



## The economic value of higher teacher quality<sup>☆</sup>

Eric A. Hanushek<sup>a,b,c,d,\*</sup>

<sup>a</sup> Hoover Institution, Stanford University, Stanford, CA 94305-6010, United States

<sup>b</sup> University of Texas at Dallas, Dallas, TX, United States

<sup>c</sup> National Bureau of Economic Research (NBER), Cambridge, MA United States

<sup>d</sup> CESifo, Munich, Germany

### ARTICLE INFO

#### Article history:

Received 15 December 2010

Accepted 16 December 2010

#### JEL classification:

H4

I2

J3

#### Keywords:

Performance pay

Teacher labor markets

Salaries

### ABSTRACT

Most analyses of teacher quality end without any assessment of the economic value of altered teacher quality. This paper combines information about teacher effectiveness with the economic impact of higher achievement. It begins with an overview of what is known about the relationship between teacher quality and student achievement. This provides the basis for consideration of the derived demand for teachers that comes from their impact on economic outcomes. Alternative valuation methods are based on the impact of increased achievement on individual earnings and on the impact of low teacher effectiveness on economic growth through aggregate achievement. A teacher one standard deviation above the mean effectiveness *annually* generates marginal gains of over \$400,000 in present value of student future earnings with a class size of 20 and proportionately higher with larger class sizes. Alternatively, replacing the bottom 5–8 percent of teachers with average teachers could move the U.S. near the top of international math and science rankings with a present value of \$100 trillion.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

It has become widely accepted that high quality teachers are the most important asset of schools, but this recognition has not led to any consensus on the appropriate policies that should be followed to ensure that we have a good stock of teachers. The policy proposals range quite broadly, although generally they call either for closer regulation of quality or for more use financial incentives with little in between these two poles. Remarkably, these

policy deliberations seldom include even the most rudimentary economic analyses or evaluations. The focus of most educational policy research and of the majority of public discussions of school policy is simply whether or not some school input has a significant positive impact on student achievement and not what it might cost or the economic benefits it might produce.<sup>1</sup> This paper focuses on the demand side of the teacher labor market in the United States and provides baseline estimates of the economic value of improving teacher quality.

Much of the discussion about the potential demand for teachers is framed in terms of ensuring sufficient

<sup>☆</sup> This paper benefitted from extensive and insightful comments by Martin West. Valuable research assistance was provided by Lorra de la Paz. This is a revised version of a paper originally prepared for the Conference on “Merit Pay: Will it Work? Is it Politically Viable?” sponsored by Harvard’s Program on Education Policy and Governance, Taubman Center on State and Local Government, Harvard’s Kennedy School, 2010.

\* Corresponding author at: Hoover Institution, Stanford University, Stanford, CA 94305-6010, United States. Tel.: +1 650 736 0942.

E-mail address: [hanushek@stanford.edu](mailto:hanushek@stanford.edu)

<sup>1</sup> One notable exception is the long term emphasis by Henry Levin and his co-authors on comparing benefits and costs, although this has not developed much traction in policy debates. See, for example, Levin and McEwan (2001) and Belfield and Levin (2007). An early attempt at benefit-cost analysis in the case of class size reduction is found in Krueger (2002), following a conceptually similar approach to one part of the analysis below.

numbers of trained teachers. This, however, is not really the issue, because the U.S. has have for a long time trained considerably more teachers than the number of positions that annually become open in schools. For example, in 2000 86,000 recent graduates entered into teaching, even though 107,000 graduated with an education degree the year before (see Provasnik & Dorfman, 2005; U.S. Department of Education, 2009).<sup>2</sup> At the same time, many have noted shortages of teachers in particular geographic regions or subject areas, such as math, science, or special education.<sup>3</sup> What usually gets left out is anything to do with quality.<sup>4</sup>

The analysis presented below is built on a simple premise: the key element defining a school's impact on student achievement is teacher quality. In turn, the demand for teacher quality is derived from just this impact of teachers on student outcomes. Both existing academic studies and the related policy discussions devote little thought to the economic value of outcomes, generally relying on the vague notion that higher scores on tests are better than lower scores. This analysis puts outcome gains in economic perspective.

Consideration of the economic value of teacher quality is especially relevant for the debates about performance pay for teachers and administrators. Until recently, teacher salary policies have given low priority to any consideration of merit pay, generally viewing it as a small add-on to salaries. As such, it was generally viewed as an add-on to the basic pay system and one of the first items to be eliminated at any sign of budget pressure (see, for example, Cohen & Murnane, 1986). In this context, it is useful to understand the value of keeping high quality teachers as it provides some bounds on the funding of potential pay policies to attract and retain effective teachers.

Estimates of the relevant achievement and pay-off parameters are in fact available in the literature. Moreover, the key parameters have been consistently estimated across different studies and with considerable precision. The innovation of this paper is to draw on those parameter estimates to produce plausible ranges for the underlying demand for teacher quality.

This paper begins with a general research overview of what is known about the relationship between teacher quality and student achievement. It then discusses the policy alternatives that have been pursued as a motivation for considering performance based policies. The central part of the paper is consideration of the derived demand for teachers that comes from their impact on economic outcomes.

## 2. The central importance of teachers

Literally hundreds of research studies have focused on the importance of teachers for student achievement. Two key findings emerge. First, teachers are very important; no other measured aspect of schools is nearly as important in determining student achievement. Second, it has not been possible to identify any specific characteristics of teachers that are reliably related to student outcomes. Understanding these findings is central to the subsequent discussions of policies and their underlying economics.

The general finding about the importance of teachers comes from the fact that the average gains in learning across classrooms, even classrooms within the same school, are very different.<sup>5</sup> Some teachers year after year produce bigger gains in student learning than other teachers. The magnitude of the differences is truly large, with some teachers producing 1.5 years of gain in achievement in an academic year while others with equivalent students produce only 1/2 year of gain.<sup>6</sup> In other words, two students starting at the same level of achievement can know vastly different amounts at the end of a single academic year due solely to the teacher to which they are assigned. If a bad year is compounded by other bad years, it may not be possible for the student to recover.

No other attribute of schools comes close to having this much influence on student achievement. The available estimates for, say, class size reduction do not suggest any leverage past the earliest grades of school, and then the expected effects are small. Most of the controversy about the impacts of class size focus on whether there is any discernible impact on student achievement (see, for example, Ehrenberg, Brewer, Gamoran, & Willms, 2001; Hanushek, 1999; Mishel & Rothstein, 2002). Much less attention has focused on the magnitude of any effects or the costs of any reductions. The largest commonly available estimate for class size reduction comes from Project STAR, where Krueger (1999) estimates that there is a one-time increase in achievement of approximately 0.22 standard deviations from reducing class size by one-third. Clearly, reducing class size by this much (eight students per class in the Tennessee STAR experiment) would be very expensive, and the comparison must be to alternative policy approaches.<sup>7</sup> These issues come back into the discussion below.

The related issue is what makes for an effective or ineffective teacher. The extensive research addressing this has found little that consistently distinguishes among teachers in their classroom effectiveness. Most documented has been the finding that master's degrees bear no consistent relationship with student achievement (see Hanushek

<sup>2</sup> Note that the recently graduated group entering teaching also includes a number of people who graduate with degrees other than in education, making the excess supply of education graduates even larger. Similar differentials existed throughout the 1990s, implying that the stock of trained teachers not in the teaching profession is substantial.

<sup>3</sup> Interestingly, these discussions have occurred over a long period of time. See Kershaw and McKean (1962).

<sup>4</sup> The one possible exception is attention to teachers lacking full certification. On the other hand, teacher certification has not been shown to be closely related to student achievement; see, for example, Goldhaber and Brewer (2000), Kane et al. (2008), and Boyd et al. (2008).

<sup>5</sup> A summary and evaluation of value-added studies that look at the influence of teachers on achievement gains is found in Hanushek and Rivkin (2010b) and discussed more thoroughly below.

<sup>6</sup> Hanushek (1992) finds differences of this magnitude for disadvantaged students found in Gary, Indiana. For an overview of the results of similar studies, see Hanushek and Rivkin (2010b).

<sup>7</sup> Krueger (1999) provides an economic analysis of class size reduction, suggesting that it is marginally cost-effective. These estimates can be compared to the estimates for teacher quality provided below.

& Rivkin, 2004, 2006). But other findings are equally as interesting and important. The amount of experience in the classroom – with the exception of the first few years – also bears no relationship to performance. On average, a teacher with five years experience is as effective as a teacher with 25 years of experience. But, this general result about measured characteristics of teachers goes even deeper. When studied, most evidence indicates that conventional teacher certification, source of teacher training, or salary level are not systematically related to the amount of learning that goes on in the classroom. For example, two recent high quality studies of different preparation and entry routes into teaching compare the impact on student achievement of Teach for America (TFA) and other alternative routes into teaching with traditional teacher training (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006; Kane, Rockoff, & Staiger, 2008).<sup>8</sup> They find little differences by teacher training background.

The exception as noted is that during the first one to three years of teacher experience the typical teacher will get better at her job (see, for example, Staiger & Rockoff, 2010). She will develop her craft, learn her tasks, learn classroom management, and find ways to help students learn. The existing evidence does not, however, suggest any clear way to provide this experience before entry into the classroom or to reduce the adjustments that will need to be made once in the classroom. For example, changes in teacher preparation or more extensive induction and mentoring programs, while plausibly effective policies, have yet to be shown to significantly alter the early-career learning of teachers.<sup>9</sup> Similarly, even very intensive professional development to help teachers become more effective after they are already in the classroom has shown little impact on student achievement.<sup>10</sup>

The fundamental impact of teacher quality has led many policy makers to focus on the implications for the form and level of pay of teachers. The most common expression of this focus is that we should pay sufficient salaries to ensure that there are high quality teachers in all classrooms.<sup>11</sup> Higher salaries are viewed as a way of providing incentives to attract and retain the teacher force that we need as a country. This conclusion, however, is not as straightforward as it sounds, given the current structure of schools and labor markets for teachers and given limitations on understanding of how to identify effective teachers.

<sup>8</sup> Teach for America has attracted a large amount of attention for its ability to attract very high quality entrants from the most competitive U.S. colleges and universities (see below).

<sup>9</sup> See the experiments in Isenberg et al. (2009). This study compares a comprehensive induction and mentoring program to “business as usual” and finds no gains.

<sup>10</sup> Garet et al. (2008) and Garet et al. (2010) present results of randomized evaluations of intensive professional development programs for early reading and for middle school math, respectively. Neither finds that this professional development even when combined with classroom coaching significantly affects student learning.

<sup>11</sup> In an international context, see for example Dolton and Marcenaro-Gutierrez (2011).

### 3. Problems with current teacher policies

Existing policies on the demand side of the teacher labor market combine regulations on entry into the profession and political bargaining over salaries. These policies have not proved effective in ensuring uniformly high quality teachers, and there is little evidence that the extensions of these that have been proposed will be more effective.

Most discussion of teacher policies includes very limited attention to costs except where salary is itself the focus of discussion. In these latter discussions of salaries, moreover, little attention is given to teacher effectiveness. Performance pay has not received extensive evaluation, because the limited use of such schemes has until recently constrained researchers’ ability to study alternative pay systems.

Information about alternative performance pay approaches is beginning to accumulated. The use of performance pay has been increasing in the United States (Podgursky & Springer, 2007). This increased interest has also led both to experimental studies (Springer et al., 2010) and new research efforts. Additionally, international experimentation has led to more recent analyses (see, for example, Lavy, 2009; Muralidharan & Sundararaman, 2009). For a direct cross-country analysis assessing the impact of different institutions, see Woessmann (2011). Given the level of activity in this area, information about the supply side of the market is likely to increase rapidly, but for now still remains rudimentary.

#### 3.1. Current certification policies

Today’s teacher policies begin almost everywhere with regulating who can enter teaching through a certification process that specifies entry requirements.<sup>12</sup> Yet, existing evidence indicates that using regulatory approaches to obtaining “good teachers” is extraordinarily difficult, if not impossible. The analyses of teacher characteristics give us little reason to believe that we know enough about consistent characteristics or backgrounds of good teachers to set appropriate training and hiring standards. Specifically, the underlying idea behind most certification requirements is that we can ensure that nobody gets a bad teacher, i.e., that is possible to put a floor on quality. But doing this requires knowledge of characteristics that systematically affect performance. As noted above, however, credentials, degrees, experience, and even teacher test scores are not consistently correlated with teaching skill.<sup>13</sup> Thus, requirements that only fully certified teachers can enter the classroom – such as included in federal accountability legislation (No

<sup>12</sup> For a general discussion of teacher licensing and certification, see Goldhaber (2011).

<sup>13</sup> Some attention has focused on teacher test scores as a potential indicator of teacher quality. The existing empirical evidence suggests that these scores, when available, are more correlated with student achievement than other explicit measures of teacher characteristics (Hanushek, 2003). Nonetheless, the relationship is modest with less than half of the estimated parameters of teacher tests on student outcomes being statistically significant and with little of the underlying variation in teacher effectiveness being accounted for by test scores.

Child Left Behind Act or NCLB) – may have little impact on student performance, even if achieved.

Nonetheless, some policy proposals argue for strengthening credentials, i.e., for making the standards higher and more rigorous. For example, some propose ensuring that *all* certified teachers have a master's degree (National Commission on Teaching and America's Future, 1996). Indeed, five states including New York currently require teachers to have an earned master's degree in order to receive permanent certification; an additional 11 states require master's for optional advanced licenses (National Council on Teacher Quality, 2009). But, since past evidence shows teachers with a master's degree are not on average more effective than teachers lacking such a degree, these proposals would raise the cost of becoming a teacher without a strong expectation that quality would improve.

Heightened screens for entry into teaching are also likely to be very costly. Tightening up on entry requirements typically involves increased undergraduate course requirements, perhaps requirements for a master's degree before entering into teaching, and higher test score requirements for entry into teacher training or for certification. Each of these makes entry into teaching relatively more costly. Other things equal, this would reduce the supply of potential teachers by altering choices among existing potential teachers. Of course salary increases could offset any alterations in supply. The magnitude of the needed increase would depend on the responsiveness of prospective teachers to salary changes—something about which we currently have only rudimentary knowledge.

A variety of experiments with alternative routes to teaching have gone beyond traditional certification. Indeed, a large portion of current teachers do not come through traditional teacher training institutions (Walsh & Jacobs, 2007). While these programs and their requirements vary widely, they do open the door to a broader range of individuals. The existing evidence on their success or failure is nonetheless limited. Some evidence does come from the Teach for America (TFA) program that recruits high performing graduates from generally elite colleges and universities and places them in hard-to-staff schools even though they lack traditional preparation and certification.<sup>14</sup> Three careful studies of the performance of the TFA teachers show generally positive results for math and equal results for reading when compared to traditionally trained teachers (see Decker, Mayer, & Glazerman, 2004; Raymond, Fletcher, & Luque, 2001; Xu, Hannaway, & Taylor, 2009). Further, Boyd et al. (2006) and Kane et al. (2008) compare New York City teachers who entered through different routes including Teach for America and other alternative certification programs and find little average difference in effectiveness.<sup>15</sup> These results do, however, cast further doubt on the idea of improv-

ing teacher quality by building on the existing certification requirements.

### 3.2. Current salary policies

The second important aspect of teacher labor markets is compensation. Standard economic models suggest that, in a competitive economy, individuals' salaries will generally be aligned to their productivity. If one firm does not pay a worker a salary that matches her value in terms of output, another competitive firm would pay that amount. If the firm pays the worker too much for her value, it will not be competitive with other firms and will be prone to going out of business.

Teacher labor markets, however, differ. Salaries are determined by collective bargaining between teachers' organizations and their employing school districts. School districts are not prone to going out of business if they pay the wrong amount.<sup>16</sup> Being public activities, schools are always subject to political forces, and the goals for school quality depend on governmental decision making about how much quality is desired. As a result, teacher salary decisions are only partially driven by the economic forces that underlie salary determination in private, competitive industries. Demand for school quality is filtered through the political process and may or may not accurately represent the true benefits and costs of varying amounts of quality, including possible externalities from quality.<sup>17</sup>

Perhaps the most notable recent pattern in teacher salaries is that they have fallen dramatically in relation to the rest of the economy. The changing position of teachers is clear in salary trends since the beginning of World War II. Compared to the earnings of male college graduates, the average male teacher was slightly above the 50th percentile in 1940. The average female teacher was close to the 70th percentile among college-educated females. But then male teachers fell precipitously to the bottom third of the earnings distribution for college graduates, and female teachers were below average during the 1960s and close to the relative male position by 1990. In 2000, less than 30 percent of young males and less than 40 percent of young females with a bachelor's degree earned less than the average teacher (age 20–29).<sup>18</sup> (Note, however, that these data do not include deferred compensation for teachers, which appears to have grown faster in teaching than in private sector employment; Podgursky, 2011.)

This decline in salaries is mirrored by other measures. While it is somewhat difficult to trace general measures of achievement and ability over time, it appears that teachers are drawn from deeper in the group of people going to

<sup>14</sup> For a description of the program, see <http://www.teachforamerica.org/>.

<sup>15</sup> Kane et al. (2008) additionally provide some idea of the cost trade-off of different teacher selection policies. Staiger and Rockoff (2010) evaluate much more stringent retention programs based on value-added information.

<sup>16</sup> Additionally, there are questions about whether the negotiations between school boards and teacher unions are arms-length transactions. See the various discussions in Howell (2005) and particularly Moe (2005).

<sup>17</sup> The public demand considerations are also influenced by the fact that aggregate importance of school quality appears to exceed the private benefits – probably reflecting externalities through the development of ideas that influence economic growth. See Hanushek and Woessmann (2008, 2011).

<sup>18</sup> Hanushek and Rivkin (2006).

college and that the best college graduates are not the ones going into teaching.<sup>19</sup>

Some analysts have focused on the current position of teachers in the wage/ability distribution as a fundamental issue driving student outcomes. Barber and Mourshed (2007), for example, identify the fact that Finland and South Korea attract top graduates into teaching as a key ingredient to their success on international tests. The source of this recruiting success is, however, unclear, because Finnish teachers are not paid high salaries (though Korean teachers are). Mourshed, Chijioke, and Barber (2010) further emphasize the importance of high level recruitment in their analysis of school systems around the world that have shown substantial improvement.

The costs of returning U.S. teachers to their former position in the salary structure are clearly enormous. While some have suggested that this would be a reasonable policy, no analysis is available to indicate what gains in teacher quality would result.

Two factors appear to be important in evaluating these salary trends. First, by most accounts the skills needed to be an effective teacher are not necessarily those needed to be successful elsewhere in the economy.<sup>20</sup> While there is uncertainty in this statement because we do not have any clear description of what skills are needed to be an effective teacher, we do not find for example that standard measures of teacher achievement or ability are closely related to student outcomes even though it is closely related to earnings elsewhere in the economy.<sup>21</sup>

Second, the current structure of teacher pay enters into the political determination of salaries and appears to hold down teacher salaries. The single salary structure that pays all teachers (with the same experience and degree level) the same amount quite likely acts to restrain the pay of teachers. The argument from a very simple political economy model, while not fully tested, is straightforward. Since salary contracts are negotiated politically with little discipline from the market, politicians negotiating salaries must be able to defend the idea that salary increases are related to improved student outcomes. But, under the single salary schedule, teachers (with the same experience and graduate training) receive the same pay and increase, regardless of the teacher's effectiveness. This situation makes large salary increases difficult because the factors that determine pay are unrelated to teacher effectiveness in the classroom, and effective teachers would see the same salary increases as ineffective teachers.

<sup>19</sup> Hanushek and Pace (1995), Corcoran, Evans, and Schwab (2004), Bacolod (2007).

<sup>20</sup> The relative earnings of teachers who leave teaching have not been thoroughly investigated. Scafidi et al. (2006) suggest that teachers do not in general move to higher paying jobs when they leave teaching. On the other hand, Chingos and West (2009) find that more effective teachers are likely to leave for better jobs.

<sup>21</sup> The relationship between measures of teacher test scores and student achievement proves to be very imprecise, with teacher tests generally being statistically insignificant in educational production functions and with variations in teacher tests explaining little of the variance in teacher effectiveness; see Hanushek (2003). At the same time, as discussed below, there are strong rewards to individual achievement in the general labor market.

Certainly higher levels of salaries would tend to increase the pool of potential teachers, but the impact of that on overall teacher quality depends on the ability of principals and human resource teams in districts to choose the best teachers. Existing evidence, while not definitive, suggests that schools are not very effective at choosing the best teachers among the pool of eligible applicants (Ballou, 1996; Ballou and Podgursky, 1997; Staiger and Rockoff, 2010).

A variant of policy discussions about salaries is to argue that the right way to set the salaries of teachers is to use the market salary for professionals in the open economy as a guide. At its heart this is simply one notion of how to determine a salary level, but it is generally unrelated to any arguments about the productivity of people in alternative occupations or about the relative attractiveness of different occupations. To the extent that the overall compensation levels of teachers would be raised by such calibration to other professions without affecting the distribution of pay across teachers, it is subject to the same questions as the previous arguments about restoring relative pay.

There is one aspect of this that has specific relevance, however. It is unclear precisely which professional occupations would provide the appropriate comparison. If the standard is privately employed professionals – say, lawyers, doctors, and accountants in private employment – a feature of the comparison is the overall structure of employment. Most private professionals have their salaries set much more in line with their individual productivity, so that these occupations have much large discrepancies in salaries and have noticeably higher employment risks than are found in teaching. Thus, even if the comparison set of alternative professions were clear, the appropriate way to compare salaries under different employment conditions is not.

A significant feature of virtually all existing discussions of teacher salaries, however, is the absence of any linkage to quality as seen through student outcomes. Certainly discussions of quality are used to motivate salary considerations, but little of the existing analysis is very relevant.

#### 4. What is the economic value of quality teachers?

An alternative way to think of the salaries for teachers is to consider the derived demand for quality teachers, because that is a way to assess the range of salary options that politicians might reasonably consider. The simplest way to value effective teachers is to note that the demand for teachers can be derived from the demand for their product—educated students. For the most part, teacher demand has never been evaluated in terms of the potential gains for students implied by the economic value of better performance. Such evaluations, however, provide some idea of the social value of highly effective teachers, even if they do not necessarily pinpoint the efficacy of particular salary structures.<sup>22</sup>

<sup>22</sup> Chetty et al. (2010) note the analogy to pay for CEOs, where their ability to create or destroy significant value does not dictate the optimal form of contract incentives.

#### 4.1. The demand side based on expected student earnings

Consider the economic returns to a student that follow having greater cognitive skills. In fact, these returns have been frequently estimated through empirical analyses of the earnings gains from increased skills. The most common estimation begins with a standard Mincer earnings model with the addition of a measure of cognitive skills (CS) such as:

$$\ln Y_i = \alpha_0 + rS_i + \alpha_1 \text{Exper}_i + \alpha_2 \text{Exper}_i^2 + \phi \text{CS}_i + \varepsilon_i \quad (1)$$

where  $Y_i$  is earnings of individual  $i$ ,  $S$  is school attainment,  $\text{Exper}$  is potential labor market experience, and  $\varepsilon$  is a random error. When cognitive skills are standardized to mean zero and standard deviation of one,  $\phi$  is interpreted simply as the percentage increase in annual earnings that can be attributable to a one standard deviation increase in achievement. This will understate the full impact of achievement to the extent that higher achievement leads to higher levels of schooling, but that is generally not considered.<sup>23</sup>

Three parallel U.S. studies provide very consistent estimates of the impact of test performance on earnings ( $\phi$ ) for young workers (Lazear, 2003; Mulligan, 1999; Murnane, Willett, Duhaldeborde, & Tyler, 2000). These studies employ different nationally representative data sets that follow students after they leave school and enter the labor force. When scores are standardized, they suggest that one standard deviation increase in mathematics performance at the end of high school translates into 10–15 percent higher annual earnings.<sup>24</sup>

Murnane et al. (2000) provide evidence from the High School and Beyond and the National Longitudinal Survey of the High School Class of 1972 (NLS72). Their estimates suggest that males obtain a 15 percent increase and females a 10 percent increase per standard deviation of test performance. Lazear (2003), relying on a somewhat younger sample from National Educational Longitudinal Study of 1988 (NELS88), provides a single estimate of 12 percent. These estimates are also very close to those in Mulligan (1999), who finds 11 percent for the normalized AFQT score in the National Longitudinal Study of Youth (NLSY) data. Note that these returns can be thought of as how much earnings would increase with higher skills every year throughout the persons' working career. The estimates do, however, come early in the worker's career, suggesting the impact may actually rise with experience.<sup>25</sup>

<sup>23</sup> Murnane et al. (2000) is an exception for tracing through the indirect effects. See also the discussion of the form of estimation in Hanushek and Zhang (2009).

<sup>24</sup> It is convenient to convert test scores into measures of the distribution of achievement across the population. A separate review of the normalized impact of measured cognitive skills on earnings by Bowles et al. (2001) finds that the mean estimate is only 0.07, or slightly over half of the specific studies here.

<sup>25</sup> These estimates are derived from observations at a point in time. Over the past few decades, the returns to skill have risen. If these trends continue, the estimates may understate the lifetime value of skills to individuals. On the other hand, the trends themselves could change in the opposite direction. For an indication of the competing forces over a long period, see Goldin and Katz (2008).

Altonji and Pierret (2001) consider the possibility of statistical discrimination that leads to increased returns to cognitive skills over time. Specifically, when young workers first go to an employer, it is difficult for the employer to judge the skills of the worker. Over time, the employer can more accurately assess the skills of the worker, and, if worker skills are related to cognitive skills as measured by the tests, the returns to test scores will rise with experience. Their analysis supports the idea that these estimated returns to skills could be an understatement, with the returns to cognitive skills rising and the returns to school attainment falling with labor market experience. When the model was tested across countries, however, it seemed most important for the United States but not for other countries (see Hanushek & Zhang, 2009).

In a different set of estimates using data on a sample of workers for the U.S., Hanushek and Zhang (2009) provide estimates of returns ( $\phi$ ) of 20 percent per standard deviation. One distinguishing feature of these estimates is that they come for a sample of workers throughout the career, as opposed to the prior estimates that all come from early-career earnings.<sup>26</sup>

Using yet another methodology that relies upon international test scores and immigrants into the U.S., Hanushek and Woessmann (2009) obtain an estimate of 14 percent per standard deviation. That analysis begins with a standard Mincer earnings model but estimates the returns to skills from a difference-in-differences formulation based on whether the immigrant was educated in the home country or in the United States. They find that skills measured by international math and science tests from each immigrant's home country are significant in explaining earnings within the United States.

Finally, Chetty et al. (2010) look at how kindergarten test scores affect earnings at age 25–27 and find an increase of 18 percent per standard deviation. These estimates do not control for any intervening school attainment differences but do control for a rich set of parental characteristics.

The finding that moving a standard deviation in cognitive skills yields 10–20 percent higher income may sound small, but these increments apply throughout the lifetime. In 2010, the average present value of income for full-time, full-year workers age 25–70 is \$1.16 million.<sup>27</sup> Thus, one standard deviation higher performance even at a low return of 13 percent per standard deviation amounts to over \$150,000.

These estimates of the labor market returns to higher cognitive skills can be merged with evidence about variation in teacher effectiveness to calculate the derived demand for teacher quality. The basic approach to

<sup>26</sup> The data from the International Assessment of Adult Literacy (IALS) provide both tests of reading and numeracy skills but also assess a range of adult workers. The estimates in Hanushek and Zhang (2009) come, like the previously mentioned studies, from adding cognitive skills to a standard Mincer earnings function. But that paper also discusses alternative ways to obtain estimates of the schooling gradient ( $r$  in Eq. (1)).

<sup>27</sup> Calculations use average income by age for all fulltime, full-year workers in the labor force in the first quarter of 2010. It is assumed that incomes rise 1 percent per year because of overall productivity improvements in the economy and that future incomes are discounted at 3 percent.

**Table 1**  
Estimates of within school variation in teacher effectiveness ( $\sigma_w$ ).

Study	Location	Test subject	
		Reading	Math
Rockoff (2004)	New Jersey	0.10	0.11
Nye, Konstantopoulos, and Hedges (2004)	Tennessee	0.26	0.36
Rivkin, Hanushek, and Kain (2005)	Texas	0.15	0.11
Aaronson, Barrow, and Sander (2007)	Chicago		0.13
Kane et al. (2008)	New York City	0.08	0.11
Jacob and Lefgren (2008)	Undisclosed city	0.12	0.26
Kane and Staiger (2008)	Los Angeles	0.18	0.22
Koedel and Betts (2009)	San Diego		0.23
Rothstein (2010)	North Carolina	0.11	0.15
Hanushek and Rivkin (2010a)	Undisclosed city		0.11
Average		0.13	0.17

Source: Hanushek and Rivkin (2010b).

Note: All estimates indicate the standard deviation of teacher effectiveness in terms of student achievement standardized to mean zero and variance one. All variances are corrected for test measurement error and except Kane and Staiger (2008) are estimated within school-by-year or within school-by-grade-by-year. Corrected reading estimates are included for Rivkin et al. (2005).

estimating teacher effectiveness begins with a model of student achievement ( $A$ ) for student  $i$  in grade  $g$  as a function of lagged achievement, a fixed effect for each teacher ( $\delta_j$ ), and other factors ( $X$ ) that might affect performance as in:

$$A_{it} = (1 - \theta)A_{it-1} + \delta_j + \beta X_i + v_{it} \quad (2)$$

A central issue in past estimation has been identifying the standard deviation of teacher effectiveness from variations in  $\delta_j$ . A key additional parameter is  $\theta$ , which indicates the depreciation rate on prior learning, because this indicates how much of the learning attributable to a teacher carries over after the student leaves the classroom.

Hanushek and Rivkin (2010b) review recent estimates of the standard deviation in teacher quality ( $\sigma_w$ ), and their estimates are reproduced in Table 1.<sup>28</sup> These estimates, however, look at the variations in teacher effectiveness found within schools (hence the subscript  $w$ ) and do not include any differences between schools.<sup>29</sup> The average within school variation in recent studies is 0.17 s.d. for math and 0.13 s.d. for reading. The focus on within school variance reflects a concern about identifying teacher quality as opposed to unobserved differences among students and families who have selected their school, largely through residential location.

For the subsequent estimation of the impact of teacher quality, an estimate of the total variation of quality ( $\sigma_T$ ) is used. In reality, because of difficulties in identifying the between-school variance in quality, the subsequent analysis relies on bounds the plausible values of total variation. A lower bound of 0.2 s.d. is used and is matched with a plausible upper bound of 0.3 s.d.<sup>30</sup> In this, teacher effec-

tiveness is measured in terms of standard deviations of student achievement. Thus, a teacher who is one standard deviation above the mean to the distribution of teachers in terms of quality (i.e., comparing the 84th percentile teacher to the 50th percentile teacher) is estimated to produce marginal learning gains of 0.2–0.3 standard deviations of student achievement compared to the average teacher.<sup>31</sup>

As a base case, consider a teacher who is 0.5 s.d. above average in the teacher effectiveness distribution (i.e., at the 69th percentile), and this level of effectiveness is maintained across school years. She would, according to the above estimates, annually lead to a 0.1 s.d. average improvement in cognitive skills of her students (assuming that the standard deviation of teacher effectiveness in units of student achievement is 0.2 s.d.).

The implication for earnings depends on the persistence of this learning into the future and how this increased learning in any given year carries through into the labor market. The baseline calculations presume that 70 percent of the added learning persists into the future, i.e., that  $\theta$  in Eq. (2) is 0.3. The 70 percent persistence of the annual growth in achievement comes from standard estimates of depreciation of learning in educational production functions, but this is subject to uncertainty. First, the estimates of  $\theta$  can be directly influenced by differences in tests across grades, by test measurement errors, and by other nonlearning matters. Second, while estimates of  $\theta$  are not always reported in the relevant empirical literature, there is a clear distribution of estimates in the literature.  $\theta = 0.3$  is roughly consistent with estimates in Hanushek (1971, 1992), Armor et al. (1976), and Boyd et al. (2006). Hanushek, Kain, and Rivkin (2009) find estimates closer to 0.4. Kane and Staiger (2008) find estimates of depreciation near 0.5 (with large standard errors) for math, but lower in language arts. The estimates of Jacob, Lefgren, and Sims (2010) for the teacher component of persistence are significantly lower than this

<sup>28</sup> Estimation of  $\sigma$  is actually done in a variety of ways and frequently makes some effort to eliminate biases and measurement error. See Hanushek and Rivkin (2010b).

<sup>29</sup> Note that all estimates are corrected for measurement error and, except for Kane and Staiger (2008), rely on just the within school variation in teacher effectiveness.

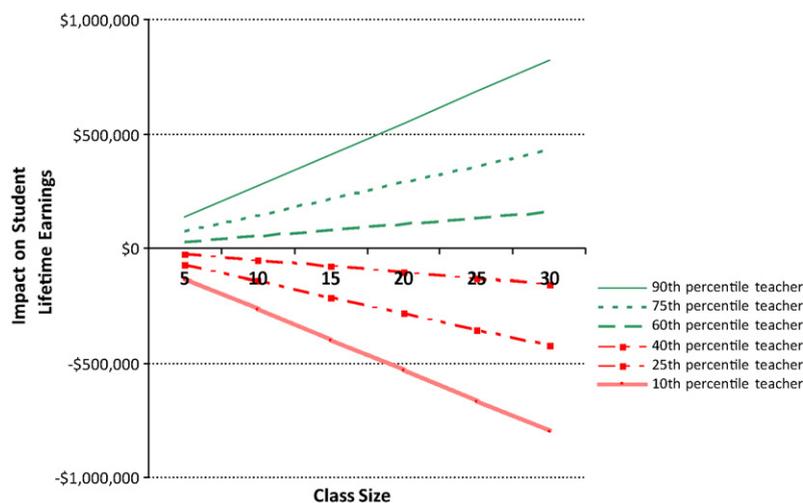
<sup>30</sup> Comparisons of estimated variations within and between schools can be found in Hanushek and Rivkin (2010a).

<sup>31</sup> In terms of the student achievement distribution, this would move a student from the 50th percentile to the 58th to 62nd percentile.

**Table 2**Baseline marginal annual economic value based on student lifetime incomes ( $\sigma_T=0.2$ ;  $\phi=0.13$ ;  $\theta=0.3$ ).

Class size	Teacher effectiveness in standard deviations from mean (percentile in parentheses)					
	0.25 (60th)	0.5 (69th)	0.75 (77th)	1.0 (84th)	1.25 (89th)	1.5 (93rd)
5	\$26,458	\$53,036	\$79,735	\$106,556	\$133,500	\$160,566
10	\$52,915	\$106,071	\$159,470	\$213,113	\$267,000	\$321,132
15	\$79,373	\$159,107	\$239,205	\$319,669	\$400,499	\$481,698
20	\$105,830	\$212,143	\$318,941	\$426,225	\$533,999	\$642,264
25	\$132,288	\$265,179	\$398,676	\$532,781	\$667,499	\$802,831
30	\$158,745	\$318,214	\$478,411	\$639,338	\$800,999	\$963,397

Note:  $\theta$ , depreciation rate;  $\sigma_T$ , standard deviation of teacher quality;  $\phi$ , labor market return to one standard deviation higher achievement.



Source: Author calculations.

**Fig. 1.** Impact on student lifetime incomes by class size and teacher effectiveness (compared to average teacher).

Source: Author calculations.

with  $\theta$  in the range of 0.7–0.8. Chetty et al. (2010) find that kindergarten scores carry through to young adult earnings – suggesting a much higher persistence of early skills – even though later test scores for students do tend to fade. As a result of this imprecision, the impact of larger depreciation than that in the baseline is also investigated in the subsequent sensitivity analysis.

It is now possible to calculate the value of an above average teacher in terms of effectiveness. Combining the improvement in scores for an individual with a conservative estimate of a impact on future individual earnings of 13 percent per standard deviation of achievement yields a present value of \$10,600 over a lifetime of work for the average worker.

But this is not yet the full impact of the above average teacher. The impact on one student is replicated across all of the other students in the class. Thus, calculation of the impact of a teacher depends directly on class size. Table 2 provides the calculated economic value of teachers at different points in the distribution and with different class sizes. Fig. 1 displays the impact of different quality teachers according to class sizes at varying percentiles of the distribution.<sup>32</sup> A teacher who is at the 60th percentile (0.25

s.d. above average) raises individual earnings by \$5292, and this translates into a present value of \$105,830 for a class size of 20 students. A teacher who is one standard deviation above the mean (84th percentile) produces over \$400,000 in added earnings for her class of twenty.<sup>33</sup>

The first thing to note is that this is an *annual* increment by the teacher. Any teacher who stays at the given level of performance produces such an amount each year.

The second thing to note from the bottom of Fig. 1 is that a below average teacher leads to a similar decrease in lifetime earnings.<sup>34</sup> Thus, having an effective teacher followed by an equally ineffective teacher will cancel out the gains.

The precise marginal economic value depends crucially on the three parameters of the teacher distribution and of how achievement evolves over time and affects earnings:

controversial and does not readily permit explicit analysis; see, for example, Hanushek (1999), Krueger (1999), Mishel and Rothstein (2002).

<sup>33</sup> Chetty et al. (2010), extrapolating from their data on early career earnings, estimate the impact of a high quality teacher at about \$214,000 per class of 20 for a teacher one s.d. above the mean. This is very close to the lower bound estimate in Table 3.

<sup>34</sup> The decrease is slightly different because the estimates come from Mincer earnings functions which relate the logarithm of earnings to the level of cognitive skills and thus to a slight different percentage change when evaluated at a different place in the distribution.

<sup>32</sup> None of these estimates introduce any possible offset that would come from direct class size effects. The magnitude of any such effects has been

$\sigma_T = 0.2$ ;  $\phi = 0.13$ ; and  $\theta = 0.3$ . It is useful to see how the specific baseline parameters affect the results when we use alternative but plausible values.

The impact of the different parameters is straightforward. A lower depreciation rate (higher persistence of achievement), a wider distribution of the teacher effectiveness distribution, and a larger labor market payoff to skill lead to a larger economic value of teacher effectiveness. All of the prior estimates were based on rather conservative estimates of  $\sigma_T$ , the variation in teacher effectiveness; one standard deviation in teacher effective translates into 0.2 standard deviations in annual student growth. As indicated, a plausible upper bound on the variations in effectiveness would be 0.3 standard deviations in annual student growth, which would be consistent both with the larger estimates in Table 1 and with a more significant between-school variation in effectiveness. Additionally, the return to skill of  $\phi = 0.13$  most closely mirrors the labor market estimates for young workers and for time periods in the past when the demand for skill was less. More recent estimates and consideration of the full age range of workers yields larger estimates, suggesting that  $\phi = 0.2$  is a plausible upper bound on the estimates. The baseline estimates do use a depreciation rate of 0.3, whereas a subset of existing production function estimates suggest larger depreciation, particularly of achievement gains induced by the teacher. We thus also look at  $\theta = 0.6$ , or a depreciation rate that is twice as large.

Table 3 presents alternative estimates of marginal impacts evaluated at one point in the teacher distribution—one standard deviation above the mean, or the 84th percentile. Compared to the baseline, a higher depreciation rate on achievement obviously lessens the impact of teacher quality on earnings, because this effectively reduces the impact of different teachers. Nonetheless, even at the lower bound in column (1) of the table defined by the previous quality and earnings parameters ( $\sigma_T$  and  $\phi$ ) but higher depreciation ( $\theta$ ), a good teacher with a class of 15 annually produces \$182,000 more in present value than the average teacher. If we scan across the marginal annual economic value of a good teacher (compared to the average) evaluated at a given class size – say 20 students per class – we see that the parameters do make a large difference in the estimated impact. The annual economic value with class size of 20 ranges from a quarter of a million dollars to a million dollars at the top of the range for the three parameters together. (The final column is an upper bound on estimates based on current empirical work.)

While the difference in estimates across the parameters is large, the more striking feature of the table is the magnitude of the lower bound. A teacher in the top 15 percent with a class of 20 or more students yields at least \$240,000 in economic surplus each and every year compared to an average teacher.

As suggested, the persistence of the annual teacher effects implied by these estimates is an open question. All of the calculations in Fig. 1 presume that 70 percent of a teacher's addition to knowledge carries over permanently (except as modified by subsequent school and family inputs). In reality, maybe all carries over, or maybe only a small part carries over. A host of unknown

factors—including compensatory behavior of parents and schools, the cumulative nature of skills, the specific attributes valued in the labor market, and the nature of peer-classroom interactions come into play in determining the long run impact of specific teachers. But even twice the depreciation of achievement that was used in the baseline yields very large estimates of the value of an effective teacher—say, \$150,000 per year present value for a 75th percentile teacher with a class of 20 students.

#### 4.2. The demand side based on aggregate economic growth

An alternative way of estimating the derived demand for effective teachers focuses on the impact of student performance on economic growth. Recent analysis has demonstrated a very close tie between cognitive skills of a country's population and the country's rate of economic growth (see the review in Hanushek and Woessmann, 2008). In particular, countries that perform better on international math and science tests have stronger growth of their economies. These analyses suggest that the aggregate impact of increased skills is noticeably larger than the individual impact from the prior calculations.<sup>35</sup>

The magnitude of the effects is truly large. For the United States, Hanushek and Woessmann (2011) calculate that the present value of increased Gross Domestic Product (GDP) from improving scores by 0.25 standard deviations would be \$44 trillion.<sup>36</sup> To get some idea of what a difference of 0.25 s.d. on the international tests means in substantive terms, it is useful to note that Canada is approximately 0.4 s.d. ahead of the U.S. and that Finland – the current world leader – is approximately 0.58 s.d. ahead.<sup>37</sup>

Now consider what would be possible if we could eliminate the bottom end of the teacher quality distribution and replace these teachers with average teachers. Following the estimates in Hanushek (2009), it is possible to bound the increases in overall performance that could be expected

<sup>35</sup> The precise reasons for the larger estimates of aggregate effects compared to the micro effects from individual earnings are not clear. These estimates are consistent with substantial externalities from higher cognitive skills, but independent estimates of these are unavailable. The macro estimates reported here assume an endogenous growth formation such that increased cognitive skills translate into permanently higher rates of long run growth in GDP per capita. An alternative neoclassical version would relate increased skills to increased factor endowments, leading to movement to a higher level of income but one with the pre-reform rate of long run growth. This latter model yields somewhat smaller estimates of the economic gains, but they remain at 70 percent of endogenous growth model and still considerably above what would be estimated from the individual earnings parameters. The alternative approaches to estimation are discussed in Hanushek and Woessmann (2008, 2011).

<sup>36</sup> The key assumptions, described in detail in Hanushek and Woessmann (2011), are that future growth follows the patterns of growth for 1960–2000, that school improvement takes 20 years and that the higher skilled people replace existing workers as they retire after a 40 year career, and that present values are calculated through 2090 using a 3 percent discount rate.

<sup>37</sup> These variations come from math performance on the 2006 tests in the Programme for International Student Assessment, or PISA (see summary data in Organisation for Economic Co-operation and Development, 2010). There are some variations in average country scores over time and across subjects, but these do not affect the calculations here.

**Table 3**  
Sensitivity of demand based on earnings to key parameters (marginal annual economic value of teacher one standard deviation above mean).

Class size	$\theta = 0.6$				$\theta = 0.3$			
	$\sigma_T = 0.2$		$\sigma_T = 0.3$		$\sigma_T = 0.2$		$\sigma_T = 0.3$	
	$\phi = 0.13$	$\phi = 0.2$						
5	\$60,652	\$93,573	\$91,215	\$140,923	\$106,556	\$164,741	\$160,566	\$248,858
10	\$121,303	\$187,145	\$182,430	\$281,847	\$213,113	\$329,482	\$321,132	\$497,715
15	\$181,955	\$280,718	\$273,645	\$422,770	\$319,669	\$494,223	\$481,698	\$746,573
20	\$242,607	\$374,290	\$364,860	\$563,693	\$426,225	\$658,964	\$642,264	\$995,431
25	\$303,259	\$467,863	\$456,075	\$704,617	\$532,781	\$823,706	\$802,831	\$1,244,288
30	\$363,910	\$561,435	\$547,290	\$845,540	\$639,338	\$988,447	\$963,397	\$1,493,146

Note:  $\theta$ , depreciation rate;  $\sigma_T$ , standard deviation of teacher quality;  $\phi$ , labor market return to one standard deviation higher achievement.

from school improvement. Using the reasonable estimates (above) of variations in teacher effectiveness as measured by achievement growth – specifically, 0.20–0.30 s.d. – it is possible to see the impact of the least effective teachers.

Fig. 2 plots the impact on overall student learning of “deselecting” (i.e., moving out of the classroom) varying proportions of ineffective teachers and replacing them with an average teacher. These calculations come from using the prior variance estimates to judge the impact of truncating the distribution. The analysis applies to all teachers, so it can be thought of improving the effectiveness of teachers throughout the system. As such, it is assumed that the quality of teachers reinforces any gains that students make and the impacts of good instruction are not assumed to die out as the student progresses to a higher grade. Instead later teachers build upon the stronger average achievement of all children and set their instruction accordingly.

The figure shows upper and lower bounds on the improvements corresponding to standard deviations of 0.3 and 0.2, respectively. The wider the distribution of teacher effectiveness the greater is the improvement from eliminating the bottom tail of the distribution. As an example, consider what would happen to average student performance if we could eliminate the least effective 5 percent of teachers from the distribution. The estimates of the impact of teachers on student achievement indicate that students would on average gain 0.28–0.42 s.d. of performance by

the end of their schooling, depending on the bounds of the teacher quality estimates.

These estimates of the importance of teacher quality permit some calculations of what would be required to yield various improvements in student performance. To begin with, consider what magnitude of teacher deselection might yield an improvement in student performance to the level of Canada (0.4 s.d. of student achievement). Fig. 2 shows that eliminating the least effective 5–8 percent of teachers would bring student achievement up by 0.4 s.d. If the upper bound on teacher effectiveness, corresponding to larger differences in effectiveness, is appropriate, replacing the bottom 8 percent of teachers with an average teacher would bring the U.S. up to the level of Finland.

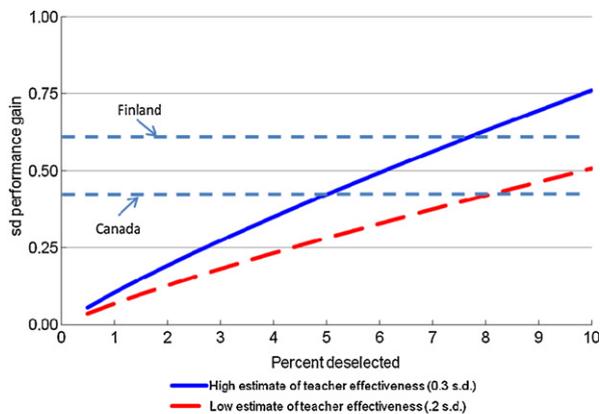
The estimates of the growth impacts of bringing U.S. students up to Finland imply astounding improvements in the well being of U.S. citizens. The present value of future increments to GDP in the U.S. would amount to \$112 trillion (Hanushek & Woessmann, 2011). These returns dwarf, for example, all of the discussions of U.S. economic stimulus packages related to the 2008 recession (\$1 trillion).

The estimates are so large for two reasons—the U.S. is currently far from Finland in achievement and the U.S. economy is very large. The increase in achievement for the U.S. would, according to historic growth patterns, lift the annual U.S. growth rate by over one percent.<sup>38</sup>

### 5. Costs and the timing of benefits

It is clear from the prior calculations that improvements in teacher effectiveness would lead to very large economic gains. The estimates of the economic gains are all put in terms of present values, but they do not accrue for some years into the future. The estimates of individual earnings gains cover the entire work life of a current student. The estimates of the economic gains to the nation consider gains across the entire lifetime for somebody born today.

But it is not appropriate to presume that these changes occur without cost. At a very simple level, if 5–10 percent

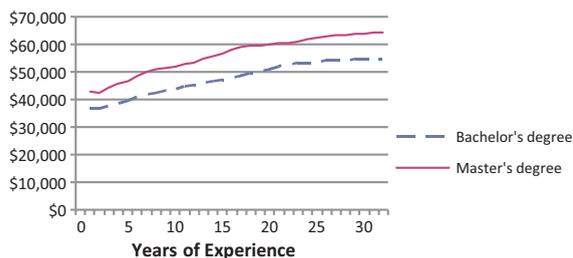


Source: Author Calculations

**Fig. 2.** Alternative estimates of how removing ineffective teacher affects student achievement.

Source: Author calculations.

<sup>38</sup> These estimates, particularly for the U.S., are sensitive to the assumptions about the form of the growth model. Under the neoclassical model, the low achievement of the U.S. is consistent with its currently being above its long run income level. The U.S. is presumed to be one of the prime contributors to the growth of the technological frontier, but the lower implied growth under this model would still yield a present value of economic improvement from achievement at the Finnish level of \$62 trillion.



Source: U.S. Department of Education (2010), Table 74

Fig. 3. Average Teacher Salary by Degree and Experience, 2007.

Source: U.S. Department of Education (2010), Table 74.

of teachers were deselected, the risk of entering a teaching career would increase, and it is natural to presume that salaries would have to rise to offset this increase in risk. More generally, it is necessary to consider how it might be possible to finance monetary incentives for altering the current teacher workforce. If there are fiscal restraints on governments, say from lowered tax revenues during recessionary periods, it would be important to find financing within the current operating budgets for schools.

The current structure of salaries for teachers pays bonuses for advanced degrees and for added teaching experience. Over time, the teachers with advanced degrees have increased as a proportion of the teacher force. Less than a quarter of all teachers having a master's degree or more in 1960, but in 2007 over half of all teachers had some sort of advanced degree.<sup>39</sup> Against this increase, as indicated previously, few studies have suggested that having a master's degree implies higher effectiveness. Similarly, median experience has progressively increased since 1960, and currently over 85 percent of teachers have more than three years of teaching experience. Again, little evidence indicates that experience after the first few years has any systematic impact on performance.<sup>40</sup>

The important thing about this increase in teacher education levels and in teacher experience is that salaries rise with these factors even though they have no systematic influence on student achievement. Fig. 3 shows average teacher salaries by degree and years of experience.<sup>41</sup> A teacher with 25 years of experience earns 35 percent more than a teacher with 5 years of experience. The average teacher with a master's degree earns 18 percent more than a teacher with just a bachelor's degree. But, neither higher levels of experience nor advanced degrees are related to teacher effectiveness.

In 2008, 9.5 percent of total teacher salaries went to bonuses for advanced degrees, while 27 percent of total salaries went for experience bonuses for teachers with greater than two years of experience. Eliminating or reducing these bonus payments for unproductive background characteristics of teachers could obviously free up

<sup>39</sup> Information on teacher degrees and experience is found in U.S. Department of Education (2010), Table 68.

<sup>40</sup> See Hanushek (2003) and, more recently, Chingos and Peterson (2011).

<sup>41</sup> The information on salaries is found in U.S. Department of Education (2010), Table 74.

substantial amounts of funds that could be re-directed toward policies to improve the quality of teachers. The national expenditure in 2007 on bonuses for advanced degrees amounted to approximately \$19 billion.<sup>42</sup> The total bonuses for teacher experience are roughly three times as large.

The larger problems may nonetheless revolve around the political costs of any reforms. The previous calculations suggest that considerable value could accrue to improving the quality of teachers. Yet the pattern of benefits imply that they are achieved far in the future, long after much of the initial costs for reform must be paid and beyond the electoral period for most politicians. Many politicians have in fact pursued school improvement, and spending on schools has risen sharply over the five decades (Hanushek & Lindseth, 2009). The policies introduced have, however, been ones that have direct benefits to current school personnel, such as reduced class size or higher overall salaries, although these policies have not been ones that have led to higher student achievement. Moving to alternate policies such as differential retention and performance pay of teachers involves greater political costs because these policies are generally not supported by the teacher unions.

## 6. Policy conclusions

The key to interpreting the prior calculations is to recognize that they flow directly from increasing teacher effectiveness. They do not flow from increased teacher salaries unless such salaries are used to attract and retain more effective teachers.

This paper has concentrated on the demand side of the teacher labor market. The underlying idea is that knowing the impact of teacher quality on economic outcomes provides immediate information about what kind of rational changes in teacher incentives and salaries are economically desirable.

Unfortunately, we know little about the supply function for teacher quality.<sup>43</sup> Thus, it is not possible to predict what kinds of pay changes would be needed to ensure any given quality of teacher force.

The standard arguments for performance pay suggest the potential value of differential pay based on effectiveness in the classroom. We actually have little empirical evidence about how to structure any such pay systems or about what the effects might be.<sup>44</sup> The evi-

<sup>42</sup> Total expenditure on instructional salaries in public schools in 2007 was \$197 billion, not counting any benefits and any degree bonuses to administrators or those providing instructional staff services. See U.S. Department of Education (2010), Table 180.

<sup>43</sup> There are actually different ways to think about the supply function of teacher quality. One can put the supply of quality into terms related to salary arguments, where selection of teachers in both hiring and retention decisions is central. On this score, no systematic research exists. Alternatively, one could relate quality to the effort made by existing teachers. This focus is central to the early work on merit pay (e.g., Murnane and Cohen, 1986), but has also been the key element of more recent evaluations such as Lavy (2002, 2009) and Muralidharan and Sundaraman (2009). See also the review of performance incentives in Lavy (2007).

<sup>44</sup> A discussion of current pay schemes can be found in Podgursky and Springer (2007). See also the various discussions in Springer (2009).

dence presented in this paper simply suggests that the economically appropriate rewards for particularly effective teachers in the context of a performance pay plan could be very large.

The foregoing analysis has also implicitly suggested an alternative approach to simple performance pay that could be more cost effective. If there is an accurate screen on teacher effectiveness, many of the properties of a performance pay scheme can be achieved by eliminating low performing teachers and paying the remaining teachers higher but relatively flat salaries.

The policy of eliminating the least effective teachers is very consistent with the McKinsey analysis of the policies found in high-performing school systems around the world (Barber & Mourshed, 2007). Their analysis suggests that the best school systems do not allow ineffective teachers to remain in the classroom for long. These conclusions are also consistent with more U.S. evidence, such as that for New York City, in Kane et al. (2008) and the related policy prescriptions in Gordon, Kane and Staiger (2008) and Staiger and Rockoff (2010).

Policies of making active performance-based decisions on retention and tenure are uncommon in the current school system. A number of states have laws and regulations that lead to tenure decisions as early as two years into a teacher's career, with the mode being just three years (National Association of State Boards of Education, 1997; National Council on Teacher Quality, 2009). On top of that, the teacher evaluation process is typically very cursory (Toch & Rothman, 2008). There is also evidence that common evaluation criteria identify very few teachers as being anything but very good (Weisberg, Sexton, Mulhern, & Keeling, 2009). These realities are inconsistent with the goal of providing a quality education to all students, because some students must necessarily be relegated to these ineffective, and damaging, teachers.

The consideration of the impact of the most ineffective teachers suggests substantial economic gains from instituting policies to identify the most ineffective teachers and to move them out of the classroom. Developing such policies, negotiating them with teachers, and implementing them in the schools would clearly take time. It would also require both severance packages for those deselected and higher pay for those who would then have a more risky job.

But there are also other policies that are suggested by the economic aspects of teacher quality. Specifically, it is important to consider the significant interaction between teacher effectiveness and class size—since all of the impacts on individuals are magnified across entire classrooms. A simple conclusion from the estimates is that, even without eliminating any teachers, the most effective teachers should be assigned larger classes and the least effective should be assigned smaller classes. In that way, the aggregate impact of less effective teachers is lessened, and the more effective teachers are better utilized. Of course, any direct impacts of altered class size would be relevant, but the existing research makes it difficult to include that in any systematic manner. Further, the more effective teachers might react badly to having larger classes, which in turn require more work. Indeed anecdotal evidence suggests that schools may try to do the opposite. If pay is

completely constrained, schools may reward the better teachers by giving them smaller classes. These concerns could be eliminated if teachers are paid a portion of their economic returns.

In the end, there is ambiguity in policy because we have never been able to effectively evaluate what the supply function for teacher quality looks like. This lack of information could, of course, be eliminated by a set of pay experiments. Unfortunately, the current negotiated pay alternatives do not seem to be providing much information—in part because they imply salaries that are relatively insensitive to effectiveness.

The bottom line remains that much higher teacher salaries would be economically justified if salaries reflected teacher effectiveness more closely. Without that linkage, we should expect our schools to underperform, and we might also expect teacher salaries to lag those in the general labor market.

## References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135.
- Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, 116(February (1)), 313–350.
- Armor, D. J., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., et al. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools*. Santa Monica, CA: Rand Corp.
- Bacolod, M. P. (2007). Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality. *Review of Economics and Statistics*, 89(November (4)), 737–751.
- Ballou, D. (1996). Do public schools hire the best applicants? *Quarterly Journal of Economics*, 111(February (1)), 97–133.
- Ballou, D., & Podgursky, M. (1997). *Teacher pay and teacher quality*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. McKinsey and Company.
- Belfield, C. R., & Levin, H. M. (Eds.). (2007). *The price we pay: economic and social consequences of inadequate education*. Washington, DC: Brookings Institution Press.
- Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: a behavioral approach. *Journal of Economic Literature*, 39(December (4)), 1137–1176.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, 1(Spring (2)), 176–216.
- Boyd, D., Lankford, H., Rockoff, J. E., Loeb, S., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, 27(Fall (4)), 793–818.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2010). *How Does Your Kindergarten Classroom Affect Your Earnings? Evidence From Project STAR*. NBER WP16381. Cambridge, MA: National Bureau of Economic Research (September).
- Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, 30(June (3)).
- Chingos, M. M., & West, M. (2009). Teacher effectiveness, mobility, and attrition in Florida. In M. G. Springer (Ed.), *Performance incentives: their growing impact on American K-12 education* (pp. 251–271). Washington, DC: Brookings.
- Cohen, D. K., & Murnane, R. J. (1986). Merit pay and the evaluation problem: understanding why most merit pay plans fail and a few survive. *Harvard Educational Review*, 56(February (1)), 1–17.
- Corcoran, S. P., Evans, W. N., & Schwab, R. S. (2004). Changing labor market opportunities for women and the quality of teachers 1957–1992. *American Economic Review*, 94(May (2)), 230–235.
- Decker, P. T., Mayer, D. P., & Glazer, S. (2004). *The effects of Teach For America on students: findings from a national evaluation*. Princeton, NJ: Mathematica Policy Research, Inc. (June 9)

- Dolton, P., & Marcenaro-Gutierrez, O. D. (2011). If you pay peanuts do you get monkeys? A cross country analysis of teacher pay and pupil performance. *Economic Policy*, 65(January).
- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest*, 2(May (1)), 1–30.
- Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. (September).
- Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Walters, K., Song, M., et al. (2010). *Middle school mathematics professional development impact study: findings after the first year of implementation, NCEE 2010-4009*. Washington, DC: Institute of Education Sciences, U.S. Department of Education. (April).
- Goldhaber, D. (2011). Licensure: exploring the value of this gateway to the teacher workforce. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 315–339). Amsterdam: North Holland.
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(Summer (2)), 129–145.
- Goldin, C., & Katz, L. F. (2008). *The race between education and technology*. Cambridge, MA: Harvard University Press.
- Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: estimation using micro data. *American Economic Review*, 60(May (2)), 280–288.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, 100(February (1)), 84–117.
- Hanushek, E. A. (1999). The evidence on class size. In S. E. Mayer, & P. E. Peterson (Eds.), *Earning and learning: how schools matter* (pp. 131–168). Washington, DC: Brookings Institution.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113(February (485)), F64–F98.
- Hanushek, E. A. (2009). Teacher deselection. In D. Goldhaber, & J. Hanaway (Eds.), *Creating a new teaching profession* (pp. 165–180). Washington, DC: Urban Institute Press.
- Hanushek, E. A., & Lindseth, A. A. (2009). *Schoolhouses, courthouses, and statehouses: solving the funding-achievement puzzle in America's public schools*. Princeton, NJ: Princeton University Press.
- Hanushek, E. A., & Pace, R. R. (1995). Who chooses to teach (and why)? *Economics of Education Review*, 14(June (2)), 101–117.
- Hanushek, E. A., & Rivkin, S. G. (2004). How to improve the supply of high quality teachers. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp. 7–25). Washington, DC: Brookings Institution Press.
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. A. Hanushek, & F. Welch (Eds.), *Handbook of the economics of education*, vol. 2 (pp. 1051–1078). Amsterdam: North Holland.
- Hanushek, E. A., & Rivkin, S. G. (2010a). *Constrained job matching: does teacher job search harm disadvantaged urban schools?* NBER w15816. Cambridge, MA: National Bureau of Economic Research (March).
- Hanushek, E. A., & Rivkin, S. G. (2010b). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2), 267–271 (May).
- Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(September (3)), 607–668.
- Hanushek, E. A., & Woessmann, L. (2009). *Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation*. NBER Working Paper 14633. Cambridge, MA: National Bureau of Economic Research (January).
- Hanushek, E. A., & Woessmann, L. (2011). *How much do educational outcomes matter in OECD countries?* Economic Policy.
- Hanushek, E. A., & Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital*, 3(Summer (2)), 107–143.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2009). New evidence about *Brown v. Board of Education*: the complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(July (3)), 349–383.
- Howell, W. G. (Ed.). (2005). *Besieged, school boards and the future of education politics*. Washington DC: The Brookings Institution.
- Isenberg, E., Glazerman, S., Bleeker, M., Johnson, A., Lugo-Gil, J., Grider, M., Dolfin, S., & Britton, E. (2009). *Impacts of comprehensive teacher induction: results from the second year of a randomized controlled study*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. (August).
- Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(January (1)), 101–136.
- Jacob, B. A., Lefgren, L., & Sims, D. (2010). The persistence of teacher-induced learning. *Journal of Human Resources*, 45(Fall (4)), 915–943.
- Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: an experimental evaluation*. NBER w14607. Cambridge, MA: National Bureau of Economic Research (December).
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27(December (6)), 615–631.
- Kershaw, J. A., & McKean, R. N. (1962). *Teacher shortages and salary schedules*. NY: McGraw-Hill.
- Koedel, C., & Betts, J. R. (2009). *Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique*. Department of Economics WP 09-02. Columbia, MO: University of Missouri (July).
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114(May (2)), 497–532.
- Krueger, A. B. (2002). Understanding the magnitude and effect of class size on student achievement. In L. Mishel, & R. Rothstein (Eds.), *The class size debate* (pp. 7–35). Washington, DC: Economic Policy Institute.
- Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy*, 110(December (6)), 1286–1317.
- Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The Future of Children*, 17(Spring (1)), 87–109.
- Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 99(December (5)), 1979–2011.
- Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*, 10(3), 179–214.
- Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: methods and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Mishel, L., & Rothstein, R. (Eds.). (2002). *The class size debate*. Washington, DC: Economic Policy Institute.
- Moe, T. M. (2005). Teacher unions and school board elections. In W. G. Howell (Ed.), *Besieged, school boards and the future of education politics* (pp. 254–287). Washington DC: The Brookings Institution.
- Mourshed, M., Chijioko, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. McKinsey and Company.
- Mulligan, C. B. (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy*, 107(December (6 pt. 2)), S184–S224.
- Muralidharan, K., & Sundararaman, V. (2009). *Teacher performance pay: experimental evidence from India*. NBER Working Paper 15323. Cambridge, MA: National Bureau of Economic Research.
- Murnane, R. J., & Cohen, D. K. (1986). Merit pay and the evaluation problem: why most merit pay plans fail and a few survive. *Harvard Educational Review*, 56(February (1)), 1–17.
- Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19(Fall (4)), 547–568.
- National Association of State Boards of Education. 1997. "Teacher tenure". Policy Updates 5(February (3)): 1.
- National Commission on Teaching America's Future. (1996). *What matters most: teaching for America's future*. New York: NCTAF.
- National Council on Teacher Quality. (2009). *State teacher policy yearbook, 2009*. Washington: National Council on Teacher Quality.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(January (3)), 237–257.
- Organisation for Economic Co-operation Development. (2010). *Education at a glance 2010: OECD indicators*. Paris: Organisation for Economic Co-operation and Development.
- Podgursky, M. (2011). Teacher compensation and collective bargaining. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education*, vol. 3 (pp. 279–313). Amsterdam: North Holland.
- Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: a review. *Journal of Policy Analysis and Management*, 26(4), 909–949.
- Provasnik, S., & Dorfman, S. (2005). *Mobility in the teacher workforce: findings for the condition of education 2005*. NCES 2005-114. Washington: National Center for Education Statistics (June).
- Raymond, Margaret E., Fletcher, Stephen, & Luque, Javier A. (2001). *Teach for America: an evaluation of teacher differences and student outcomes*

- in Houston, Texas. Stanford, CA: Center for Research on Education Outcomes (CREDO) Stanford University.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(March (2)), 417–458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review*, 94(May (2)), 247–252.
- Rothstein, J. (2010). Teacher quality in educational production: tracking, decay, and student achievement. *Quarterly Journal of Economics*, 25(1).
- Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2006). Do teachers really leave for higher paying jobs in alternative occupations? *Advances in Economic Analysis & Policy*, 6(1).
- Springer, M. G. (Ed.). (2009). *Performance incentives: their growing impact on American K-12 education*. Washington: Brookings Institution Press.
- Springer, M. G., Ballou, D., Hamilton, L., Le, V.-N., Lockwood, J. R., McCaffrey, D. F., et al. (2010). *Teacher pay for performance: experimental evidence from the project on incentives in teaching*. Nashville, TN: National Center on Performance Incentives, Vanderbilt University.
- Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, 24(Summer (3)), 97–118.
- Toch, T. and Rothman, R. 2008. "Rush to judgment: teacher evaluation in public education". Education Sector Reports Washington: Education Sector (January).
- U.S. Department of Education. (2009). *Digest of education statistics, 2008*. Washington, DC: National Center for Education Statistics.
- U.S. Department of Education. (2010). *Digest of education statistics, 2009*. Washington, DC: National Center for Education Statistics.
- Walsh, K., & Jacobs, S. (2007). *Alternative certification isn't alternative*. Washington, DC: Thomas B. Fordham Institute. (September).
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: our national failure to acknowledge and act on differences in teacher effectiveness* (2nd ed.). New York, NY: The New Teachers Project.
- Woessmann, L. (2011). *Cross-country evidence on teacher performance pay*. *Economics of Education Review*, 30(June (3)).
- Xu, Z., Hannaway, J., & Taylor, C. (2009). *Making a difference?: the effects of Teach for America in high school*. CALDER Working Paper 17 (revised). Washington, DC: Urban Institute (March).