowski, "Teacher Motivation and School-Based Performance Awards," *Education Administration Quarterly* 38, no. 3 (2002): 372–401.

30. Ibid.

31. A. Odden, "New and Better Forms of Teacher Compensation Are Possible," *Phi Delta Kappan* 81, no. 5 (2000): 361–66.

32. Cutler and Waine, "Mutual Benefits or Managerial Control?" (see note 12).

33. A. Mohrman, S. Mohrman, and A. Odden, "Aligning Teacher Compensation with Systemic School Reform: Skill-Based Pay and Group-Based Performance Awards," *Educational Evaluation and Policy Analysis* 18, no. 1 (1996): 51–71.

34. Brian Jacob and Lars Lefgren, "Principals as Agents: Subjective Performance Measurement in Education," Working Paper 11463 (Cambridge, Mass.: National Bureau of Economic Research, May 2005).

35. Several past studies in education found little correlation between principal-based teacher evaluations; see, for example, D. M. Medley and H. Coker, "The Accuracy of Principals' Judgments of Teacher Performance," *Journal of Educational Research* 80 (1987): 242–47. However, David Armor and others, *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools*, Report R-2007-LAUSD (Santa Monica, Calif.: RAND, 1976), found that principal evaluations of teachers predicted student achievement even after conditioning on prior student test scores and a host of other student and classroom-level demographic controls. The authors argue that these results indicate that principals *can* identify effective teachers. While these findings are suggestive, this literature has many limitations. Most of the studies involved extremely small, often nonrepresentative samples, and the methodologies used often do not account for selection and measurement error issues.

36. E. Cohn, "Methods of Teacher Remuneration: Merit Pay and Career Ladders," in *Assessing Educational Practices: The Contribution of Economics*, edited by W. Becker and W. Baumol (New York: Russell Sage Foundation, 1996); Mohrman, Mohrman, and Odden, "Aligning Teacher Compensation with Systemic School Reform" (see note 33).

37. See W. Bengt Holmstrom, "Managerial Incentive Problems—A Dynamic Perspective," republished in *Review of Economic Studies* 66 (1999): 169–82.

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

39. See Eugene Kandel and Edward Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy* 100, no. 4 (1992): 801–17.

40. D. Ballou and M. Podgursky, "Teacher Pay and Teacher Quality" (Michigan: W. E. Upjohn Institute for Employment Research, 1997).

41. Ibid.

42. Ibid.

43. Additional details and evidence on the effects of some of these programs, as well as of other performance-related pay programs, are presented in Peter Dolton, Steve McIntosh, and Arnaud Chevalier, *Pay and Performance*, Bedford Way Papers (University of London, Institute of Education, 2003).

44. Kelley, "The Motivational Impact of School-Based Performance Awards" (see note 6); H. G. Heneman and A. T. Milanowski, "Teachers' Attitudes about Teacher Bonuses under School-Based Performance Award

Programs," *Journal of Personnel Evaluation in Education* 12 (1999): 327–41; Kelley, Heneman, and Milanowski, "Teacher Motivation and School-Based Performance Awards" (see note 29).

45. S. Smith and R. Mickelson, "All That Glitters Is Not Gold: School Reform in Charlotte-Mecklenburg," *Educational Evaluation and Policy Analysis* 22, no. 2 (2000): 101–27, evaluated the Charlotte-Mecklenburg program outcomes by contrasting them to outcomes observed in nonparticipating North Carolina urban school districts. They compared progress on SAT scores and dropout rates for a range of age levels in program schools against statewide averages. However, they did not adequately control for potential differences between program schools and nonprogram schools. Further, Charlotte-Mecklenburg is a complex reform program that involves many policy developments in addition to performance-based rewards. This makes it difficult to identify the unique effect of the performance-related pay programs. More recently in North Carolina, another interesting incentive program was implemented for a three-year period beginning in 2001. The program awarded an annual bonus of $1,800 to certified math, science, and special education teachers working in high-poverty public secondary schools. Charles Clotfelter and others, "Compensating Salary Differentials for Teachers in High-Poverty Schools: Evidence from a Policy Intervention in North Carolina" (Sanford Institute of Public Policy, Duke University, 2006), suggest that this sum was sufficient to reduce mean turnover rates by 12 percent, and even more among experienced teachers.

46. H. Ladd, "The Dallas School Accountability and Incentive Program: An Evaluation of Its Impacts on Student Outcomes," *Economics of Education Review* 18, no. 1 (1999): 1–16.

47. C. Clotfelter and H. Ladd, "Recognizing and Rewarding Success in Public Schools," in *Holding Schools Accountable: Performance-Based Reform in Education,* edited by H. Ladd (Brookings, 1996).

48. S. T. Cooper and E. Cohn, "Estimation of a Frontier Production Function for the South Carolina Educational Process," *Economics of Education Review* 16 (1997): 313–27; M. Boozer, "The Design and Evaluation of Incentive Schemes for Schools: Evidence from South Carolina's Teacher Incentive Pay Project," paper prepared for the National Academy of Sciences Conference on Devising Incentives to Improve Human Capital, Irvine, Calif., December 17–18, 1999.

49. Details of this program are provided in a publication issued by the chief scientist of the Israeli Ministry of Education: *The Differential Compensation: Principles for Allocation* (Jerusalem, 1995; in Hebrew).

50. This program closely fits the framework of a rank order tournament as analyzed in E. Lazear and S. Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy* 89 (1981): 841–64; and Jerry Green and Nancy L. Stokey, "A Comparison of Tournaments and Contracts," *Journal of Political Economy* 91 (1983): 349–64. In such schemes, prizes depend on the rank order of the winner of a contest or tournament.

51. See Victor Lavy, "Evaluating the Effect of Teachers' Performance Incentives on Pupils' Achievements," *Journal of Political Economy* 110 (December 2002): 1286–317.

52. Glewwe, Ilias, and Kremer, "Teacher Incentives" (see note 15).

53. R. Eberts, K. Hollenbeck, and J. Stone, "Teacher Performance Incentives and Student Outcomes," WP00-65 (Michigan: Upjohn Institute for Employment Research, 2000).

54. D. N. Figlio and L. W. Kenny, "Do Individual Teacher Incentives Boost Student Performance?" (University of Florida, 2003).

55. T. Dee and B. Keys, "Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment," *Journal of Policy Analysis and Management* 23, no. 3 (2004): 471–88.

56. For more details, see Israel, Ministry of Education, High School Division, *Individual Teacher Bonuses Based on Student Performance: Pilot Program* (Jerusalem, December 2000; in Hebrew).

57. Victor Lavy, "Paying for Performance and Teachers' Effort, Productivity, and Grading Ethics," Working Paper 10622 (Cambridge, Mass.: National Bureau of Economic Research, 2004).

58. See Dolton, McIntosh, and Chevalier, "Pay and Performance" (see note 43), for an extensive survey and discussion of the performance-related pay program for teachers in the United Kingdom.

59. Adele Atkinson and others, "Evaluating the Impact of Performance-Related Pay for Teachers in England," CMPO Working Paper 04/113 (Bristol, 2004); and S. Burgess and others, "The Intricacies of the Relationship between Pay and Performance for Teachers: Do Teachers Respond to Performance Related Pay schemes?" CMPO Working Paper 01/35 (Bristol, 2001).

60. Dale Ballou, "Pay for Performance in Public and Private Schools," *Economics of Education Review* 20 (2001): 51–61.