January, 2012

Recruiting, Retaining, and Creating Quality Teachers

C. Kirabo Jackson, *Northwestern University*

Recruiting, Retaining, and Creating Quality Teachers

C. Kirabo Jackson
Northwestern University, IPR, and NBER
kirabo-jackson@northwestern.edu

Keywords: Teacher quality, education reform.

JEL classification numbers: J2 , J45, I20, I28.

Summary

This article synthesizes the research literature on how to ensure that the teaching workforce is effective. It offers three approaches to improving effectiveness: attract talented individuals into the profession, create incentives for exerting optimal effort, and provide professional development so that teachers have the skills to be effective. The research literature reveals that each approach can yield meaningfully improved student outcomes and that no one strategy is clearly more effective. The policy implication of these findings is that a multifaceted approach would improve teacher effectiveness and student outcomes. However, although there are examples of successful policies, there is no consensus on the most effective practices. The paper concludes that schools need to be data driven, flexible in their employment practices, and open to experimentation.

Policy makers in many nations are grappling with how to improve the quality of public schools, and there is consensus among policy makers, parents, educators, and researchers that teachers are one of the most important components of the school environment. As such, many education policy reforms have the expressed aim of improving the quality of teachers to whom students are exposed. For example, the No Child Left Behind Act of 2001 (arguably the most significant federal intervention into education in the United States over the last 40 years) required states to ensure that all teachers of the core academic subjects were "highly qualified" by the end of the 2005–06 school year. Similarly, one of the four specific areas that states must advance to receive federal grant funding under Race to the Top (a $4.35 billion U.S. Department of Education competition designed to spur reform in state and local district education) is "recruiting, developing, rewarding, and retaining effective teachers and principals." Similar policy pushes are under way in the United Kingdom, Australia, Mexico, Chile, and other nations.

Although the objective is clear, what is unclear is how to ensure that all teachers are highly qualified and how to go about "recruiting, developing, rewarding, and retaining effective teachers." There are several proposed policies aimed at improving teacher quality and several conflicting research findings based on different samples and different research designs speaking to what strategies may be most effective. As such, figuring out exactly what policies one should pursue to improve the quality of teachers is neither straightforward nor obvious. However, because some studies are more conclusive than others, and theory suggests that certain policies may be relatively more or less effective in certain contexts, the academic research does provide some guidance.

There are several approaches to improving teacher quality that can roughly be put into one of three categories. The three main approaches to ensuring that students are exposed to high-

quality teachers are motivated by three views of the world: In one view of the world, the primary drivers of differences in teaching performance are differences in the innate ability of teachers, which is determined before entering the profession. Under this view, the most effective policies will be those that attract high-quality prospective teachers to the profession and better retain those effective teachers already in the profession. In the second view of the world, differences in teaching performance are primarily due to differences in teacher effort. Under this view, the most effective policies will be those that provide incentives for teachers to exert more effort. In the third view, differences in teacher effectiveness are due primarily to differences in teaching skills or know-how. Under this scenario, the most effective policies will be those that facilitate the most rapid acquisition of teacher skill for the greatest number of teachers.

In this article, I call on the existing theoretical and empirical research to provide some guidance on what are likely to be the most fruitful kinds of policies for improving teacher quality. In section 1, I survey the evidence on the importance of teachers for student outcomes, and in sections 2, 3, and 4, I discuss the main economic forces at play for each of the preceding proposal types and survey the relevant empirical research in the area. Finally, section 5 presents an overview of all the evidence, discusses the specific need for further research, and highlights those policies that are most likely to succeed at improving teaching quality.

1. The importance of teachers

The notion that teacher quality is the most important component of school quality comes from evidence from value-added models that measure the improvements in test scores that students make from the beginning to the end of the year. In principle, because a student's test score at the beginning of the school year should reflect all that a student knows before entering

the classroom that year, and because a student's test score at the end of the year should reflect all that a student knows after being in the classroom with his or her teacher for the year, the *growth* in the student's test scores reflects the additional learning that took place that year.[1] By focusing on test score growth rather than simply comparing test score levels, one avoids penalizing those teachers who teach students with low levels of academic preparedness. With this in mind, researchers take the average growth in test scores for a given teacher as a measure of his or her *value added*, that is, how much learning took place among all students in the teacher's classroom in a given year.[2] These analyses have shown that some teachers are systematically associated with consistently high test-score growth (high value-added teachers), whereas others are systematically associated with consistently low test-score growth (low value-added teachers).

Using these value-added models, Rivkin et al. (2005), Rockoff (2004), and other researchers find that moving from an average elementary school teacher to an elementary school teacher at the 85th percentile of value added leads to between a 0.1 and 0.2 standard deviation improvement in student test scores. This is roughly equal to an improvement of between 4 and 8 percentile points in a single year. While the use of statistical models to measure teacher quality is not uncontroversial, Kane and Staiger (2008) find that these value-added measures, based on test-score gains, have tremendous power in predicting a teacher's *future* success in the classroom. Jacob and Lefgren (2008) find that they are correlated with principals' subjective evaluations of teachers, and Jackson and Bruegmann (2009) find that teachers with high estimated value added improve the test scores of their colleagues' students. To put the size of

[1] See Todd and Wolpin (2006) for a discussion of the specific scenario in which test scores at the beginning of the year are a sufficient statistic for the entire history of educational inputs.

[2] Note that most researchers do not merely take the average test-score growth for a given teacher but also make adjustments for differences in the levels of parental education, family income levels, student ethnic origins, and school characteristics when computing average test-score growth. In practice, the additional adjustments do improve the estimate somewhat but do not make a very large practical difference, as long as one focuses on test-score *growth*.

these estimates in context, Hanushek et al. (2005) and Jackson (2010c) find that the test-score improvements associated with being exposed to a great teacher versus an average teacher are larger than the differences associated with being in a great school versus an average school—so that a student would be better off at a bad school with a great teacher than at a great school with a bad teacher.[3]

While this evidence is compelling, there are some reasons why it may be premature to conclude that teachers are important in all contexts for all students and in all grades. The first is that most studies of teacher value added have focused on elementary school teachers so that it is not obvious that high school teachers have as large an effect on student outcomes as elementary school teachers. Two exceptions are Aaronson et al. (2007) and Koedel (2009), who find similar effects on math and reading achievement for high school teachers as for elementary schools teachers. However, Jackson (2011a) uses detailed information about student class, taking into account additional sources of bias,[4] and finds that while high school math teachers have important effects on math test scores, high school English teachers do not have any systematic effect on student English test scores, suggesting that while elementary school teachers and high

[3] Readers may wonder whether this example is contrived given that the best schools may also be those that attract the best teachers, as suggested in Jackson (2009). However, there is more variation in teacher quality within schools than there is variation in mean teacher quality across schools (Hanushek et al., 2005). Consistent with this, Sass et al. (2010) find that the average effectiveness of teachers in high-poverty schools is in general less than that of teachers in other schools, but only slightly and not in all comparisons.

[4] Jackson (2011a) points out that because elementary and secondary schools differ in important ways, methodologies designed for elementary school teachers may be inappropriate for measuring teacher quality in high school for two reasons. First, while much of the literature on teacher effects has focused on biases due to the nonrandom matching of students to teachers in elementary school settings (Rothstein, 2010; Koedel and Betts, 2009), in a high school setting, *even with random assignment to teachers*, if different teachers teach in different tracks, and students in different tracks are exposed to different treatments (e.g., students in the gifted and talented track receive extra college preparatory sessions or students in the remedial math class all have a bad social studies teacher), there will be bias due to *track treatment effects*; that is, in high schools, there is possible selection bias due to nonrandom placement of students to tracks and omitted variables bias due to track-specific treatments. This second source of bias creates additional challenges to identifying teacher effects in high school. The extant literature has not accounted for this second source of bias and may therefore be misleading about the importance of teachers in high school.

school math teachers are important for student test score outcomes, further research may be needed for understanding the importance of non-math teachers in high school.

The other important reason to question the broad importance of teachers is that test scores are an imperfect measure of student achievement so that while teachers are important for improving test scores, other outcomes may be more important for the overall education of the child. There is mounting evidence that the effects of high schools on test scores are not the best predictors of high school's effects on overall educational attainment (Booker et al., 2008; Deming et al., 2011). As such, the emphasis on teacher quality as measured by the ability to improve test scores alone might be unwarranted and too narrow.

In response to this second critique, there is a nascent literature showing that teachers have effects on outcomes other than test scores. Jackson (2011b) finds that English teachers in 9th grade have effects on attendance and discipline problems. Similarly, Koedel (2008) finds that a student's algebra teacher in high school affects the likelihood that a student will drop out of high school by between 4.2 and 14.1 percentage points and that those teachers associated with higher test-score growth are also those associated with reduced dropout. Also, Chetty et al. (2011) find that high value-added elementary school teachers increase their students' probability of college attendance, raise earnings, reduce teenage pregnancy rates, and improve the quality of the neighborhood in which their students live at age 25. Taken together, these studies indicate that both elementary and high school teachers have important effects on long-run outcomes and that estimated value added based on test scores is a reasonable measure of the importance of teachers for a student's long-term well-being.

Despite strong evidence that teachers make important contributions to student outcomes, there are few observable characteristics that are strongly associated with teacher effectiveness.

Clotfelter et al. (2007, 2010) find that a teacher's experience, test scores, and regular licensure all have positive effects on student achievement, with larger effects on math than on reading. They conclude that, taken together, the various teacher credentials exhibit quite large effects on math achievement. To get a sense of the magnitude of teacher quality that can be explained by observable teacher characteristics, Jackson and Bruegmann (2009) summarize all observable teacher characteristics (experience, licensure, score of license exams, certification, degree level) into a single index[5] and find that a 1 standard deviation increase in own-teacher value added due to *observable* characteristics is associated with a 3.6 and 2.6 percent of a standard deviation increase in math and reading test scores, respectively. These estimates are about one-third of the size of the variability in estimated value added overall, indicating that although observable teacher characteristics can explain some of the differences in teacher effectiveness, much of what makes a good teacher is still largely unknown. Recent evidence suggests that relating detailed preservice survey information from teachers and classroom observations with student achievement may allow us to better understand why some teachers are more effective in the classroom than others (Lavy, 2011; Rockoff and Speroni, 2011).

In sum, though researchers may quibble about the methods used to estimate teacher quality, and although results vary across samples, studies, and empirical designs, the empirical research indicates that teachers are quite determinative of student test score outcomes and also longer-term well-being. Even though most typically collected objective measures of teacher qualifications cannot explain the large differences in outcomes across teachers, it is clear that there would be large improvements in student outcomes if one could improve the quality of

[5] They predict student test score growth as a function of observable teacher characteristics and then take the value-added associated with each characteristic and compute the "predicted" value-added for each teacher based on their observed characteristics. This approach weights the characteristics by their importance in improving test scores.

teacher to which students are exposed. In sections 2, 3, and 4, I discus the three main types of strategies that may achieve this.

2. Attracting and retaining quality teachers

The first group of strategies aimed at improving teacher quality are predicated on the notion that a large part of the teacher quality differences we observe are due to differences in innate teaching ability. Under this view, because the variables that are typically collected *before* the time of hire for teachers do not allow one to identify who will be an effective teacher with much precision, being able to attract talented individuals into the teaching profession and retaining the most successful teachers will be effective ways to improve teacher quality. In this section, I discuss strategies to attract talented prospective teachers into the profession and also strategies to help retain the most effective teachers.

2.1 Attract talented prospective teachers into the teaching profession

Much recent interest in attracting and retaining quality teachers is driven by the observation that there has been a decline in the size and quality of the pool of available teachers; that is, both the fraction of college graduates going into teaching and the measured pre-service aptitudes of teachers have declined over time in many countries. Corcoran et al. (2004) use survey data of U.S. high school graduates spanning the years 1957–1992 and find that (1) while in 1964, more than half of working female college graduates were teachers, by 2000, this percentage had dropped to 15 percent, and (2) the likelihood that a woman from the top of her high school class would eventually enter teaching had fallen from almost 20 percent in 1964 to under 4 percent in 1992. Consistent with this, using data on female college graduates between 1961 and 1997, Hoxby and Leigh (2004) find that the share of teachers coming from the most

selective colleges has fallen over time. Similarly, in Sweden, Fredriksson and Öckert (2008) find that (1) in the cohort born in 1948, approximately 40 percent of university graduates earned teaching degrees, while for the cohort born in 1975, the corresponding figure was just 14 percent, and that (2) between the birth cohorts of 1948 and 1977, the average rank of teachers in ability tests conducted at age 13 had fallen by 10 percentile points. This decline in teacher aptitudes relative to other occupations is not confined to the United States and Sweden but has also been documented in the United Kingdom (Nickell and Quintini, 2002) and Australia (Leigh and Ryan, 2006).

This decline in teacher aptitude is only problematic insofar as it has deleterious effects on student outcomes. The empirical evidence does suggest that students have better outcomes with teachers who graduate from more selective colleges and have high scores on teacher license examinations (Brewer and Ehrenberg, 1994; Clotfelter et al., 2007; Jackson and Bruegmann, 2009). As such, teacher aptitude does matter. However, the decline in aptitude is also worrisome because it is indicative of a general decline in the quality (broadly construed) of students entering teaching. As such, while much of the empirical evidence is focused on aptitude as measured by achievement tests, the patterns observed likely relate to talent broadly construed (of which test scores and college selectivity are just two measures).

Teacher pay and teacher talent

In debates about teacher quality, the claim is always made that teacher pay is too low. What people really mean by this is that high-ability (talented) individuals make too much more outside of teaching for teaching to be an attractive occupational choice, and as a result, only those who would not earn high wages outside of teaching (the less talented) decide to enter the profession. It is reasonable to look to the reasons for the decline in the supply and aptitude of

teachers for guidance on how policy makers might better attract and retain high-aptitude individuals. While the timing of the declines and the exact policy or social changes vary from country to country, there are two explanations for these trends. Specifically, (1) average teacher pay (relative to other occupations) has declined over time and (2) the relative monetary benefits associated with ability have declined in teaching over time. Even though most discussions of teacher pay are focused on the low level of teacher pay on average, economic theory indicates that while the level of average teacher pay will be important for the number of individuals who decide to enter teaching, it is the dispersion of teacher pay that is likely to have a much larger effect on the aptitudes of prospective teachers.

It is helpful to lay out a stylized and simple model of occupation choice following Roy (1951) to illustrate the forces at play. Consider a world in which there are two occupations: teaching and nonteaching. Students go to college and then decide on their occupation. Each student has some known ability level (motivation, cognitive skills, social skills, scholastic preparation) that is associated with better performance and therefore higher wages in both occupations. As such, those who would be the most effective teachers are also those who would be the most effective at the nonteaching profession. The empirical evidence suggests that this is a reasonable description of the world. Specifically, both higher achievement test scores and graduating from a prestigious college are associated with higher wages outside of teaching and better classroom performance within the teaching profession. Also, Chingos and West (2010) examine earnings records for classroom teachers employed by Florida public schools and find that among teachers in grades 4–8 leaving for other industries, a 1 standard deviation increase in estimated value added to student achievement is associated with about 8 percent higher earnings outside of teaching; that is, teachers who increase student achievement by 4

percentile points more than the average teacher earn 8 percent more than the average teacher outside of teaching when they both leave the profession.

Suppose that wages in both occupations follow a bell curve–shaped distribution, and both teaching and nonteaching have the same median wage, but nonteaching has a more dispersed earnings distribution (i.e., the difference in earnings between the high-ability types and the low-ability types is larger in nonteaching than in teaching). The empirical literature indicates that this is a reasonable description of the world. Chingos and West (2010) find that while among teachers leaving for other industries, a 1 standard deviation increase in estimated value added to student achievement is associated with 6 to 9 percent higher earnings outside of teaching, *it is not associated with any higher earnings within teaching.* Similarly, Fredriksson and Öckert (2008) find that teacher earnings are less dispersed than in nonteaching professions and that the earnings increase associated with higher cognitive test scores (a measure of talent) is much smaller in teaching than in other professions. This is depicted in Figures 1 and 2.

In this stylized world, all students care about when choosing an occupation is how much they will earn so that they will chose the occupation that will grant them the highest earnings. A large body of empirical work suggests that there are several additional considerations that go into the decision to teach;[6] however, this model is meant to illustrate the economic forces at play rather than to be a perfect description of the real world. A key aspect of this model is that from an individual student's standpoint, what matters is what the student could earn in teaching versus in nonteaching rather than what the average teacher salary is relative to the average nonteaching salary; that is, a student who knows that he or she is high-ability will respond to the wages paid

[6] For example, Jackson (2009) finds that, holding salary constant, teachers change their school choices in response to changes in student demographics. Also, Ladd (2011) uses rich administrative data for North Carolina and documents that working conditions are highly predictive of teachers' intended movement away from their schools, independent of other school characteristics such as the racial mix of students.

in teaching compared to nonteaching for high-ability workers, whereas a student who knows that he or she is low-ability will respond to the wages paid in teaching compared to nonteaching for low-ability workers.

Consider a world in which an individual who lies in the *x*th percentile of the ability distribution will earn at the *x*th percentile of the earnings distribution of whichever occupation the individual chooses. A student who scores at the 10th percentile on some ability test is choosing between being at the bottom of a compressed teacher distribution (where all values are close to the median) and being at the bottom of a dispersed nonteacher distribution (where the 10th percentile lies far below the median). Even where the median wage for teachers and nonteachers is the same, low-ability students will be attracted to teaching because teaching offers higher wages for low-ability workers due to the compressed wage distribution. See Figure 3 (top) for an illustration of this, where median earnings are assumed to be $30,000 for both teaching and nonteaching.

By the same logic, a high-aptitude student who scores at the 90th percentile on the ability test is choosing between being at the top of a compressed teacher distribution (where all values are close to the median) and being at the top of a dispersed nonteacher distribution (where the 90th percentile lies far above the median). Even where the median wage for teachers and nonteachers is the same, high-ability students will be attracted to nonteaching because nonteaching offers higher wages for high-ability workers due to the dispersed wage distribution.

In this stylized world, even if median wages (and average wages) are the same in both occupations, because wages are more dispersed in nonteaching, high-ability types will choose to enter nonteaching, whereas low-ability types will choose to enter teaching. It is helpful to use this stylized framework to analyze two reforms to teacher pay: (1) an increase in the pay of all

teachers by $6,000 (about a 25 percent increase for the median-ability student in teaching) and (2) an increase in the pay of high-ability teachers and a decrease in the pay of low-ability teachers, while keeping pay for the median student in teaching the same.

What happens when all teachers receive an across-the-board increase?

While giving an increase to all teachers would certainly make teaching more attractive for all prospective teachers, it is not obvious that it will lead to a large increase in the ability level of teachers. To illustrate this point, consider the preceding setup, in which salaries follow a normal distribution. The median teacher and nonteacher salaries are each $30,000 per year, that is, the median-ability student will earn $30,000 in teaching or in nonteaching. In this hypothetical example, nonteacher salaries are more dispersed than teacher salaries such that the standard deviation of teacher salaries is $3,000, while that of nonteaching is $9,000. As such, in this hypothetical world, 95 percent of teacher salaries lie between $24,000 and $36,000, while 95 percent of nonteacher salaries lie between $12,000 and $48,000 per year. This is illustrated in Figure 3 (top). If one were to give all teachers a sizable pay increase of $6,000, it would shift the entire distribution of teacher earnings to the left by $6,000. This shift is illustrated in Figure 3 (bottom).

After the shift, the median-ability student who goes into teaching earns more than the median-ability person who goes into nonteaching by $6,000. In this example, after the increase, the median-ability student who teaches (and would now earn $36,000 in teaching) will make the same as the 75th percentile person in nonteaching so that the 75th percentile–ability student will choose to enter teaching over nonteaching after the $6,000 across-the-board increase in teacher pay. In fact, one has to go all the way up to the 82nd percentile of the ability distribution for the teacher salary and the nonteacher salary to be the same. As such, all ability levels below the 82nd

percentile would earn higher wages in teaching after the increase, and they would all choose to enter teaching; that is, now students with ability between the 50th and 82nd percentiles will choose to enter teaching after the across-the-board increase in teacher pay but would not have chosen teaching before. This example also illustrates that owing to the higher dispersion in nonteaching, despite the sizable increase in teacher pay, the highest-ability students will still not decide to enter teaching. Clearly, with a large enough increase in teacher pay (such as increasing teacher pay by $200,000 per year), virtually all students will be attracted to teaching (both high- and low-ability students). However, even in this implausible and expensive scenario, because students of all ability levels will choose to enter teaching, unless schools and districts can identify and select the most talented applicants, the average teacher will be of similar ability level to the average student.

What happens when the dispersion in teacher salaries grows?

Another approach to altering teacher pay would be to make it more responsive to ability so that the dispersion of teacher pay is increased. Consider the simple case, in which median-ability teacher salaries are kept the same in both occupations but pay is reduced for low-ability teachers and increased for high-ability teachers to the point where the lowest-ability (lowest paid) teachers earn less than the lowest-ability nonteachers and the highest-ability teachers earn more than the highest-ability nonteachers. This situation is analogous to the situation outlined earlier (depicted in Figure 3, top), but with teaching and nonteaching reversed. Because the median-ability student will earn the same in teaching and nonteaching, he or she would be indifferent between the two occupations. However, students with below-median ability will all choose nonteaching, and students with above-median ability will all be attracted to teaching. Under this pay scheme, only students who are above the median will choose to enter teaching so

that, even if schools and districts cannot identify and select the most talented applicants, the average ability of teachers will be *higher* than the ability of the average student.

Comparing the two strategies

Whereas increasing teacher pay across the board attracts more students with abilities above the median into teaching, it does not dissuade low-ability students from teaching, and it does not attract the highest-ability students. In contrast, increasing the dispersion in teacher pay by making it more sensitive to ability attracts more high-ability students into teaching (even those of the highest ability) and dissuades low-ability students from entering teaching. As such, by increasing teacher pay for everyone (even the less effective teachers), one both pays more to all teachers and has a lower-ability teaching pool than under the system that increases pay dispersion and only pays high-ability students high wages in teaching. While these examples are stylized, they highlight the economic forces at play and make clear that while the average level of teacher pay will affect the ability levels of those going into teaching, the dispersion in teacher pay will generate much larger selection effects and is likely to be more important for the average ability levels of teachers. In the next section, I outline the main empirical evidence supporting this model of the world.

Empirical evidence on wage structure and teacher ability

Across a variety of nations, the decline in the aptitude of teachers is coincident with (1) declines in the level of average teacher pay, (2) a narrowing of the wage distribution in teaching, and (3) a widening of the wage distribution in nonteaching; that is, across a variety of nations, the wages paid to teachers on average have declined over time relative to nonteacher wages, the difference in wages between the highest-ability teachers and the lowest-ability teachers has declined over time, and the difference in wages between the highest-ability nonteachers and the

lowest-ability nonteachers has increased over time.[7] The Roy models predicts that each of these should have led to a decline in the aptitudes of teachers. While Hoxby and Leigh (2004) find that pay compression for teachers in the United States increased the share of the lowest-aptitude female college graduates who became teachers by about 9 percentage points and decreased the share of the highest-aptitude female college graduates who become teachers by about 12 percentage points, Leigh and Ryan (2006) find that the rise in pay differentials associated with ability in nonteaching occupations is responsible for the decline in the academic aptitude of new teachers in Australia. Irrespective of exactly how one apportions "blame", fundamentally, what has happened across a variety of nations is that for an individual with the potential to earn a wage at the top of the earnings distribution in any occupation, teaching has become a much less lucrative option today than it was in the past.

Policy implication

The policy implication of the fact that the relative compression of the teacher pay distribution is the primary driving force behind the decline in teacher aptitude over time for a variety of nations is that teacher pay should be structured to attract high-aptitude students and deter low-aptitude students. Simply put, teaching should be lucrative financially for those who are good at it (presumably, high-aptitude students) and not lucrative for those who are bad at it (presumably, low-aptitude students). In most countries, however, teacher pay is almost completely unrelated to aptitude or performance on the job. In the United States, for example, in most districts, teachers are paid based on the number of years of experience, being regularly

[7] In Australia, Leigh and Ryan (2006) document an increase in the ratio of pay at the 90th percentile relative to the 50th percentile (so that the highest-paid nonteachers saw relative increases in pay compared to the average nonteachers). In contrast, they see a decrease in the ratio of pay at the 90th percentile relative to the 10 percentile (so that the highest-paid teachers saw relative decreases in pay compared to the lowest-paid teachers). Similarly, Hoxby and Leigh (2004) document that in the United States, the wages of female teachers from the most selective universities became more similar to those of teachers from the least selective universities between 1963 and 2000, at the same time that the wages of female nonteachers from the most selective universities were rising.

certified, and having a master's degree. This is also true in Israel (Lavy, 2002), Sweden (Hensvik, 2010), and other nations. Teachers who perform well on the job earn the same amount, have the same likelihood of being fired, and receive the same raises as those who perform poorly, conditional on these three characteristics.

There are a variety of ways one could make teacher pay more responsive to skills such that high- and low-ability workers will be able to earn higher and lower wages in teaching than they currently do, respectively. The first is to use *performance-based pay*. While much of the research on performance-based teacher pay has focused on the short-term incentive effects (i.e., a teacher may exert more effort and have better outcomes to receive higher pay under performance-based pay), there could also be potential long-term selection effects associated with performance-based teacher pay; that is, if high-ability workers are more likely to have good performance and therefore earn higher pay than is comparable to what nonteaching professions pay, they will be attracted to teaching. In contrast, low-ability teachers who know that they will not have good performance have no additional incentive to pursue a career in teaching. While this is certainly a real possibility, sadly no empirical evidence exists to date, to my knowledge, documenting this.

Another approach is to use *performance-based promotions* (such as in academia). In most systems, all teachers receive the same raise based on how long they have been teaching. A simple tweak to this would reward high performers (who are most likely to be high ability) with larger raises and weak performers with no raises. Under such an approach, high-ability students may select entering into teaching even if starting salaries are low if they know that if they work hard, their earnings will rise rapidly over time. Such an approach has the added benefit of giving an increased incentive for high-ability teachers to remain in the profession. Related to this,

another approach could give high-ability teachers retention bonuses to keep those who enter the profession from leaving. Clotfelter et al. (2006) find that bonus payments of $1,800 (about 2 percent of the average teacher's salary) were sufficient to reduce mean turnover rates of the targeted teachers by 17 percent. This suggests that a targeted retention bonus program based on teacher performance could alter the composition of teacher staffs by retaining effective teachers.

The other side of this coin is *performance-based firing.* If we think of a teacher's expected wages as being the teacher's wage if employed multiplied by the likelihood of being employed, one way to make expected teacher pay more sensitive to ability without altering the structure of teacher pay is to make the likelihood of future employment responsive to teacher ability. If low-aptitude or low-quality teachers know that they will be fired if they do not perform well in the classroom, they will be less likely to enter the profession and will likely seek employment elsewhere. It is worth noting that in addition to the selection effect due to a teacher's own actions, this policy will also have a direct effect on the composition of teachers by removing low-performing teachers from the teaching pool—I discuss this mechanism separately in section 2.2.

While there is relatively little evidence of the effect of increasing the dispersion in teacher pay on student outcomes, the empirical evidence supports the idea that increasing teacher pay can be beneficial. A number of studies document that geographic areas that experience larger increases in teacher pay over time experience larger improvements in student outcomes (Heckman et al., 1995; Card and Krueger, 1992), and Loeb and Page (2000) find that once one adjusts for labor market factors, raising teacher wages by 10 percent reduces high school dropout rates by 3 to 4 percent. Because these increases in teacher pay were not random, it is not obvious that this relationship is causal; however, the robustness of the positive association between

teacher pay and student outcomes using different models and different samples suggests that the relationship is positive. As such, policies that increase teacher pay will likely have a positive impact on student outcomes.

Because most evaluations of performance-based teacher pay have been on short-run interventions that are unlikely to generate any selection effects, there is little conclusive direct evidence of the selection effects of performance-based pay and the associated improvement in student outcomes. However, Figlio and Kenny (2007) use survey data from across the United States and find that test scores are higher in schools that offer individual financial incentives to teachers for good performance. Similarly, using country-level performance-pay measures with international achievement microdata, Woessmann (2011) finds that the use of teacher salary adjustments for outstanding performance is significantly associated with better math, science, and reading achievement across countries. Of course, these studies capture both the incentive effects of performance pay (due to teachers working harder) and the selection effects on the quality of teachers (due to higher-ability students entering teaching) and may be affected by other differences that are correlated with introducing teacher performance pay. However, the results are suggestive that making teacher pay more responsive to ability and performance might attract better prospective teachers into the profession and, accordingly, may improve student outcomes.

2.2 Identify and retain good teachers and dismiss bad teachers

The previous section deals with how one might use the teacher compensation system to induce self-selection of talented individuals into teaching and to encourage good teachers to remain in teaching. Another strategy to change the composition of teachers is to actively retain good teachers and dismiss or fire bad teachers. This strategy is predicated on the notion that we

have reliable information about which teachers are good and which teachers are not good. As discussed previously, teachers exhibit considerable heterogeneity in their ability to raise student test scores. In principle, if teacher quality were perfectly observable, one could simply hire ony the best teachers, and one's problem would be solved. Unfortunately, the world is not so simple because teacher quality is difficult to determine before a teacher is hired. To investigate the extent to which schools might be able to predict teacher effectiveness before teachers are hired, Jacob et al. (2011) administered an in-depth survey to new elementary and middle school math teachers in New York City that assessed a host of teacher qualities at the time of hire, including cognitive ability, content knowledge, personality traits (e.g., openness, conscientiousness, extraversion, agreeableness, and neuroticism), and feelings of self-efficacy (a person's belief in his or her own competence). They find modest and marginally statistically significant relationships between student achievement and several nontraditional predictors of teacher effectiveness and conclude that while schools wishing to increase the effectiveness of their teacher force may benefit from gathering a broad set of information on new applicants, data on job performance will still be a more powerful tool for improving teacher selection than data available at the recruitment stage.

The most straightforward way to use on-the-job performance to remove bad teachers is to determine some cutoff value-added level and fire or dismiss all teachers with estimated value added below this cutoff level. Rockoff and Staiger (2010) point out that if there were no hiring costs and teachers did not improve with experience because replacement teachers will, on average, be as effective as the average effectiveness in the population, principals should set the cutoff where the effectiveness of the *marginal* teacher (the least productive teacher not fired) is equal to the effectiveness of the *average* teacher in the population; that is, principals should keep

only those teachers who are expected to be better than the average replacement teacher. If the principal were to set a cutoff above the level of the effectiveness of the average teacher, then by lowering the cutoff slightly, the principal could have more above-average teachers. As such, the cutoff should not be above the average effectiveness. Similarly, if the principal were to set a cutoff below the level of the effectiveness of the average teacher, then by increasing the cutoff slightly, the principal could hire fewer below-average teachers. As such, the cutoff should not be above the average effectiveness. Taken together, in this simple world with no costs to hiring and no experience effects, the cutoff should be at the average effectiveness in the population.

In the real world, there *are* costs to hiring, and teachers *do* learn on the job. Specifically, Barnes et al. (2007) and Milanowski and Odden (2007) study recruitment, hiring, and replacement costs and find that replacing a teacher costs a school anywhere between $4,000 and $15,000, with an average estimate of roughly $8,200. It is worth noting, however, that some of these monetary costs will be defrayed by the lower salaries earned by new first-year teachers. Also, because teachers tend to improve substantially in their first year of teaching, there is an additional cost in that students assigned to the replacement rookie teacher over the first two years of teaching lose about 0.07 standard deviations in student achievement. The students' lost lifetime earnings associated with these achievement losses are likely larger than the direct replacement costs. When all is taken into account, the optimal dismissal rule should set the cutoff effectiveness level to maximize student achievement while taking into account the costs of recruiting, training, and hiring rookie teachers.

The effect of this optimal dismissal policy on student outcomes will depend primarily on two important factors: (1) how long teachers stay in the profession and (2) how stable teacher effectiveness is from year to year. I briefly describe the implications of each of these factors, and

I then present the research findings on what kinds of improvements one could expect from imposing the optimal policy of removing the least effective teachers.

How long teachers stay in the profession is important for the effect of the dismissal policy on student achievement. If a large fraction of teachers leave the profession for other employment and have to be replaced with average teachers, then a large share of the above-average teachers not fired will leave the profession and have to be replaced with average teachers—dampening the effect of firing the ineffective teachers. In the extreme hypothetical case in which all teachers in one year leave the profession the following year, firing ineffective teachers will have no effect on the quality of teachers the following year. By the same logic, if all the high-quality teachers not fired remain in the profession, the policy will have maximal effect, all else being equal. The U.S. Department of Education estimates that among public school teachers who were teaching during the 2007–08 school year, 84.5 percent remained at the same school, 7.6 percent moved to a different school, and 8.0 percent left the profession during the following year.[8] As such, from the perspective of a school, the loss rate of those retained would be approximately 15 percent, while the loss rate for the teaching profession as a whole would be 8 percent. This suggests that while high levels of teacher turnover would mute the effects of any policy of removing bad teachers, the aggregate teacher attrition rate is relatively low such that this is likely to be a second-order concern.

The stability of teacher effectiveness from year to year is also important for the effect of this policy on student achievement. Consider the extreme case in which a teacher's effectiveness does not change at all over time so that a teacher who is estimated to be very strong in one year will be very strong the following year. In this example, all the teachers who are found to be above average today will perform above average next year so that performance will improve

[8] http://nces.ed.gov/pubs2010/2010353.pdf.

relative to the policy whereby no teachers are fired. Consider the other extreme example, in which a teacher's effectiveness in one year is completely unrelated to his or her effectiveness the following year. In this example, all the teachers who are found to be above average today will perform like average teachers next year so that student performance will be the same under the dismissal policy as under a policy in which no teachers are fired (under the simplifying assuming that teachers do not improve with experience). These examples illustrate that the more stable teacher effectiveness is over time, the more effective the optimal dismissal policy will be. A variety of empirical studies find that more than half of the variation in estimated individual teacher impacts on tests in math and English in a given year is not persistent over time. Specifically, estimates of the correlation between estimates for the same teacher from one year to the next range between 0.2 and 0.4 (Sass, 2008; Goldhaber and Hansen, 2008; Rockoff and Staiger, 2010). Put another way, somewhere between about one-third and one-half of the estimated effectiveness from one year can be expected in another year. This suggests that while the benefits to the optimal dismissal policy may not be enormous, teacher effectiveness estimates are sufficiently reliable over time that the dismissal policy is likely to yield economically meaningful benefits.

Policy implications

While this policy has never been carried out, Rockoff and Staiger (2010) take the available empirical estimates about teacher turnover, the increase in teacher productivity associated with experience, and the persistence of teacher effectiveness over time to simulate the effects on steady state (long-run) teacher quality under the optimal teacher quality cutoff rule.[9]

[9] They set the standard deviation of the persistent teacher effect (in student-level standard deviation units) equal to 0.15 and the reliability of the value-added measure (the ratio of the persistent variance to total variance) equal to 40 percent. For the return to experience, they assume that a first- and second-year teacher's value added is –0.07 and –0.02 student standard deviations below the value added of teachers in their third year or higher. They ignore the

Using Monte Carlo simulations, they find that teacher value added in steady state could be increased by 0.08 standard deviations (in student achievement units) using the optimal dismissal strategy after the first year of teaching. Note that this estimate is based only on the changing composition of teachers and does not include any benefits due to incentive effects. They also analyze the effects of the optimal dismissal strategy whereby the required dismissal only occurs after a teacher's second, third, and fourth year. While postponing the required dismissal year allows one to get more reliable information on teachers, this comes at the cost of keeping bad teachers on staff for longer. As such, if the policy requires that all dismissals take place at a certain point in a teacher's career, dismissal after the first year is best. Finally, they analyze the effects of a policy that allows dismissal at any point in time up until the fourth year of teaching. This is a more flexible policy as it allows one to remove teachers who are clearly below the threshold after one year but allows one to gather more reliable estimates of teacher effectiveness for other teachers. With this more flexible optimal dismissal policy, they estimate that steady state teacher effectiveness could be increased by 0.1 standard deviations on average in student achievement units. To put these estimates into perspective, the optimal flexible cutoff dismissal policy would raise student achievement by about 4 percentile points on average. Under this strategy, 67 percent of teachers would be dismissed after their first year of teaching, 8 percent after their second year, 4 percent after their third year, and 5 percent after their fourth year.

Adopting a policy that would dismiss more than 80 percent of all early-career teachers may not be politically palatable. As such, it is helpful to describe the gains that one could expect in steady state with a more modest approach. It turns out that the increase in steady state value added is roughly linear to the proportion of teachers dismissed (below the optimal dismissal

---

direct costs of hiring a new teacher. Finally, they assume a maximum teaching career of 30 years and an exogenous turnover rate of 5 percent, which is approximately the proportion of experienced teachers who leave the Los Angeles and New York City districts each year.

rate).[10] Therefore, as a rule of thumb, steady state value added increases by approximately 1.1 percentage points for every 10 percent of low-performing early-career teachers dismissed. This suggests that a less aggressive policy of removing the bottom third or half of teachers would increase steady state teacher value added by 3.5 and 5.5 percent of a standard deviation in student achievement units, respectively. This is between 1.5 and 2.5 percentile points on average.

This simulation is predicated on a variety of assumptions such as no spillovers across teachers, no incentive effects, and that there is an unlimited supply of teachers. If effective teachers improve the outcomes of their peers as suggested in Jackson and Bruegmann (2009), the achievement gains could be as much as 30 percent greater than those calculated. Also, if teachers work harder because they know that they might be dismissed for poor performance, there could be some positive incentive effects that would make the increases in student achievement even greater. However, if there is not a large supply of prospective teachers (as may be the case for certain remote locations and for difficult to staff schools), the improvements in teacher quality may be limited. In any case, the simulated effects under reasonable assumptions are sufficiently large that dismissal policies based on estimated effectiveness should be considered.

3. Getting teachers to work harder

The discussion thus far has described teaching quality as an immutable characteristic of teachers and has been focused on altering the composition of teachers. However, there is mounting evidence that how well teachers perform in the classroom varies over time and across classroom and schooling contexts. Specifically, Aaronson et al. (2007) and Koedel and Betts (2009) estimate the teacher-year effects based on student achievement data; they group teachers into performance quintiles based on fixed-effects estimates of teacher performance and find

[10] On the basis of the author's own calculations and those in Rockoff and Staiger (2010).

considerable movement between quintiles—suggesting that an excellent teacher one year is not necessarily a wonderful teacher the following year. Also, Jackson (2010c) finds that teacher performance varies systematically across schools. Jackson finds that the match between teacher and school can "explain away" a quarter of the difference and is as economically important as teacher quality. In light of these findings, it is clear that thinking of teachers as fixed quantities of value added that can be moved around and organized *may* be overly simplistic. These studies also suggest that there may be substantial benefits to improving teacher performance through mechanisms other than changing the composition of teachers. The variation in individual teacher performance across schooling contexts and time may be due to differences in teacher *effort* across contexts and time. Insofar as this is true, there may be sizable gains to increasing the effort levels of the teachers we already have.

3.1 Teacher incentives

One policy aimed at improving student outcomes by increasing teacher effort is teacher incentive pay or pay for performance. Pay for teacher performance has been adopted in numerous districts in the United States since the 1990s, by the Pay Performance and Management Reform in the United Kingdom, by the Victorian Government Schools Agreement in Australia, by the Carrera Magisterial Program in Mexico, and by National System of School Performance Assessment in Chile as well as through reforms in other nations.

The use of teacher incentive pay is predicated on the notion that teacher outcomes are a function of both teacher ability and teacher effort and that teachers do not expend the optimal level of teacher effort. Data from the United States, Israel, Sweden, and the United Kingdom show that most public school teachers are employed by districts or schools that use salary schedules to determine pay (Podgursky, 2007; Lavy, 2002; Atkinson et al., 2009; Hensvik,

2010). Thus most teachers are paid primarily based on years of experience and education level. As such, teacher pay is largely unresponsive to actual teacher performance. Consistent with this, Jackson (2010c) and Chingos and West (2010) find that the correlation between teacher pay and teacher performance in the classroom is weak. If a large component of teacher performance is effort, the unresponsiveness of teacher salary to classroom effectiveness would imply that the *monetary* benefits to exerting more effort to improve student outcomes are close to zero. If teacher effort is responsive to monetary incentives, student outcomes could possibly be improved by changing the compensation scheme such that teachers benefit monetarily from their own effort. In other contexts, researchers have found that worker effort and output are higher when workers are paid for the output directly versus with an hourly wage (Foster and Rosenzweig, 1994; Lazear, 2000). If teaching is like other occupations, rewarding teachers for their performance may increase teacher effort and improve student outcomes.[11]

Where there is a close correspondence between teacher effort and performance, and where performance is well measured, a pay-for-performance contract should elicit more effort and improve student outcomes more than a standard wage or salary contract (as most teachers have). However, in reality, neither condition is met because we cannot observe teacher effort, and teacher performance is imperfectly measured. Instead, principals are forced to infer teacher effort based on student outcomes, such as student test scores, which themselves are not a perfect measure of student learning (the outcome we really care about). This leads to three problems with pay for performance in teaching.

[11] Readers may wonder whether the need for test-based accountability arises from the fact that school principals do not bother to observe teacher performance in the classroom. However, this is not the case. In many school districts in the United States, principals do observe and evaluate teacher performance. While these principal assessments do predict a teacher's subsequent classroom performance, estimated value added is a much better predictor of a teacher's future success in the classroom (Jacob and Lefgren, 2008). Also, as a practical matter, because teachers are often public employees, it is politically difficult to dismiss a teacher based on subjective assessments because there is always the worry of bias.

The first problem is that if the outcome measure (say, test scores) can be improved without exerting effort, teachers will have an incentive to do so. This would involve practices, such as "teaching to the test" and cheating, that do not require the teacher to work harder and do not lead to greater learning (which is not measured) but do improve student test scores (which are measured). Evidence of this kind of behavior has been documented in high-stakes test-taking contexts. For example, Figlio (2006) finds that low-performing students are relatively more likely to be suspended than high-performing students around the time period when accountability tests are administered. Jacob (2005) finds that test score gains associated with school-level accountability were driven largely by increases in test-specific skills and student effort and did not lead to comparable gains on a state-administered low-stakes exam. Also, Jacob and Levitt (2003) find evidence of teacher cheating on high-stakes student tests. These examples underscore the fact that an effective performance pay system should be based on measures of teacher performance that cannot be easily gamed (i.e., improved without increasing effort).

The second problem stems from the fact that we often care about outcomes other than the measured outcome (such as test scores). Holmstrom and Milgrom (1991) point out that incentives to perform on one dimension may cause agents to withdraw effort from other dimensions. For instance, if teacher pay is based on student standardized test scores, teachers may spend more time teaching the topics covered on the test to improve test performance (which is easily measured) and less time nurturing a deep understanding of, and interest in, the subject (which are difficult to measure). If these difficult-to-measure aspects of performance are sufficiently important, one may be better off without any performance rewards at all. This point highlights that pay-for-performance systems will not unambiguously lead to improved student

outcomes and that any well-designed pay-for-performance scheme must be based on outcomes that are a good measure of student learning.

The third problem stems from the fact that effort and outcome measures may not be closely related. Performance pay is predicated on the notion that when a teacher exerts more effort, student outcomes—in this case, test scores—will tend to improve. This may not be true for two reasons. The first is that student test scores are influenced by a variety of factors that are outside the control of the teacher. If these outside influences change sufficiently often, then the relationship between outcomes and teacher effort will be weak. Knowing this, a teacher may not be compelled to exert effort if he or she knows that it will likely not result in better outcomes and increased pay. This problem can be reduced statistically by accounting for the influence of student attributes and family influences so that one can net out the influence of the teacher (much like the value-added approach). The second problem, raised by Murnane and Cohen (1986), is that merit pay may not be effective at improving teacher performance if individual teachers do not know what to do to improve their teaching performance. If teachers do not know *how* to improve student outcomes, even if they exert more effort, one may have small or no improvements in student outcomes. The implication of this is that pay for performance is likely to be most effective in contexts in which the outcome measures are well understood (such as a standardized test that has been administered for several years so that there is a general familiarity with how to improve performance) or in which some guidance or coaching is provided so that teachers know how to turn their increased effort into better student outcomes.

3.2 Empirical evidence

The preceding discussion highlights that the effects of performance pay are likely to be context specific and dependent on how the incentive scheme is designed. As such, one might

expect that across different studies from different nations and different outcomes, the results will be mixed. This is the case. Where one would expect the effects to be context specific, the correct question to ask is not "does it work?" but rather "can it work?" and "in what contexts does it work?" I review the empirical literature from this perspective.

While there are examples of teacher performance pay schemes that did not improve student outcomes (e.g., Fryer, 2011b; Goodman and Turner, 2009; Springer et al., 2010), the empirical evidence indicates that individual teacher performance pay for individual teacher outcomes can work. Lavy (2009) analyzes an experimental program in Israel that offered individual teachers bonus payments on the basis of the performance of their classes on high school graduation exams in English and mathematics. The bonuses were structured so that teachers were rewarded on the basis of their performance relative to other teachers of the same subjects in the same school. The rewards ranged between 6 and 25 percent of the average teacher's salary. Lavy finds that the intervention increased overall pass rates by 12 percent and average math scores by 10 percent. Effects were less precisely estimated and about half the size in English. Using survey data, he finds that improvements were mediated through changes in teaching methods, enhanced after-school teaching, and increased responsiveness to students' needs. Similarly, positive results have been found for a pay-for-performance scheme in England that rewarded individual teachers at least an 8 percent permanent salary increase for improving student performance. Atkinson et al. (2009) find that this payment scheme improved secondary school test scores, and value added increased on average by about 40 percent of a grade per pupil. Similarly, Muralidharan and Sundararaman (2009) analyze an experimental program in India that provided bonus payments to individual teachers based on the average improvement of their students' test scores in independently administered learning assessments (with a mean

bonus of 3 percent of annual pay). After two years, students in incentive schools performed significantly better than those in control schools by 0.28 and 0.16 standard deviations in math and language tests, respectively (improvements of 12 and 6 percentage points in math and reading, respectively).

Taken together, these studies show that individual teacher incentive pay can lead to sizable improvements in student outcomes. They demonstrate that this is true both in developing nation contexts (where one may expect large effects due to improvements in teacher absenteeism) and in developed nations, such as Israel and the United Kingdom, where improved outcomes likely come from other changes in teacher behavior.

Where there are positive spillovers across teachers, and teachers learn from each other, systems that reward individual teachers for group outcomes may be more effective. However, Lazear (2000) points out that under such group incentives, teachers may put in less effort because they can free ride off other teachers. Studies have found positive effects of individual teacher pay for *group outcomes.* Lavy (2002) analyzed a program that provided both the school and its teachers with monetary performance incentives for school-leaving-exam performance and the school dropout rate. The bonuses paid to teachers were between 1 and 3 percent of the average teacher salary and were associated with test score improvements of between 3 and 6 percent and reductions in the dropout rate of about 2 percentage points.

It is well documented that both individual and group level incentives can improve student outcomes in meaningful ways, but this is not always the case. Glewwe et al. (2008) analyze a randomized evaluation that provided primary school teachers in Kenya with group incentives based on test scores and find that, while program schools saw improved test scores in the short run, students did not retain the test-score gains after the incentive program ended. They conclude

that the results are consistent with teachers expending effort towards short-term increases in test scores but not towards deeper longer-term learning.

To help understand the situations in which performance pay is likely to improve outcomes, I discuss the results from papers that find differences in treatment heterogeneity across different incentive designs within the same study and look for similarities in the contexts among those studies that find positive effects. Muralidharan and Sundararaman (2009) compare individual and group incentives within the same experiment in India and find that group and individual incentive schools perform equally well in the first year of the program but that the individual incentive schools significantly outperform group incentive schools in the second year. This suggests that incentive pay schemes for individual outcomes are likely to produce better outcomes than those for group incentives. Consistent with this, Sojourner et al. (2011) compare the effects of different kinds of pay-for-performance schemes in Minnesota and find that districts offering greater rewards for teacher-level goals experienced large gains in reading, whereas those offering rewards based on school-wide goals or subjective evaluations did not. Also, an evaluation of District Awards for Teacher Excellence, a state-funded program in Texas, finds that the schools and districts with larger maximum awards to individual teachers had larger test score gains in math. Taken together, these results suggest that (1) individual-level incentives are likely to produce better outcomes and (2) higher-powered incentives (i.e., larger bonuses) are likely to produce better outcomes.[12]

Among studies looking at group-level incentives, Fryer (2011b) and Goodman and Turner (2009) analyze group incentive in New York City and find that while the average effects

[12] Because teachers may not always be able to translate increased effort into improved outcomes, it is not obvious that higher-powered incentives will lead to improved outcomes. In fact, there is evidence that higher-powered incentives could actually lead to worse outcomes either through increasing the stress level of teachers or by exacerbating any undesirable distortionary behaviors.

of performance pay were zero, there were some positive effects in smaller schools. This suggests that group incentives may be effective in small-group contexts in which the free-rider problem is less severe.

As mentioned previously, one would expect better outcomes in contexts in which there is support for teachers or where there is widespread familiarity with the test because merit pay may lead to improved teacher performance when teacher have ways to figure out *how* to improve student outcomes (such as other teachers or test manuals). Although no studies have manipulated the level of teacher support directly, the body of evidence suggests that this may be important. Specifically, some of the most successful studies are based on well-established national tests (Israel and England) for which teachers have a great deal of institutional knowledge at their disposal about how to improve outcomes. In contrast, many of the unsuccessful performance pay schemes in the United States are based on new tests that are not well known so that knowledge of the test is limited. While not discussing a pure teacher performance pay intervention, Jackson (2010b) provides suggestive evidence of this. Specifically, Jackson analyzes the longer-run effects of a high school intervention that includes cash incentives for both teachers and students for each passing score earned on Advanced Placement exams, teacher training, and curricular oversight. Jackson finds the largest positive effects on college going and college performance at those schools that had a long-standing and well-developed Advanced Placement infrastructure before the intervention was adopted. Taken as a whole, the empirical evidence points to a need for teachers to know how to improve the outcome being rewarded.

3.3 Policy implications

In summary, the empirical evidence on the effectiveness of teacher performance-based pay suggests three things: (1) pay for performance can lead to large improvements in student

outcomes, (2) pay for performance is not effective in all contexts, (3) performance schemes that reward individual-level outcomes are likely to be more effective than those that reward group outcomes, (4) performance schemes that reward group outcomes are more likely to be effective when group sizes are small so that teachers do not free ride, (5) schemes that have higher rewards are likely to be more effective than those with lower rewards, and (6) schemes that are based on well-understood outcomes and provide guidance about how to turn increased effort into improved outcomes are likely to be most effective. The range of estimates indicate that a well-designed pay-for-performance policy could have sizable positive effects on student outcomes. To put the range of positive estimated effects in context, Krueger (1999) analyzes the well-known project STAR (Student/Teacher Achievement Ratio) class size reduction experiment and finds that reducing class size by seven students increased average test scores by 4 percentile points. The most successful interventions yield larger effects than this and are likely to be less costly.[13] Also, Lavy (2002) compares the costs and benefits of teacher performance pay to those of spending an equivalent amount on school resources and concludes that expenditures on teacher performance pay are "much more cost effective" that those on additional school resources.

Because teacher performance pay is based on performance at least in part based on a test, an important consideration is test design. Neal and Schanzenbach (2011) illustrate that test-based accountability systems based on the numbers of students who score at or above specified proficiency levels provide incentives for teachers and principals to target children near current proficiency levels for extra attention and weak incentives to devote extra attention to students who are clearly proficient already or who have little chance of becoming proficient. As such, performance pay should ideally be based on an outcome measure that is sensitive to the

[13] Note that the causal estimates from project STAR are among the larger of the estimated class size effects. However, because they are based on an experimental design, they are likely to be the most internally valid.

improvement of all students (such as average test score growth) so that teachers have an inventive to spend extra effort on all students. In fact, Barlevy and Neal (2011) point out that because the effort required to improve test scores by 1 point may differ for low- and high-achievement students, the ideal incentive pay scheme would link teacher compensation to the ranks of their students relative to observationally similar students at the beginning of the school year. Under such a scheme, teachers will have a clear incentive to improve the outcomes of all students and will be evaluated relative to teachers with very similar student populations.

Another practical consideration is how often students need to be tested. Often standardized tests are not administered every year. In principle, one could base teacher performance pay on the results of a single test. The problem with this is that those teachers who have less well-prepared students will be penalized, whereas those with better prepared students will be rewarded. If students were randomly assigned to teachers, then annual testing would not be necessary. However, because students typically sort into schools and into classrooms, it will be important to have some sort of pre-exposure measure of student ability so that teachers can be compared to other teachers with similarly well prepared students. As such, as a practical matter, the use of test-based teacher pay will likely necessitate the use of two tests: one before exposure to teachers to determine pre-exposure proficiency and one after exposure to teachers to be used to determine teacher value added (based on some measure of student achievement growth).

4. Give teachers the knowledge and skills they need to be effective

The final approach to improving the outcomes of existing teachers is to improve their teaching skills. This approach is predicated on the notion that while innate teacher ability and teacher effort are important for student outcomes, teachers can become more effective in the

classroom through improving their skills. Several studies find that while there is wide variation in effectiveness among teachers, the benefits to additional years of experience are small beyond the first few years of teaching (e.g., Rockoff, 2004; Rivkin et al., 2005). However, recent work by Wiswall (2011) and Papay and Kraft (2010) find that teacher effectiveness improves even after the first few years of teaching. Wiswall (2011) in particular finds that after accounting for the quality of teachers who remain in the profession, the returns to later career teacher experience is sizable.[14] This implies that teaching ability is highly malleable so that interventions that aim to increase teacher knowledge and skills could potentially be effective at increasing average teacher effectiveness and overall student outcomes.

4.1 Evaluation programs

Some argue that teacher evaluation can improve teacher performance due to both the incentives inherent in the evaluation and to increases in teachers' professional and skill development. Specifically, because teachers may have a desire to have positive performance evaluations, they may exert more effort be more likely to invest in their skills. Also, because evaluations may provide teachers with helpful information on ways in which to improve, teachers may use this information to improve their teaching skills. Despite these possibilities, most of the existing evaluation programs are viewed as being poorly designed and therefore ineffective (Donaldson, 2009; Weisburg et al., 2009). In an overview of teacher evaluation systems in the United States, Donaldson (2009) describes most evaluation systems as follows:

> External constraints decrease evaluators' inclination to evaluate rigorously—vague district standards, poor evaluation instruments, overly restrictive collective bargaining

[14] He estimates that a teacher with 30 years of experience has over 1 standard deviation higher measured quality than a new, inexperienced teacher and about 0.75 standard deviations higher measured mathematics effectiveness than a teacher with 5 years of experience. In comparison, estimates on the same data sample using previously restricted models suggest that experienced teachers have between 0.1 and 0.2 standard deviations higher quality than new teachers, with almost all of these gains in the first few years of teaching.

> agreements, and a lack of time all contribute to this problem. Internal constraints, such as the absence of high-quality professional development for evaluators, a school culture that discourages critical feedback and negative evaluation ratings, and a district culture that offers little oversight and few incentives for administrators to evaluate accurately, also contribute to inflated ratings. (p. 2)

He also found that in most cases, "evaluators rarely provide teachers with substantive feedback" (p. 2). This highlights the fact that the relative lack of evidence supporting the use of evaluation systems may not be due to them being ineffective per se but rather to the fact that they are often poorly implemented.

Consistent with this notion, Taylor and Tyler (2011) study the effects of a seemingly well-designed teacher evaluation scheme on teacher performance in Cincinnati Public Schools. They find evidence that quality classroom observation–based evaluation and performance measures can improve mid-career teacher performance both during the period of evaluation and in subsequent years, consistent with human capital investment. Specifically, students assigned to a mid-career teacher after he or she participates in evaluation score about 0.1 of a standard deviation higher in math than similar students taught by the same teacher prior to participation. This is the same size effect for replacing an average teacher with a teacher at the 85th percentile of the distribution and is equivalent to improving test scores by 4 percentile points. This is about the same size effect as reducing class size by about 30 percent and is notable given that the sample being studied comprised mid-career teachers.

A few key aspects of this program likely led to its success: (1) teachers were evaluated based on rubric criteria that have been demonstrated to produce higher achievement;[15] (2)

[15] Teachers are evaluated on dozens of specific skills and practices covering classroom management, instruction, content knowledge, and planning, among other topics. Evaluators use a scoring rubric, based on Charlotte

evaluation outcomes were linked to career development such that teachers who did not have strong evaluations had to undergo a year-long process of intensive assistance from a mentor that included another full year of evaluation with more frequent classroom observations, along with other meetings and writing assignments; and (3) teachers received detailed feedback about where they were deficient and how to improve. In essence, this program appeared to overcome many of the shortcomings cited as responsible for the failure of other evaluation systems and led to meaningfully improved student outcomes. This suggests that similarly well designed programs might also be successful.

4.2 Peer mentoring

Another approach to improving teacher skills is to use peer mentoring. In the United States in 2003–04, nearly 70 percent of recently hired teachers reported that they received help from a mentor in their first year of teaching, and a majority of states require mentoring programs for new teachers (Rockoff, 2008). Despite the prevalence of mentoring programs, many of the empirical evaluations of mentoring programs are based on research methodologies that do not provide credible causal estimates.[16] In one, the first evaluation based on quasi-experimental variation, Rockoff (2008) studies the impact of mentoring in New York City. He finds strong relationships between measures of mentoring quality and teachers' claims regarding the impact of mentors on their success in the classroom but weaker evidence of effects on teacher absences, retention, and student achievement. However, he does find that retention within a particular school is higher when a mentor has previous experience working in that school, suggesting that an important part of mentoring may be the provision of school-specific knowledge. While the relationship may not be causal, he also documents that 10 additional hours of mentoring are

Danielson's (1996) *Enhancing Professional Practice: A Framework for Teaching*, which describes performance of each skill and practice at four levels: Distinguished, Proficient, Basic, and Unsatisfactory.

[16] See Strong (2005) for a review of this literature.

associated with 0.05 standard deviations higher student achievement in math and 0.04 standard deviations higher achievement in reading. If this suggestive relationship is causal, these are important effect sizes.

Although there is only suggestive evidence of benefits to formal peer mentoring, Daly et al. (2010) and Bakkenes et al. (1999) use network analysis to find that teachers are most likely to interact with teachers in the same grade level. As such, informal peer mentoring may be important for how teachers acquire skills. Using other teachers at the same school in the same grade as a measure of peers, Jackson and Bruegmann (2009) study the effect of working around more effective colleagues. They find that exposure to better teacher peers improves a teacher's own performance. Specifically, a 1 standard deviation increase in mean teacher-peer quality was associated with a 0.04 and 0.02 standard deviation increase in math and reading achievement, respectively. After two years, a 1 standard deviation increase in mean teacher-peer quality that persists over time was associated with a 0.078 standard deviation increase in student math test scores and a 0.072 standard deviation increase in reading test scores. Importantly, these effects persisted even after the working relationships ended, suggesting that they reflected peer learning.

Taken together, these findings suggest that peer learning and peer mentoring can be fostered to improve teacher effectiveness. In fact, Jackson and Bruegmann (2009) find that about 20 percent of teacher effectiveness can be explained by the value added of a teacher's peers over the previous three years, suggesting a significant role for peer learning. The evidence also suggests that the most effective peer mentors are likely to be teachers from the same school and the same grade who share similar experiences and can impart both pedagogical skills and specific knowledge about the grade and school.

4.3 Formal teacher training

The most obvious way to improve teacher quality would be to provide formal teacher training. Indeed, formal training is prevalent among teachers. In the United States, 72 percent of teachers report having engaged in training related to their main teaching assignment during the previous year (Parsad et al., 2001), and a similar fraction reports receiving training on new teaching methods. Unfortunately, most of the existing research on in-service teacher training suffers from the fact that the training is determined by teachers and schools so that it is unclear whether one can credibly compare the outcomes of teachers who undergo training to the outcomes of those who do not. There are two notable exceptions. First, Jacob and Lefgren (2004), who use a credible regression discontinuity design and find no statistically significant effect of receiving additional hours of teacher training on student outcomes in Chicago. Also, using a quasi-experimental design, Harris and Sass (2007) study the effects of various types of education and training on teacher productivity. Their results suggest that teacher training generally has little influence on productivity but that content-focused teacher professional development appears to be positively associated with productivity in middle and high school math. However, Angrist and Lavy (2001) find that an in-service teacher training program in Israel designed to improve the teaching of language and math rather than content was associated with between 0.2 and 0.4 standard deviations higher student test scores in non-religious schools. While these effects are not very precisely estimated, these are sizable effects of between 4 and 8 percentile points on average. This seemingly successful training program involved a mixture of counseling and feedback sessions for teachers, changes in the organization of class time, and training in the use of instructional aids.

Taken together, the results suggest that although many of the teacher training programs that have been evaluated have not been wildly successful at improving student achievement, the

results from Israel indicate that a well-designed in-service training program can be an effective tool for improving teacher quality. The challenge is to identify these effective programs so that best practices can be replicated in other settings.

5. Conclusions

Under the views that (1) the primary drivers of differences in teaching performance are differences in the innate ability of teachers, which is determined before entering the profession; (2) differences in teaching performance are primarily due to differences in teacher effort; and (3) differences in teacher effectiveness are primarily due to differences in teaching skills or know-how, there are three main approaches to improving teacher quality. They are to (1) change the composition of teachers so that only the most able are in the profession, (2) provide appropriate incentives so that teachers exert the optimal amount of effort to improve student outcomes and become more effective, and (3) provide teachers with the knowledge and skills they need to be successful in the classroom.

The empirical research indicates that there are large and persistent differences in teaching ability such that having a teacher at the 85th percentile of the value-added distribution versus a teacher at the median would be associated with 0.1 standard deviations higher student achievement. This implies that obtaining the correct mix of teachers could lead to sizable achievement gains. At the same time, the empirical research shows that by exerting more effort, teachers *can* increase their effectiveness by the same amount as having a teacher at the 85th percentile of the value-added distribution versus a teacher at the median—implying that while having a high-ability teacher is important, having a hardworking teacher may be equally important for student achievement. Also, the empirical evidence shows that teachers *do* become

more effective over time as they acquire more skills and that the achievement gains associated with well-designed mentoring, training, and evaluation programs are similar in magnitude to the gains associated with having a teacher at the 85th percentile of the value-added distribution versus a teacher at the median. As such, it is clear that all three approaches, if done correctly, may be successful at improving student achievement.

One common element of the different approaches to improving teacher quality is that most of them require that teaching effectiveness be measured in some way; that is, making pay more responsive to performance, using incentive-based pay, evaluation, and targeted peer mentoring all require the collection of data on teacher performance. As such, one of the most important policies that districts and schools should adopt is the use of data on teacher performance that can be used either as a diagnostic tool, as an outcome measure on which to base pay or personnel decisions, or as an evaluative tool. With this infrastructure put in place, most of the proposed education reforms can be implemented with relatively low cost.

In discussing the different approaches to improving teacher effectiveness, it is also apparent that many strategies are related to each other in meaningful ways such that some policies might work better when others are in place. For example, teacher training may be more effective when teachers are rewarded for their improved classroom effectiveness and therefore have an increased incentive to learn. Similarly, pay for performance may be more effective when coupled with teacher evaluations because the evaluations may provide teachers with much-needed information about where they are deficient and what they need to improve. Also, optimal dismissal policies and targeted retention bonuses may be more effective where teachers acquire many teaching-specific skills.

It is also the case that some policies might improve outcomes through all three mechanisms (changing the composition of teachers, changing the effort levels of teachers, and increasing the skills of teachers). In light of evidence that all three mechanisms can lead to meaningfully improved student outcomes, strategies that operate through all three mechanisms might be most useful. For example, if districts were to provide teachers with performance-based[17] raises (as in many occupations), while also providing high-quality teacher training, this could induce the selection of high-quality applicants to teaching, lead to an increase in effort for existing teachers, and also give teachers an incentive to invest in their skills and undergo high quality training. The combination of all three mechanisms seems the most likely to be effective. The empirical literature suggests that schools or districts that use performance-based measures in the dismissal and retention of teachers, make teaching financially lucrative for high-ability students, and provide high-quality professional development and training would likely experience large improvements in student outcomes.

Finally, the fact that there are few rigorous studies on best practices for evaluation programs or training programs or for pay-for-performance programs suggests that policy makers might wish to experiment with different versions of these programs and then adopt those practices that yield the best results. While this may not be a short-term solution, if districts and schools were more experimentally minded and were to adopt programs in ways that could be rigorously evaluated, in the long-run, the successful interventions (that often led to improved student outcomes) could be implemented in many schools.

[17] These performance measures need not be test based but can be based on classroom observations and take the form of principal or peer evaluation—allowing for the evaluations to serve not only as an evaluative tool but also as a diagnostic tool and for useful feedback.

References

Aaronson, D., Barrow, L. and Sander, W. (2007), Teachers and student achievement in the Chicago public high schools, Journal of Labor Economics 25, 95–135.

Angrist, J.D. and Lavy, V. (2001), Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools, Journal of Labor Economics 19, 343–369.

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H. and Wilson, D. (2009), Evaluating the impact of performance-related pay for teachers in England, Labour Economics 16, 251–261.

Bakkenes, I., De Brabander, C. and Imants, J. (1999), Teacher isolation and communication network analysis in primary schools, Educational Administration Quarterly 35, 166–202.

Barlevy, G. and Neal, D. (2011), Pay for percentile, American Economic Review, forthcoming.

Barnes, G., Crowe, E. and Schaefer, B. (2007), The cost of teacher turnover in five school districts, report, National Commission on Teaching and America's Future.

Booker, K., Sass, T.R., Gill, B. and Zimmer, R. (2008), Going beyond test scores: Evaluating charter school impact on educational attainment in Chicago and Florida, RAND Corporation, Santa Monica, CA.

Brewer, D.J. and Ehrenberg, E.G. (1994), Do school and teacher characteristics matter? Evidence from high school and beyond, Economics of Education Review 13, 1–17.

Card, D. and Krueger, A.B. (1992), Does school quality matter? Returns to education and the characteristics of public schools in the United States, Journal of Political Economy 100, 1–40.

Chetty, R., Friedman, J.N. and Rockoff, J. (2011), The long-term impacts of teachers: Teacher value-added and student's outcomes in adulthood, unpublished mimeo.

Chingos, M.M. and West, M.R. (2010), Do more effective teachers earn more outside of the classroom?, CESifo Working Paper 2996.

Clotfelter, C.T., Glennie, E., Ladd, H.F. and Vigdor, J.L. (2006), Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina, NBER Working Paper 12285.

Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2007), How and why do teacher credentials matter for student achievement?, NBER Working Paper 12828.

Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2010), Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects, Journal of Human Resources 45, 655–681.

Corcoran, S., Evans, W. and Schwab, R. (2004), Changing labor-market opportunities for women and the quality of teachers 1957–2000, American Economic Review Papers and Proceedings 94, 230–235.

Daly, A.J., Moolenaar, N.M., Bolivar, J.M. and Burke, P. (2010), Relationships in reform: The role of teachers' social networks, Journal of Educational Administration, 48, 359–391.

Danielson, Charlotte. (1996). *Enhancing Professional Practice: A Framework for Teaching*. Alexandria, Va.: Association for Supervision and Curriculum Development.

Deming, D., Hastings, J., Kane, T. and Staiger, D. (2011), School choice, school quality and academic achievement, unpublished mimeo.

Donaldson, M.L. (2009), So long, Lake Wobegon? Using teacher evaluation to raise teacher quality, report, Center for American Progress.

Figlio, D.N. (2006), Testing, crime and punishment, Journal of Public Economics 90, 837–851.

Figlio, D.N. and Kenny, L.W. (2007), Individual teacher incentives and student performance, Journal of Public Economics 91, 901–914.

Foster, A. and Rosenzweig, M. (1994), A test for moral hazard in the labor market: Contractual arrangements, effort, and health, Review of Economics and Statistics 76, 213–227.

Fredriksson, P. and Öckert, B. (2008), The supply of skills to the teacher profession, mimeo, Uppsala University.

Fyer, R. (2011a), Financial incentives and student achievement: Evidence from randomized trials, Quarterly Journal of Economics, forthcoming.

Fryer, R. (2011b), Teacher incentives and student achievement: Evidence from New York City public schools, Harvard University, Cambridge, MA.

Glewwe, P., Ilias, N. and Kremer, M. (2008), Teacher Incentives, Harvard University Press, Cambridge, MA.

Goldhaber, D. and Hansen, M. (2008), Is it just a bad class? Assessing the stability of measured teacher performance, CRPE Working Paper 2008-5.

Goodman, S. and Turner, L. (2009), Group incentives for teachers: The impact of the NYC school-wide bonus program on educational outcomes, Columbia University Department of Economics Discussion Paper 0910-05.

Hanushek, E. A., Kain, J. F., Rivkin, S. G., (2005). Teachers, Schools, and Academic Achievement. *Econometrica , 73* (2), 417-458.

Harris, D. and Sass, T. (2007), Teacher training, teacher quality, and student achievement, Calder Working Paper 3.

Heckman, J., Layne-Farrar, A. and Todd, P. (1995), Does measured school quality really matter? An examination of the earnings-quality relationship, NBER Working Paper 5274.

Hensvik, L. (2010), Employer competition and wages: Lessons learned from Swedish teachers, Working Paper Uppsala University.

Holmstrom, B. and Milgrom, P. (1991), Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design, Journal of Law, Economics, and Organization 7, 24–52.

Hoxby, C.M. and Leigh, A. (2004), Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States, American Economic Review Papers and Proceedings 94, 236–240.

Jackson, C.K. (2009), Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation, Journal of Labor Economics 27, 213–256.

Jackson, C.K. (2010a), A little now for a lot later: A look at a Texas Advanced Placement Incentive Program, Journal of Human Resources 45, 591–639.

Jackson, C.K. (2010b), The effects of an incentive-based high-school intervention on college outcomes, NBER Working Paper 15722.

Jackson, C.K. (2010c), Match quality, worker productivity, and worker mobility: Direct evidence from teachers, NBER Working Paper 15990.

Jackson, C.K. (2011a), Do high-school teachers really matter? The need to account for tracking, unpublished mimeo.

Jackson, C.K. (2011b), High-school teachers and non-test score outcomes?, unpublished mimeo.

Jackson, C.K. and Bruegmann, E. (2009), Teaching students and teaching each other: The importance of peer learning for teachers, American Economic Journal: Applied Economics 1, 85–108.

Jacob, B.A. (2005), Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools, Journal of Public Economics 89, 761–796.

Jacob, B.A., Kane, T.J., Rockoff, J.E. and Staiger, D.O. (2011), Can you recognize an effective teacher when you recruit one? Education Finance and Policy 6, 43–74.

Jacob, B. and Lefgren, L. (2004), The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago, Journal of Human Resources 39, 50–79.

Jacob, B.A. and Lefgren, L. (2008), Can principals identify effective teachers? Evidence on subjective performance evaluation in education, Journal of Labor Economics 26, 101–136.

Jacob, B.A. and Levitt, S.D. (2003), Rotten apples: An investigation of the prevalence and predictors of teacher cheating, Quarterly Journal of Economics 118, 843–877.

Kane, T.J. and Staiger, D.O. (2008), Estimating teacher impacts on student achievement: An experimental evaluation, NBER Working Paper 14607.

Koedel, C. (2008). Teacher Quality and Dropout Outcomes in a Large, Urban School District. *Journal of Urban Economics , 64* (3), 560-572.

Koedel, C. (2009), An empirical analysis of teacher spillover effects in secondary school, Economics of Education Review 28, 682–692.

Koedel, C. and Betts, J. (2009), Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique, Department of Economics, University of Missouri Working Paper 0902.

Krueger, A.B. (1999), Experimental estimates of education production functions, Quarterly Journal of Economics 114, 497–532.

Ladd, H. (2011), Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? Educational Evaluation and Policy Analysis 33, 235–261.

Lavy, V. (2002), Evaluating the effect of teachers' group performance incentives on students' achievements, Journal of Political Economy 110, 1286–1318.

Lavy, V. (2009), Performance pay and teachers' effort, productivity, and grading ethics, American Economic Review 99, 1979–2011.

Lavy, V. (2011), What makes an effective teacher? Quasi-experimental evidence, unpublished mimeo.

Lazear, E. (2000), Performance pay and productivity, American Economic Review 90, 1346–1361.

Leigh, A. and Ryan, C.A. (2006), How and why has teacher quality changed in Australia?, Australian Economic Review 41, 141–159.

Loeb, S. and Page, M. (2000), Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation, Review of Economics and Statistics 82, 393–408.

Milanowski, A.T. and Odden, A.R. (2007), A new approach to the cost of teacher turnover, School Finance Redesign Project Working Paper 13.

Muralidharan, K. and Sundararaman, V. (2009), Teacher performance pay: Experimental evidence from India, NBER Working Paper 15323.

Murnane, R. and Cohen, D. (1986), Merit pay and the evaluation problem: Why most merit pay plans fail and few survive, Harvard Education Review 56, 1–17.

Neal, D. and Schanzenbach, D.W. (2011), Left behind by design: Proficiency counts and test-based accountability, Review of Economics and Statistics, forthcoming.

Nickell, S.J. and Quintini, G. (2002), The consequences of the decline in public sector pay in Britain: A little bit of evidence, Economic Journal 112, F107–F118.

Papay, J.P. and Kraft, M. (2010), Do teachers continue to improve with experience? Evidence of long-term career growth in the teacher labor market, Harvard University Working Paper.

Parsad, B., Lewis, L., Farris, E. and Greene, B. (2001), Teacher preparation and professional development 2000, National Center for Educational Statistics Working Paper 2001-088.

Podgursky, M. (2007), Teams versus bureaucracies: Personnel policy, wage-setting, and teacher quality in traditional public, charter, and private schools, in M. Berends et al. (eds.), Charter School Outcomes, Lawrence Erlbaum Associates, Mahwah, NJ.

Rivkin, S.G., Hanushek, E.A. and Kain, J.F. (2005), Teachers, schools, and academic achievement, Econometrica 73: 417–458.

Rockoff, J. (2004), The impact of individual teachers on student achievement: Evidence from panel data, American Economic Review, Papers and Proceeding, 94(2): 247–252.

Rockoff, J. (2008), Does mentoring reduce turnover and improve skills of new employees? Evidence from teachers in New York City, unpublished mimeo.

Rockoff, J. and Speroni, C. (2011), Subjective and objective evaluations of teacher effectiveness, Labour Economics, forthcoming.

Rockoff, J. and Staiger, D. (2010), Searching for effective teachers with imperfect information, Journal of Economic Perspectives 24(3): 97–118.

Rothstein, J. (2010), Teacher quality in educational production: Tracking, decay, and student achievement, Quarterly Journal of Economics 125, 175–214.

Roy, A.D. (1951), Some thoughts on the distribution of earnings, Oxford Economic Papers 3, 135–146.

Sass, H., Hannaway, J., Xu, Z., Figlio, D. and Feng L. (2010), Value added of teachers in high poverty schools and lower poverty schools, Working paper, The Urban Institute.

Sass, T.R. (2008), The stability of value-added measures of teacher quality and implications for teacher compensation policy, policy brief, Urban Institute, Center for Analysis of Longitudinal Data in Education Research, Washington, DC.

Sojourner, A.J., West, K.L. and Mykerezi, E. (2011), When does teacher incentive pay raise student achievement? Evidence from Minnesota's Q-COMP program, unpublished mimeo.

Springer, M.G., Ballou, D., Hamilton, L.S., Le, V.-N., Lockwood, J.R., McCaffrey, D.F., Pepper, M. and Stecher, B.M. (2010), Teacher pay for performance: Experimental evidence from the project on incentives in teaching, Working Paper, CALDER.

Strong, M. (2005), Induction, Mentoring and Teacher Retention: A Summary of the Research, ATE Commission on Mentoring and Teacher Induction and the New Teacher Center, University of California, Santa Cruz.

Taylor, E.S. and Tyler, J.H. (2011), The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers, NBER Working Paper 16877.

Todd, P. and Wolpin, K. (2006), On the specification and estimation of the production function for cognitive achievement, Economic Journal 113, F3–F33.

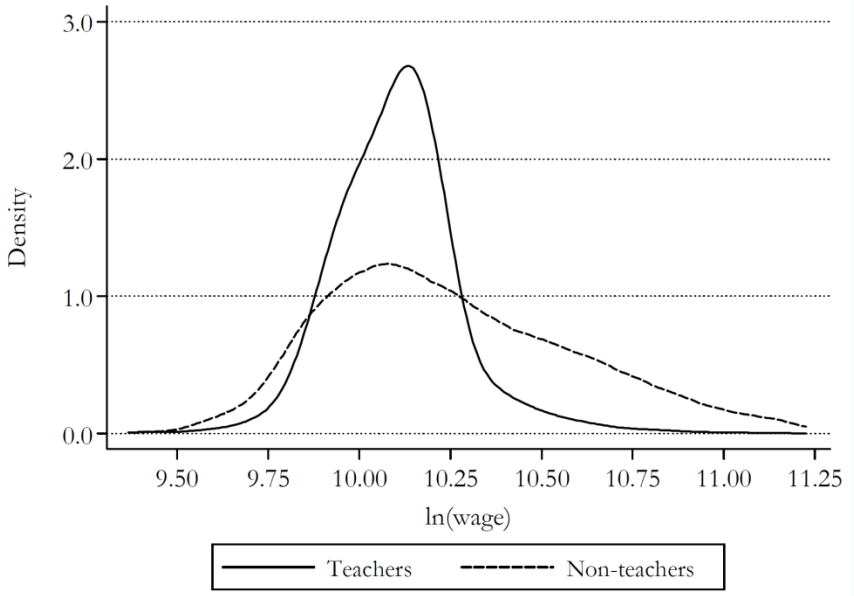
Vigdor, J.L. (2008), Teacher salary bonuses in North Carolina, Working Paper, The Urban Institute.

Weisberg, D., Sexton, S., Mulhern, J. and Keeling, D. (2009), The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness, New Teacher Project, New York.

Wiswall, M. (2011), The dynamics of teacher quality, http://ssrn.com/abstract=1911309.

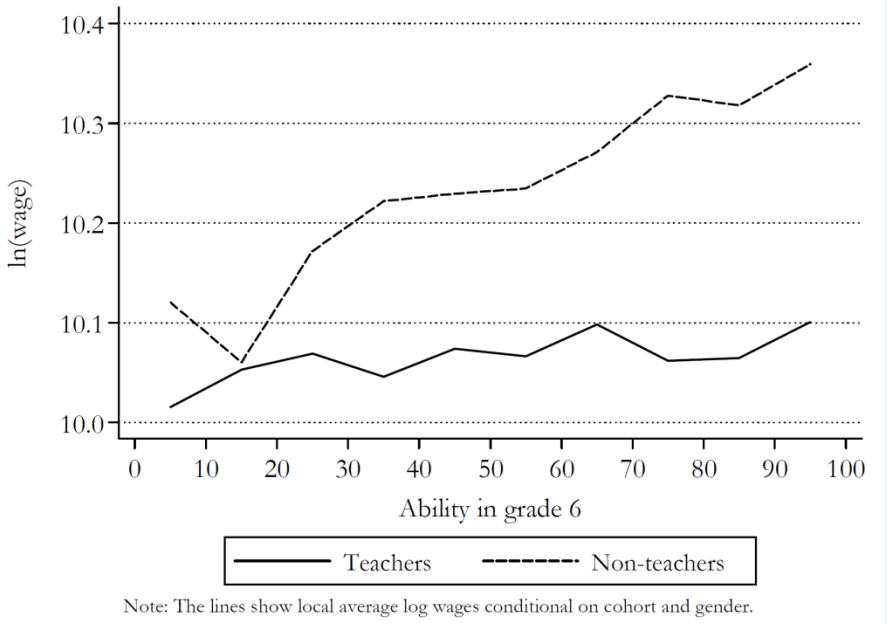
Woessmann, L. (2011), Cross-country evidence on teacher performance pay, Economics of Education Review 30, 404–418.

Figure 1: *Wage distributions for university graduates, by teacher education*
(From Fredriksson and Öckert (2008)



Notes: The figure shows smoothed Kernel-density plots using wages in 2004.

Figure 2: *The returns to ability for university graduates, by teacher education*
(From Fredriksson and Öckert (2008)



Note: The lines show local average log wages conditional on cohort and gender.

Figure 3: *Illustration of a simplified Roy Model*