Program on Education Policy and Governance Working Papers Series

Market- and Performance-Based Reforms of Teacher Compensation: A Review of Recent Practices, Policies, and Research

Michael J. Podgursky
University of Missouri

Matthew Springer
Vanderbilt University

PEPG 10-09

Prepared for the PEPG Conference
Merit Pay: Will It Work? Is It Politically Viable?

Harvard Kennedy School
Cambridge, Massachusetts
June 3-4, 2010

Market- and Performance-Related Reforms of Teacher Compensation:

A Review of Recent Practices, Policies, and Research

M. Podgursky, University of Missouri - Columbia

M.G. Springer, Peabody College of Vanderbilt University

March 12, 2010

Abstract

This paper provides a review of recent policy initiatives to reform teacher compensation systems and evidence regarding the effect of these policies. The first section examines the current structure of teacher compensation in the U.S. K-12 public education system. The compensation "system" for teachers is fragmented and uncoordinated. Teacher compensation is largely set by salary schedules that are neither market-oriented nor performance-driven. The second section reviews pay reforms being implemented in U.S. public school districts. The third section of the paper examines the small but growing evaluation literature on compensation reform, paying particular attention to evidence from studies using experimental and quasi-experimental designs to assess the impact of the program on student achievement and teacher outcomes. A final section provides observations on prospects for future reform, and suggestions for policy research.

This paper was prepared for the Harvard Program on Education Policy and Governance conference on merit pay, Cambridge, Massachusetts, June 2-4, 2010. The authors wish to acknowledge the National Center on Performance Incentives at Vanderbilt University and the Center for Analysis of Longitudinal Data in Education Research for research support. We also thank Yi Du for research assistance. The usual disclaimers apply.

1. Introduction

During the 2006-07 school year, the most current year for which national data are available, U.S. public schools spent $197 billion for salaries and $64 billion for benefits for instructional personnel (U.S. Department of Education, 2009). These compensation payments account for 55 percent of current expenditures in K-12 public schools and 90 percent of instructional expenditures. As large as these expenditures are, they do not fully capture the resources committed to K-12 compensation, since they do not include the billions of dollars of unfunded liabilities of pension funds and retiree health insurance for teachers and administrators (Pew Center on the States, 2008, 2010; Clark, 2009). If productivity doubles for an input accounting for one percent of total cost, there will be little overall efficiency gain. However, given the large share of K-12 costs that arise from teacher compensation, even small gains in efficiency would yield large benefits.

The current compensation "system" for public school teachers is neither strategic nor integrated. Rather, it is best seen as an amalgam of components, reflecting divergent stakeholder preferences, legislative tinkering, and legacies from earlier vintages of employment contracts, with little consideration for overall efficiency. For example, teacher base pay is set by salary schedules that have evolved generations of collective bargaining agreements, or in non-bargaining states like Texas, legislative fiat. Base pay is augmented by various types of district or state-wide salary supplements (e.g., coaching, career ladder). Additionally, deferred compensation in the form of retirement pay inhabits another silo altogether, with policy typically set by statewide pension boards dominated by senior teachers and administrators. Teacher compensation is the sum of all of these parts (plus fringe benefits such as health insurance).

An optimal teacher compensation system would be a pay package that is designed to recruit, retain, and develop the highest quality professional workforce for any given level of expenditures while recognizing and rewarding excellence. Teacher compensation systems also should be aligned to a school's strategic mission, supporting the creation of value for stakeholders and employees. In practice, however, the pieces of teacher compensation systems – current salary, additional compensation, benefits, and deferred compensation – are set in ad hoc ways with little coordination or consideration of strategic tradeoffs, or tested against labor market benchmarks.

In this paper we review the recent policy initiatives concerning teacher performance pay and the research evidence accumulated to date concerning the effectiveness of these policies. We conclude with some observations on prospects for future reform and suggestions for policy research.

2. The Single Salary Schedule and It's Consequences

The most important determinant of a teacher's pay is the salary schedule in the school district. District salary schedules have been nearly universal in the public school system since the early-1950s, although some locations (primarily southern states) have state-wide teacher salary schedules that set a minimum pay level but allow for local districts to supplement these minimums. During the 2003-04 school year, about 96 percent of public school districts accounting for nearly 100 percent of all public school teachers reported use of a salary schedule (Podgursky, 2007).

Table 1 displays the 2008-09 salary schedule for teachers in the Houston (TX) Independent School District (HISD). The rows represent years of teaching experience and the three blocked columns identify post-secondary degrees. Similar to most teacher compensation

systems found in U.S. public school districts, the salary schedule in HISD provides larger salaries to teachers with higher-levels of formal education and for each additional year of teaching experience. This class of salary schedule is sometimes referred to as a single salary schedule, which reflects their historical development (Podgursky and Springer, 2007; Protsik, 1996).

Insert Table 1 Here

Single salary schedules for teachers contrast with the situation in most other professions where merit or performance-related pay is more commonplace. In medicine, pay of doctors and nurses varies by specialty. Even within the same hospital or HMO, pay will differ by specialty field. In higher education there are large differences in pay between faculty by teaching field. Faculty pay structures tend to be flexible. Starting pay is usually market-driven as institutions will often match counter-offers for more senior faculty whom they wish to retain. Studies report generally similar findings for private K-12 education (Ballou and Podgursky, 1997; Ballou, 2001). Even when private schools report the use of a salary schedule to determine teacher pay levels, payments "off schedule" are frequent.

Salary schedules would not be as costly if the factors rewarded, teacher experience and graduate education, were strong predictors of teacher productivity. However, surveys of the education production function literature find little support for a masters degrees positively impacting student achievement, and teacher experience has little effect beyond the first few years (Rivkin et al., 2005; Clotfelter, Ladd, and Vigdor, 2006; Hanushek, et.al. 2005; Aaronson, Sanders, Barrow, 2007). Hanushek (2003) reports of 41 “value-added” estimates of the effect of a teacher’s education level on her effectiveness (primarily MA’s) that not a single study found a

statistically significant positive effect. In fact, ten of the studies found statistically meaningful negative effects.

In spite of the depth and consistency of this finding in the research literature, school districts continue to spend billions of dollars annually rewarding MA degrees. For example, as noted by Roza and Miller (2009), between 1997 and 2007, the education-specific master's had the highest growth rate of all master's degrees. About 90 percent of teachers' master's degrees are not subject specific. Hassell (2008) further notes that the public school system could annually award the top 50 percent of teachers performance bonuses averaging $13,000 with the top teachers earning $20,000 or more if advanced degree premiums were reduced by approximately 80 percent.

There is an adage in economics: *You can't repeal the law of supply and demand*. By this economists mean that if governments or regulatory agencies do not allow prices to clear a market then some other mechanism will. For example, if city governments use rent controls to set rates below the market clearing level, then shortages will develop. In such a case the market will clear in the sense that individuals wanting to reside within the city proper will have to invest more of their time searching for an apartment. Some prospective residents will give up and quit, while others may pay bribes or find other ways to work around the system. It is likely the overall quality of the apartment stock will decline over the long-run. Ultimately, non-price mechanisms such as rent controls can act to clear the market instead of moving it towards prices which balance the quantity supplied and the quantity demanded.

The rigidities in most teacher salary schedules have similar consequences in regard to teacher labor markets, such that the quality of the teacher labor supply suffers from the market being out of equilibrium. To illustrate, we provide information on teacher shortages by field, the

concentration of novice teachers in high-poverty schools, and the incentives (or lack thereof) for teachers to stay on the job or enter into the profession.

1.1. Shortages by Field

The training, working conditions, and non-teaching opportunities for teachers differ significantly by teaching field, yet the salary schedule within a school district treats all teachers the same. On average the non-teaching opportunities for a high school physical science teacher (or a degree in any technical field) are more remunerative than for elementary education teachers, yet the salary schedule within a school district gives them identical salaries. Since salaries are rigid, the market thus clears in a quality dimension.

Data from a nationally representative survey of school principals conducted by the U.S. Department of Education nicely illustrates the consequences of these rigidities. Principal respondents were asked a series of questions about how difficult or easy it was for them to fill teaching vacancies by fields. If a vacancy existed in a particular field, the perceived difficulty to fill the vacancies was reported by principal respondents on a four point scale where a value of one denoted it was *easy* and four meant the principal *could not fill the vacancy*.

Analysis of these data indicated that approximately 75 percent of principal respondents who needed to hire an elementary education teacher reported that it was *easy* to fill the vacancy in the 2003-04 school year. Yet only 30 to 35 percent of principal respondents who needed to fill a science, mathematics, or special education opening gave such an assessment. More to the point, two percent of principals with elementary education teacher vacancies reported that it was *very difficult* to fill the opening or that they *could not fill the vacancy*. Similar statistics were around 30 percent for science and mathematics vacancies, 21 percent for biology vacancies, and 35 percent for special education vacancies. It therefore may come as little surprise that science,

mathematics, and special education teachers tend to be less likely to have majored in their primary field of instruction and are more likely to be classified as teaching "out of field" than elementary school teachers (Podgursky, 2008; U.S. Department of Education, 2004).

1.2. Inequitable Distribution of High Quality Teachers

The inequitable distribution of high quality teachers across schools has helped perpetuate the student achievement gap in many urban school districts. An oft-cited research finding is that novice teachers (e.g., first or second year teachers) tend to produce smaller achievement gains for their students than more experienced teachers (Aaronson, Barrow, Sander, 2007; Rivkin, Hanushek, Kane, 2001). Most public schools differ in attractiveness as places to teach. Schools with larger concentrations of non-free and/or reduced-price lunch students are perceived as nicer places to work than schools serving mostly low-income students. More experienced teachers therefore leverage seniority provisions that allow teachers with the most experience to transfer to any open teaching position in a school district. Or, if they can't transfer, they will simply quit. In either case, children enrolled in economically disadvantaged schools with high concentrations of non-white students will have greater exposure to teachers graduating from less competitive colleges, novice teachers, and teachers instructing out-of-field (Lankford, Loeb, and Wyckoff, 2002; Iaterola and Steifel, 2003; Roza et al., 2007).

The inequitable distribution of high-quality teachers among schools is a direct consequence of uniform teacher salary schedules in conjunction with differences in nonpecuniary characteristics of schools. If pay is equalized then teacher quality is disequalized across schools. In order to equalize teacher quality schools have begin to experiment with a wage premium that neutralizes differences in nonwage job characteristics across schools (Prince, 2002). As there is little research on the additional amount of money needed to offset differences

in nonpecuniary workplace characteristics a number of school systems have proposed policies to increase the supply and equitable distribution of high quality teachers, including reduced entry requirements into the profession and providing high-need schools block grants so they can better meet the needs of their students.

1.3. Lack of Incentives for More Effective Teachers to Stay on Job or Enter Profession

A consistent research finding is that there is considerable variability in teacher effectiveness (Aaronson, Barrow, Sander, 2007; Hanushek, Kain, O'Brien, Rivkin, 2005; Kane, Staiger, Rockoff, 2006). Some teachers are consistently better at raising the achievement of their students than others. In fact, the difference in learning growth for a student exposed to a teacher in the top performance quintile versus the bottom quintile of effectiveness is substantial, and if cumulated over several school years, could substantially narrow or widen achievement gaps.

However, by depending so heavily on the single salary schedule, the great majority of U.S. public school systems have not been able to leverage incentive pay to encourage highly-effective teachers to remain in the profession or transfer to a high-needs school. A more efficient pay structure would attempt to retain the best teachers and find ways to shed those instructors not meeting expectation. It is well recognized in the personnel economics and general management literatures that differential recruitment and retention of more productive employees can be at least as important as performance gains attributed to the motivational response among workers (Podgursky and Springer, 2007).

III. Trends in Compensation Reform

Given the efficiency costs of rigid salary schedules and growing pressure on K-12 schools to raise student performance, interest in market and performance-related pay reforms is

growing. Several states and districts have implemented incentives to encourage experienced teachers to teach in low performing schools (Prince, 2003, 2007). Florida, Minnesota, and Texas have allocated over $550 million to incentive pay programs that reward teacher performance. Moreover, funding for the federally sponsored Teacher Incentive Fund (TIF) quadrupled in 2010, and the Obama Administration's 2011 budget request designated an additional $950 million for a new Teacher and Leader Innovation Fund that would support the development and implementation of performance-oriented approaches to recruiting, retaining, and rewarding highly effective teachers.

Perhaps more important than the direct allocation of dollars, current education reform efforts, including the Race to the Top grant competition, have focused heavily on market and performance-related pay programs. In some states, in order to compete for a piece of the coveted $4.35 billion earmarked for Race to the Top, state legislators met in special sessions to remove institutional barriers to judging teacher performance, retaining and rewarding their most effective practitioners, and counseling out the lowest performers. In fact, the largest portion of the 500-point Race to the Top rubric for grading state applications was performance-related pay (United States Department of Education 2009).

Despite the considerable financial investment in compensation reform, we do not have much "microeconomic" data on the actual design components of these programs or information about the experiences of school systems implementing various pay reform models. State data systems generally do not capture details on programs. Additionally, even in the few states that maintain more detailed records on teacher compensation payments, the policies and procedures governing data collected by and received from local education agencies typically lump incentive payments with several other expense categories.

The best data currently available on national levels and trends comes from various waves of the Schools and Staffing Survey (SASS), which is fielded every four to five years by the U.S. Department of Education. The SASS collects information from a large nationally representative sample of roughly 8,000 public schools and 43,000 public school teachers.[1] At the same time, even though the SASS covers two decades of public school experience and has included various questions about market and performance-related pay, many of the compensation-specific survey questions are longitudinally inconsistent.[2] Thus, we focus attention on data in the most recent waves of the survey, which have maintained some consistency.

Table 2 displays estimates from a series of items that asked school district administrators whether they provided pay bonuses or other rewards for certain teacher characteristics or behaviors.[3] Considering the highly skewed distribution of school districts we have chosen to report these statistics two ways: (a) as a percent of all school districts and (b) as a percent of all teachers. In the 2007-08 school year, 24.5 percent of teachers offered a bonus to teachers with NBPTS certification. However, these districts were well above average in size. As a consequence, the teacher-weighted estimates indicate that 48.5 percent of all public school teachers in the U.S. were exposed to a program that offered a bonus if they earned NBPTS certification.

Insert Table 2 Here

NBPTS certification is also the most rapidly growing form of incentive pay, with the teacher exposure rates growing by 26 percentage points between the 1999-2000 and 2007-08

[1] SASS includes private schools and teachers as well. However, the focus of this study is on trends in public schools.

[2] There have been six waves of SASS, associated with six school years: 1987-88, 1990-91, 1994-95, 1999-00, 2003-04, 2007-08.

[3] "Does the district currently use any pay incentives such as a cash bonuses, salary increase, or different steps on a salary schedule to reward …?"

school years. Incentive payments for excellence in teaching, teaching in a less desirable location, or teaching in a shortage field are less popular, but have also grown considerably over this period. Interestingly, as of the 2007-08 school year, nearly one-third of all public school teachers were employed in a school district that offered some type of incentive for teaching in a shortage field.

The middle rows of Table 2 reports estimates on the number of incentives provided to teachers excluding in-service professional development. As of the 2007-08 school year, only five percent of teachers were employed in public school districts where the district administrator responding to the survey reported all four incentives being present. The results also indicate that more than one-third of all teachers (36.1 percent) were employed in a school district that did not offer any incentives for earning NBPTS certification, excellence in teaching, teaching in a less desirable location, or teaching in fields of shortage. At the same time, the share of districts reporting none of these incentives has dropped by 17 percentage points between the 1999-2000 and 2007-08 school years.

The bottom half of Table 2 displays summary statistics for a series of questions concerning group-based incentive programs. Of most interest in the context of this paper is the question about teachers being awarded cash bonuses and/or additional resources if their school was recognized based on student achievement. Approximately five percent of all public school districts offered this type of incentive payment during the 2003-04 school year, which accounts for 15.4 percent of all public school teachers. Unfortunately, we cannot examine whether the incidence of these group incentive programs changed over time since this battery of questions only appeared in the 2003-04 school year.

The two most recent waves of the SASS asked school district administrators about various methods used in their district to recruit teachers. As displayed in Table 3, loan forgiveness programs were the most prevalent strategy used by districts to recruit teachers during the 2003-04 school year, though we cannot study the trend over time because the questions were inconsistent from the 2003-04 to the 2007-08 administration. Roughly seven percent of school districts employing 16 percent of all public school teachers used signing bonuses to recruit teachers during the 2007-08 school year. While the number of districts offering signing bonuses increased moderately over time, there was a slight decrease in the percentage of teachers employed in those districts, which aligns with the growing number of rural districts exploring financial incentives to fill vacancies.[4]

Insert Table 3 Here

While all of the SASS surveys included questions about market and performance-related pay, only a few of the questions were consistently asked from one survey administration to the next. One block of questions that was nearly identical concerned teacher recruitment bonuses by field. School district administrators were asked whether their district currently offered pay incentives to recruit or retain teachers to teach in shortage fields and, if so, to identify the fields in which incentive pay was used. Unfortunately, this set of questions was not included in the district survey administered during the 2007-08 school year.

Table 4 displays summary statistics on prevalence of rewards to recruit and/or retain teachers in fields of shortage. There is a sharp increase over the 16 year interval in the incidence of field-based incentive pay. In the 1987-88 school year, only 7.5 percent of districts (11.3

[4] Table 3 also indicates slightly more than three percent of school districts provided relocation assistance to new hires. Districts offering finder's fees to existing staff for new teacher referral were reported in less than 2 percent of school districts.

percent of teachers) provided such incentives.[5] That share climbed to 12 percent of districts employing 25 percent of teachers by the 2003-04 school year. Consistent with the recruitment difficulty responses seen in Table 3, these incentives are most commonly used in the areas of special education, mathematics, science, and English as a second language.

Insert Table 4 Here

The SASS data support the incidence of market and performance-related pay reform is increasing. However, one significant limitation of SASS is that respondents are never asked about the size of bonuses, either in dollars or as a percent of salary. Ballou (2001) uses these SASS data to compare bonus pay in public versus private schools. His analysis of teacher salary data suggest that the size of bonuses in public schools is small, both as a share of salary and in comparison to private schools. In addition to Ballou, some other researchers have analyzed factors associated with the use of performance pay in public schools. Goldhaber, Hyung, DeArmond, and Player (2005) examine factors associated with the incidence of performance pay in public schools. Podgursky (2007) shows charter schools are much more likely to make use of market or performance-related pay incentives.

IV. Evaluations of Market and Performance-Related Pay Programs

This section of the paper reviews recent evaluation studies assessing the impact of market and performance-related pay programs on student achievement and teacher outcomes, which is surprisingly thin considering the number of schools, districts, and states implementing teacher compensation reform. We focus on evaluations of performance-pay programs using experimental designs or those relying on a regression discontinuity framework. When

[5] Note that these recruitment incentives can take the form of cash bonuses, higher pay, or higher initial placement on the salary schedule. The latter is more subtle, and thus less controversial, than explicit bonuses or differentiated pay structures.

implemented properly, such designs are ideal for assessing whether an intervention truly produces changes in outcomes under study or whether observed changes in outcomes are simply artifacts of pretreatment differences between two or more groups under study. While we also include RD studies in our review, it is important to note they generate highly localized estimates of a treatment effect because they are reliant on a subset of observations immediately above and below a cutoff point.

4.1. Evaluations of Performance-Related Pay Programs – International Evidence

Table 5 summarizes key design components of the performance-pay program that were evaluated as well as the study period, sample size, dependent variable(s), and main findings. All of these studies were implemented abroad. Most report generally positive effects on student achievement, though it is less clear whether these programs actually promoted long-run learning as some studies find the effects do not persist from one year to the next or document opportunistic behaviors on the part of treatment teachers. The incentive structure facing teachers and schools in several of the studies (e.g., Andhra Pradesh, India or rural Kenya) are also much different from the operational context found within the U.S. public elementary and secondary school system.

Insert Table 5 Here

Muralidharan and Sundararaman (2008) studied the impact of two output-based incentive systems (an individual teacher incentive program and a group-level teacher incentive program) and two input-based resource interventions (one providing an extra-paraprofessional teacher and another providing block grants). In what was known as the Andhra Pradesh Randomized Evaluation Study (AP RESt), 500 rural schools in Andhra Pradesh, India, were randomly selected to participate and then assigned to one of the four treatment conditions or to the control

group. These schools had a weak incentive structure for teachers, with 90 percent of noncapital education spending going to regular teacher salary and benefits. The AP RESt intervention was developed in partnership with the government of Andhra Pradesh, a large nonprofit organization interested in education issues in India (the Azim Premji Foundation), and the World Bank.

The individual incentive program awarded bonus payments to teachers for every percentage point of improvement above five percentage points in their students' average test score. All recipients received the same bonus for every percentage point of improvement. The bonus award scheme was structured as a fixed performance standard, which means that awards were distributed to any teacher or school that was selected to be in the AP RESt intervention and that exceeded the performance threshold.

Muralidharan and Sundararaman (2008) reported that student test scores on high-stakes tests increased between 0.12 and 0.19 standard deviations in the first year of the program and between 0.16 and 0.27 standard deviations in the second. Students enrolled in classrooms presided over by teachers eligible to receive a bonus award scored 0.11 to 0.18 standard deviations higher on low-stakes tests than those students whose teachers were not eligible to earn a bonus award. Students in treatment-condition classrooms also scored higher on a separate test that assessed high-order thinking which Muralidharan and Sundararaman (2008) indicate represents "genuine improvements" in learning, as opposed to better test-taking skills or perhaps other strategies employed by teachers to increase their chances of receiving a bonus award.

Muralidharan and Sundararaman (2008) also found that the schools assigned to the output-based intervention (i.e., individual- or group-incentive conditions) outperformed those schools assigned to the input-based resource interventions (i.e., paraprofessional or block grant conditions). Students enrolled in a classroom instructed by a teacher selected for the group

incentive intervention also outperformed students in control-condition classrooms on the mathematics and language tests (0.28 and 0.16 standard deviations, respectively). At the same time, students enrolled in schools assigned to the individual incentive condition outperformed students in both the group incentive condition and the control condition following the second year of implementation.

Another interesting feature of the AP RESt study is that external evaluators collected data on intermediate outcomes in interviews and through classroom observation. Teacher interviews offered anecdotal evidence that teachers in the individual or group incentive intervention were more likely to assign homework, offer support outside of class time, have students complete practice tests, and focus attention on low-performing students. However, Muralidharan and Sundararaman (2008), using data collected by the observational protocol, found no significant differences between treatment- and control-condition classrooms.

Glewwe, Ilias, and Kremer (2008) studied the impact of the International Child Support Incentive Program (ICSIP), a group incentive intervention that randomly assigned 100 schools in rural Kenya to either a treatment or a control condition. ICSIP's bonus scheme was structured as a rank-ordered tournament, and prizes ranged between 21 percent and 43 percent of average monthly base salary.[6] The ICSIP appraised school performance on the basis of student drop-out rates and test scores, with the twelve highest-performing and the twelve most-improved schools that were assigned to the ICSIP intervention receiving a prize.

Glewwe et al. (2008) found that students enrolled in schools participating in the ICSIP intervention had noticeably higher scores on high-stakes tests than students enrolled in schools

[6] Unlike other incentive programs discussed in this section of the paper, ICSIP awarded teachers with prizes rather than cash bonuses. As noted by Glewwe, Ilias, and Kremer (2008), the ICSIP awarded prizes such as a suit worth about $50, plates, glasses and cutlery worth about $40, a tea set worth about $30, and bed linens and blankets worth about $25.

assigned to the control condition. However, when comparing the performance of students enrolled in control- and treatment-group schools on a low-stakes test, Glewwe et al. (2008) found no differences in student test scores. It appeared that students enrolled in schools participating in the ICSIP intervention were coached in test-taking skills; an analysis of item-level test data revealed, for example, that treatment-condition students were significantly less likely to leave a test question blank.

Glewwe et al. (2008) also examined the impact of the ICSIP on teacher behavior. The authors found no differences in teacher attendance or pedagogy (behavior in classroom, instructional practices, number of homework assignments) among teachers in schools assigned to the ICSIP intervention and those working in a control-condition school. At the same time, teachers working in schools eligible for an ICSIP prize were 7.4 percentage points more likely to offer test-preparation sessions for students outside of normal school hours (typically when students were on vacation). In total, Glewwe et al. (2008) question the probability of the ICSIP program's improving long-run education outcomes, given the current state of schooling in the Busia and Teso districts of western Kenya.

Unlike the above-mentioned controlled trials, in which teachers or schools were randomly assigned to research groups, the next several studies exploited the fact that teachers or schools assigned to intervention and control-group conditions differ solely with respect to a cutoff point along some pre-intervention assignment variable. When implemented properly, an RD design allows for unbiased comparison of average treatment effect on teachers or schools that fall just to the right or to the left of such selection cutoffs.[7] The remainder of this subsection presents an overview of major findings from three RD studies of education incentive

[7] For a discussion of RD designs, see Thistlewaite and Campbell (1960); Hahn, Todd, and van der Klaauw (2001); and Lee and Lemieux (2009).

interventions: two programs implemented in Israel and a program operating in Mexico since 1992.

Lavy (2002) evaluated a group incentive program that was implemented in sixty-two Israeli high schools and designed to reduce student drop-out rates and improve student achievement. The program rewarded school performance on the basis of three factors: mean test scores, mean number of credit hours, and school drop-out rate. The bonus scheme was designed as a rank-ordered tournament, with the schools in the top third of performers competing for $1.44 million in awards. Schools earning a bonus had to distribute to their teachers 75 percent of the school-level award funds in amounts proportional to their gross annual compensation, regardless of their performance during the school year; the remaining 25 percent was to be used for improving school facilities for teachers. Lavy (2002) reported that top-performing schools received between $13,000 and $105,000 during the first year of implementation, with teacher bonuses ranging from $250 to $1,000 per teacher.

Lavy (2002) found a positive and statistically significant effect on student outcomes. Following the second year of implementation, for example, the group incentive program was found to have had a positive effect on average credit hours earned, average science credits earned, average test scores, and proportion of students taking Israel's matriculation test. Estimates further indicated that the program affected particular groups of students more than others—for instance, students at the low end of the ability distribution performed much better than expected on Israel's exit tests.

Lavy (2002) also compared the effectiveness of Israel's group incentive intervention with an input-based intervention that had been implemented several years earlier. The input-based intervention provided twenty-two secondary schools with additional resources to implement

professional training programs, reduce class size, and offer tutoring to below-average students. Although both programs improved student outcomes, Lavy (2002) concluded that the group incentive program is more cost-effective per marginal dollar spent. Muralidharan and Sundararaman (2008) similarly found that both the individual and group incentive programs were more cost-effective than either the "extra-paraprofessional" teacher or block-grant treatment conditions. The relative effectiveness of these interventions is particularly relevant to U.S. education policy because input-based reforms generally have been implemented more widely than output-based interventions such as New York City's SPBP.[8]

Lavy (2008) studied an individual incentive program in Israel that awarded bonuses to high school teachers in grades ten, eleven, and twelve based on their students' performance on national exit tests. The program was structured as a rank-ordered tournament and operated for a single semester (January–June 2001). Teachers in the intervention could earn a bonus for each class of students they prepared for the national exit tests, with awards ranging from $1,750 to $7,500 per class prepared. As reported by Lavy (2008), of the 302 teachers (48 percent of eligible teachers) awarded a bonus following the June 2001 exit tests, sixteen won bonuses for two of their classes.

Lavy (2008) creatively exploited two subtle features of the pay-for-performance program—measurement error in the assignment variable and a break along the pre-intervention assignment variable—to estimate the causal impact of the incentive program by using regression discontinuity design. Estimates of the net intervention effect indicated that the number of exit-exam credits earned by students instructed by a teacher in the incentive program increased by 18

[8] Hanushek (2003) provides a critical review of evidence on input-based schooling policies in the United States and abroad.

percent in mathematics and 17 percent in English, while data from a survey of teacher attitudes and behaviors suggested positive changes in teaching practices, teacher effort, and instruction tailored to low-performing students. When investigating gaps in performance between the results of school tests and national tests taken by students enrolled in treatment and comparison schools, Lavy (2008) did not find evidence of opportunistic behavior or negative spillover effects.

Santibanez et al. (2008) used a RD design to estimate the impact of Mexico's Carrera Magisterial (CM) on student test scores. Implemented in 1992, CM is a teacher incentive program that was designed collaboratively by state and federal education departments and the national teachers' union. Teachers participating in the program can earn a financial bonus if they accumulate enough points on a variety of measures defined by CM guidelines, including input criteria such as years of experience, highest degree held, and professional development activities, as well as output criteria such as their performance on a subject-matter knowledge test and their students' test scores (Santibanez et al., 2008). Awards ranged from 24.5 to 197 percent of a teacher's annual earnings (McEwan and Santibanez, 2005; Ortiz-Jiminez, 2003).

Santibanez et al. (2008) take advantage of the financial incentive that individual teachers have to improve their students' test performance. Since the program appraises teachers on most performance measures before students take the high-stakes tests each school year, teachers participating in the CM program have a general sense of how many additional points they need to earn on the strength of their students' performance on the high-stakes test to receive an award. Santibanez et al. (2008) detected a negligible impact on test scores of students enrolled in elementary school classrooms taught by teachers facing a strong incentive, while they detected small, positive effects at the secondary level. The authors note that their identification strategy

relies on a factor in the CM program that may be worth too few points to motivate teachers to exert more effort to improve student test scores.

4.2. Evaluations of Performance-Related Pay Programs –Evidence from U.S.

Table 6 summarizes key design components of the performance-pay program that were evaluated as well as the study period, sample size, dependent variable(s), and main findings. All of these studies were implemented in the U.S. public school system. Table 6 further indicates that all of the studies employing a random assignment study design are still being evaluated. In fact, many of these studies are still being implemented and either short-run findings or no findings are available at this time.

In August 2006, the National Center on Performance Incentives (NCPI) implemented the Project on Incentives in Teaching (POINT) intervention in the Metropolitan Nashville Public Schools (MNPS) system.[9] The POINT experiment recruited 297 teachers of middle-school mathematics in grades five through eight and randomly assigned these teachers to the treatment or control condition. Teachers assigned to the intervention are eligible to receive bonuses of up to $15,000 per year for a three-year period on the basis of two factors: the progress of a teacher's math students over a year, as measured by their gains on the Tennessee Comprehensive Assessment Program (TCAP); and the progress of a teacher's nonmath students over a year, as measured by their gains on the TCAP as well.

The POINT experiment is designed as an individual incentive intervention in which performance is judged according to a fixed performance standard. Because this standard was

[9] The NCPI, a state and local policy research and development center funded by the U.S. Department of Education's Institute of Education Sciences, was established in 2006 to conduct independent and scientific studies on the individual and institutional effects of pay-for-performance programs and other incentive policies. The NCPI is located at Vanderbilt University's Peabody College and core institutional partners include the RAND Corporation and the University of Missouri – Columbia. More information can be found at www.performanceincentives.org.

determined at the beginning of the POINT experiment and will remain fixed for three years, all teachers have the opportunity to be rewarded for having improved over time. The experiment concludes following the 2008–09 school year, and preliminary results will be available sometime during the following year.

In October 2008, the NCPI implemented a demonstration project to study a group incentive intervention. Eighty-two grade-level teams of teachers in grades six, seven, or eight were randomly assigned to either the treatment or control conditions. A team is defined as a group of academic teachers who meet regularly to discuss a common set of students, performance goals, and outcomes for which they are collectively accountable. Teachers assigned to the incentive intervention are eligible to receive an award if their team is selected as one of the four highest-performing teams at their grade level, as measured by standardized achievement scores in reading, mathematics, science, and social studies. Treatment teachers are projected to earn a bonus of about $6,000 if their team qualifies for an award.

Glazerman et al. (2007) designed and implemented an impact evaluation of the Teacher Advancement Program (TAP), a program being implemented by the Chicago Public Schools using a federal Teacher Incentive Fund grant. The TAP is a comprehensive school-reform model consisting of four elements: (1) multiple career paths; (2) ongoing, applied professional growth; (3) instructionally focused accountability; and (4) performance-based compensation.[10] At the beginning of the 2007–08 school year, Glazerman and colleagues randomly assigned eight schools to receive the TAP intervention and eight schools to the control condition. The latter set

[10] More information on the TAP can be found at www.talentedteachers.org. For a recent, non-experimental evaluation of the TAP see Springer, Ballou, and Peng (2008). The Center for Teacher Compensation reform also provides an overview of a related program in Chicago's Public Schools (http://www.cecr.ed.gov/initiatives/profiles/pdfs/Chicago.pdf).

of schools delayed implementation of TAP for a two-year period while serving as controls. Another sixteen schools were then recruited and randomly assigned to the TAP intervention or control conditions for the 2009-10 and 2010-11 school years

New York City's School-Wide Performance Bonus Program (SPBP) was implemented midway into the 2007–08 school year. The SPBP was designed to provide financial rewards to teachers in schools serving disadvantaged students. The program sets expected incentive payments as a fixed performance standard, meaning that schools participating in the program are not competing against one another for a fixed sum of money. All participating schools can earn bonus awards of up to $3,000 per full-time union member working at the school if the school meets predetermined performance targets defined by the NYCDOE's accountability program, with the idea that this sum will be used to award bonuses to teachers and staff found to be deserving. The SPBP rules further mandate that schools participating in the program establish a four-person site-based compensation committee to determine how bonus awards will be distributed to school personnel.

Springer and Winters (2009) examined the impact of the SPBP on student outcomes and the school learning environment. Their sample included 186 SPBP-eligible elementary, K–8, and middle schools and 137 control-condition schools in New York City over a two-year period. Overall, they find that the SPBP had little impact on student proficiency or school environment it its first year. However, it is important to remember the short-run results reported in this study provide only very limited evidence of the SPBP's effectiveness. An evaluation of the program's impact after two years should provide more meaningful information about the impact of the SPBP.

4.3. Evaluations of Market-Oriented Pay Reforms

Our review of evaluations of market-oriented reforms adopted a more lenient evidence standard as we are aware of only one study using randomized controlled design.

V. Conclusion

Human resource (HR) policy – the recruitment, retention, and motivation of employees -- is increasingly recognized as a critical variable in the success of an organization. An integrated and coherent compensation policy is the central core of an efficient HR policy. In private and many public organizations, the compensation package is considered as a strategic whole, and carefully designed to get the most HR return per dollar of compensation. In public K-12 by contrast, the compensation "system" is fragmented and uncoordinated, each piece perhaps responding to pressures from a particular constituency or inherited from an earlier contracts, but without systematic assessment of the logic or incentive effects of the whole.

Accountability pressures are forcing school districts to address the inefficiencies built into this compensation system, and rethink how they are spending roughly $250 billion annually for compensation of instructional personnel. Federal programs such as the Teacher Incentive Fund are encouraging states to experiment with market and performance-related pay reforms. States such as Minnesota, Florida, and Texas have developed programs to encourage their districts to develop such programs. A number of large urban districts, most notably Denver, have taken important steps in this direction. Market and performance-related pay incentives are much more common in charter schools and are expanding with the charter school base (Podgursky and Springer, 2007). Less movement has occurred in the area of teacher pensions, however, large unfunded liabilities for pensions and retiree benefits are likely to force reforms in this area as well. States and districts contemplating change should consider running pilot

programs or phasing in changes (or better yet, some type of randomized pilot study) so as to permit assessment of the effectiveness of such reforms.

Taken as a whole, pay for performance is now poised to become more reality than simple rhetoric. That said, moving forward deliberately and purposefully will be important. In particular, research must play a critical role. To date, the little research surrounding pay for performance paints a mixed picture (see Table 2), with many of the most rigorous studies still under way or having been conducted abroad.

Policy makers and education stakeholders at all levels would benefit from more unbiased assessments of teacher compensation reform programs and policies, as well as the effect of their various design components. For instance, should individual teachers or teams of teachers be rewarded, or perhaps a combination of both? Should the measure be based on student growth or attainment? What criteria should be included? Should it be based strictly on student test scores, or should other measures, like principal evaluations, be included? If other measures are included, what should be the weight of each element? Since the design of an incentive program can lead to dramatic effects on students, teachers, and administrators, we must take the lessons learned from these evaluations and continue to both evaluate and refine programs to maximize their effectiveness.

In addition to ongoing research and evaluation, data systems need continued development. Even though these systems at the state- and district-level have helped drive policy innovations around teacher compensation, they haven't been designed to inform high-stake personnel decisions. Among the weaknesses are seemingly mundane errors, such as inaccurate course codes, to more significant errors, such as assigning the same unique identifier to multiple students. Many structural errors can also be found, including systems that don't account for

student mobility (both within and between schools) and too few data snapshots to accurately capture what's happening in schools and classrooms. When these systems are used to reward teachers, it's imperative they accurately capture what's happening in classrooms.

Related to the development of data systems, there needs to be a continued push for more accurate and reliable tools for assessing individual and team performance. As many teachers will attest, far more happens in a classroom than can be measured on a standardized assessment. And roughly 70% of teachers don't instruct a course or a grade covered by a standardized assessment. Research is ongoing in this area and must be included in the design of pay for performance systems.

Balancing the interests of all stakeholders is crucial when designing and developing pay for performance systems. While stakeholder engagement and buy-in is critically important, the interests and preferences of individual actors may water down the power of the incentive. For example, a study of a pilot incentive pay program in Texas, in which multiple stakeholders were engaged in plan development, found that many schools chose to distribute relatively small awards across all school personnel, regardless of individual performance. The relatively weak incentive system didn't appear to induce any significant changes in teacher productivity (Taylor and Springer 2009).

Finally, we must stay focused on the bigger picture of schooling. Even if pay for performance programs and policies ultimately play an important role in reforming K-12 public schools, these reforms can't be implemented in a vacuum. Teacher pay alone will not improve the quality of teaching and, by extension, improve levels of student learning. Compensation reform is just one element to be implemented alongside reforms that retool resource allocation

and deployment norms; teacher hiring, tenure, and dismissal practices; and the standards and assessments systems, among other areas.

Table 1. Salary Schedule for Houston Independent School District, 2007-08 School Year

10 Month Teacher – 2008-2009 School Year

10-M Bachelors			10-M Masters			10-M Doctorate		
Step	Experience	Pay Rate	Step	Experience	Pay Rate	Step	Experience	Pay Rate
1	0-1	$44,027	1	0-1	$45,057	1	0-1	$46,087
2	2	$44,477	2	2	$45,507	2	2	$46,537
3	3	$44,992	3	3	$46,022	3	3	$47,052
4	4	$45,507	4	4	$46,537	4	4	$47,567
5	5	$46,022	5	5	$47,052	5	5	$48,082
6	6-8	$47,476	6	6-8	$48,506	6	6-8	$49,536
7	9-10	$48,801	7	9-10	$49,831	7	9-10	$50,861
8	11	$49,251	8	11	$50,281	8	11	$51,311
9	12	$49,814	9	12	$50,844	9	12	$51,870
10	13-14	$51,838	10	13	$52,868	10	13	$53,318
11	15-17	$53,601	11	14	$53,368	11	14	$53,818
12	18-19	$55,138	12	15-16	$55,138	12	15-16	$56,791
13	20-22	$56,791	13	17-19	$56,791	13	17	$57,241
14	23-25	$58,444	14	20-21	$58,444	14	18-19	$60,091
15	26-27	$60,091	15	22-23	$60,091	15	20-21	$61,741
16	28-29	$61,741	16	24-25	$61,741	16	22-23	$63,395
17	30+	$65,222	17	26-27	$63,395	17	24-25	$65,045
			18	28	$63,845	18	26	$65,495
			19	29+	$68,590	19	27+	$71,960

Source: http://www.nctq.org/salary_schedule/32-08.pdf

Table 2: Incidence of Various Incentive Payment

	District-Weighted(%)				Teacher-Weighted(%)			
District Rewards Following:	1999-2000	2003-2004	2007-2008	Change	1999-2000	2003-2004	2007-2008	Change
NBPTS	8.3	18.4	24.5	16.2	22.9	39.8	48.5	25.6
Excellence in Teaching	5.5	8.0	10.3	4.8	13.6	14.0	15.3	1.7
Teach in less desirable location	3.6	4.6	5.7	2.1	11.2	13.1	16.1	4.9
Teach in fields of shortage	10.4	11.9	15.4	5.0	23.6	25.3	32.2	8.6
In-service professional development	26.4	24.2	---	---	38.8	35.9	---	---
Number of incentives (excluding in-serv. PD)								
None	78.02	68.7	61.0	-17.0	56.46	41.2	36.1	-20.4
1 incentive	17.13	23.4	27.0	9.9	26.92	36.3	35.3	8.4
2 incentives	4.03	6.2	8.0	4.0	9.8	14.4	14.5	4.7
3 incentives	0.67	1.5	3.1	2.4	2.52	6.8	8.8	6.2
4 incentives	0.15	0.3	0.9	0.7	4.3	1.3	5.4	1.1
Based on student achievement, were any schools in the district rewarded in any of the following ways?								
Cash bonus/add resources for school-wide activity	---	6.8	---		---	19.6	---	---
Cash bonus/add resources for teachers	---	4.7	---		---	15.4	---	---
Schools given non-monetary forms of recognition	---	15.8	---		---	30.4	---	---
District Has Salary Schedule for Teachers								

Source: Schools and Staffing Surveys, various years. School District Survey

Table 3: District Use of Various Recruitment Incentives

	District-Weighted(%)			Teacher-Weighted(%)		
Methods district uses to recruit teachers:	2003-04	2007-08	Change	2003-04	2007-08	Change
Signing Bonuses	4.8	6.8	2.1	17.4	15.9	-1.5
Student Loan Forgiveness	8.1	---	---	19.0	---	---
Forgiveness of student loan(s) funded by district	---	2.3	---	---	5.1	---
Relocation assistance	3.2	3.6	0.4	8.5	9.2	0.7
Finder's fee to existing staff for new teacher referrals	0.9	1.6	0.6	2.3	2.5	0.2
Available training to staff members to teach for current or anticipated shortage	---	30.7	---	---	38.3	---

Source: Schools and Staffing Surveys, various years

Table 4

Reward to recruit/retain teachers in fields of shortage						change
	1987-88	1990-91	1993-94	1999-00	2003-04	1987-88 to 2003-04
District provides incentive	11.3 %	16.6 %	18.7 %	23.6 %	25.3 %	14.0 %
Elementary	---	---	---	2.4	2.6	---
Special ed	6.7	11.8	13.4	14.3	20.6	13.9
English/language arts	---	---	---	5.3	4.2	---
Social studies	---	---	---	1.6	2.4	
Computer sci	1.4	2.9	1.3	3.4	3.4	2.0
Mathematics	5.2	5.8	3.9	8.9	15.7	10.5
Physical Sciences	3.6	5.0	3.9	8.4	13.4	9.8
Biological sci	3.8	4.3	3.7	8.4	12.8	8.9
English as Second Lang	3.3	7.6	8.1	11.1	15.5	12.2
Foreign lang	2.4	3.1	2.4	5.3	9.4	7.0
Music or art	---	---	---	4.9	6.4	---
Vocational or technical educ/	---	4.7	3.2	8.0	7.3	---
Other fields	4.2	4.2	1.6	---	---	---

Recruitment Incentives by Teaching Field*

* "Does this district currently use any pay incentives to recruit or retain teachers to teach in fields of shortage?"

Source: Schools and Staffing Surveys, various years. School District surveys.

Table 5. Summary of Experimental and Quasi-Experimental Evaluations of Teacher Pay for Performance Programs

Program	Study Design	Study Period	Sample	Performance Measures		Results
				Unit of Accountability	*Measures of Teacher Performance*	
International						
Kenya's International Christelijk Steuenfonds Incentive Program	RCT	1998 - 1999	100 primary schools; 1,000+ teachers; 50,842 students.	Group (school)	Student test score gains and student achievement levels	Modest, positive effect for high-stakes assessment. No effect on low-stakes assessment.
Andra Pradesh, India's Randomized Evaluation Project	RCT	2006 - 2008	300 schools and 68,000+ student observations.	Individual and Group (school)	Student test score gains	Modest, positive effect on high-stakes assessment (approx. 0.12 to 0.19 standard deviations after year one and 0.16 to 0.19 standard deviations after year two).
Israel's Ministry of Education's School Performance Program	RD	1994 - 1997	62 schools (37 non-religious, 18 religious and 7 Arab schools).	Group (school)	Number of credit units per student, student receiving a matriculation certification, and school dropout rate.	Modest, positive effect for average credit hours earned, average science credits earned, average test score, and proportion of students taking Israel's matriculation exam.
Israeli Teacher-Incentive Experiment	RD	2001	4,109 students and 27 schools.	Individual	Student achievement levels	Modest, positive effect for number of exit exam credits earned in mathematics (increased 18 percent) and in reading (increased 17 percent).

Mexico's Carrera Magisterial	RD	1998 - 2003	850,000+ classroom-year observations. 810 primary school teachers; 209 secondary school teachers.	Individual	Educational degrees, years of experience, professional development, principal ratings, content knowledge mastery, student performance on standardized tests.	No effect for primary school teachers; Modest, positive effect for secondary school teachers (approx. 3 to 15 percent of standard deviation).
	RD	2000 - 2002	76,567 teachers and 27,123 schools.	Individual	Educational degrees, years of experience, professional development, principal ratings, content knowledge mastery, student performance on standardized tests.	Small, positive effects (<10 percent of standard deviation).

Table 6. Summary of Experimental and Quasi-Experimental Evaluations of Teacher Pay for Performance Programs

Program	Study Design	Study Period	Sample	Performance Measures		Results
				Unit of Accountability	*Measures of Teacher Performance*	
United States						
Project on Incentives in Teaching (Nashville, TN)	RCT	2007 - 2009	147 treatment and 152 control teachers (grades 5 - 8).	Individual	Student test scores in mathematics, reading, social studies, and science.	In-progress
Project on Team-Level Incentives in Teaching (Round Rock, TX)	RCT	2009	41 treatment and 41 control group teams (grades 6 - 8).	Group (grade-level teams)	Student test scores in mathematics, reading, social studies, and science.	In-progress
Recognizing Excellence in Academic Leadership Program (Chicago, IL)	RCT	2008 - 2011	32 Teacher Advancement Program (TAP) schools.	Individual and Group (school)	Mentor review, self-review, master teacher review, administrator review, classroom observations, teacher developed portfolio, interviews, student test score gains, and overall school performance.	In-progress
School-Wide Performance Bonus Program (New York, NY)	RCT	2008 - 2009	191 treatment and 131 control group schools (elementary, middle and k-8). More than 100,000 in grades 3 through 8.	Group (school)	Student test score levels and gains, student, teacher, and principal perceptions of school environment, and external enumerators'' rating of school's instructional climate.	In-progress