

TEACHER QUALITY

ERIC A. HANUSHEK

Stanford University,

National Bureau of Economic Research

and

University of Texas at Dallas

STEVEN G. RIVKIN

Amherst College,

National Bureau of Economic Research

and

University of Texas at Dallas

Contents

Abstract	1052
Keywords	1052
Introduction	1053
1. Aggregate salary trends	1053
2. Distribution of teachers	1056
3. Teacher characteristics and student achievement	1058
3.1. Basic structure	1058
3.2. Evidence on measurable characteristics	1060
3.2.1. Teacher experience and education	1060
3.2.2. Teacher salary	1062
3.2.3. Teacher tests	1064
3.2.4. Teacher certification	1064
4. Outcome-based measures of quality	1065
5. Markets for teacher quality	1069
6. Policy connections	1071
7. Research agenda	1073
8. Conclusions	1074
Acknowledgement	1074
References	1075

Abstract

Improving the quality of instruction is a central component to virtually all proposals to raise school quality. Unfortunately, policy recommendations often ignore existing evidence about teacher labor markets and the determinants of teacher effectiveness in the classroom. This chapter reviews research on teacher labor markets, the importance of teacher quality in the determination of student achievement, and the extent to which specific observable characteristics often related to hiring decisions and salary explain the variation in the quality of instruction. The evidence is applied to the comparison between policies that seek to raise quality by tightening the qualifications needed to enter teaching and policies that seek to raise quality by simultaneously loosening entry restrictions and introducing performance incentives for teachers and administrators.

Keywords

teacher salaries, incentives, teacher experience, teacher education, teacher test scores

JEL classification: H4, I2, J4

Introduction

Teachers are central to any consideration of schools, and a majority of education policy discussions focus directly or indirectly on the role of teachers. There is a *prima facie* case for the concentration on teachers, because they are the largest single budgetary element in schools. Moreover, parents, teachers, and administrators emphasize repeatedly the fundamental role that teachers play in the determination of school quality. Yet there remains little consensus among researchers on the characteristics of a good teacher, let alone on the importance of teachers in comparison to other determinants of academic performance.

This chapter considers research related to the quality of teachers. Like many other areas where quality is important but difficult to observe, much of the evidence is indirect. Consideration of quality variation in the education sector is complicated further by the dominance of public provision of education, constraints on market operations, and the importance of nonpecuniary factors in the teacher supply decision. With public provision, schools are not necessarily operating in an efficient manner and do not necessarily make hiring decisions based on expected performance.¹

The relevant research follows three distinct lines that relate in varying ways to teacher quality. At the most aggregate level and possibly the most influential, a variety of studies have traced changes over time in the salaries of teachers relative to those in other occupations. This set of studies flows naturally into analyses of the importance of pay and nonpecuniary factors in determining the distribution of teachers among schools. A second line of research, following directly from the first, investigates the extent to which specific teacher characteristics account for differences in student achievement. Finally, the third line of research drops the parametric, input-based view of teacher quality and attempts to identify the total impact of teachers on student learning without the constraints imposed by relying on measurable characteristics.

Most of the evidence examines US schools, where data and analysis have been generally more plentiful. Relevant research on other countries is included and, where available, does not indicate qualitative differences in conclusions.

1. Aggregate salary trends

A starting point in the consideration of teacher quality is the evolution of teacher salaries over time in comparison to other workers.² Teacher salaries constitute equilibria in the teacher labor market, and both demand and supply side factors contribute to changes in relative teacher salaries. Importantly, even if the correlation between alternative employment opportunities and instructional quality is weak and school districts do not

¹ The issue of efficiency of public schools is the subject of Hanushek, this volume.

² More details on the time pattern of salaries in both the United States and the United Kingdom can be found in Dolton, this volume.

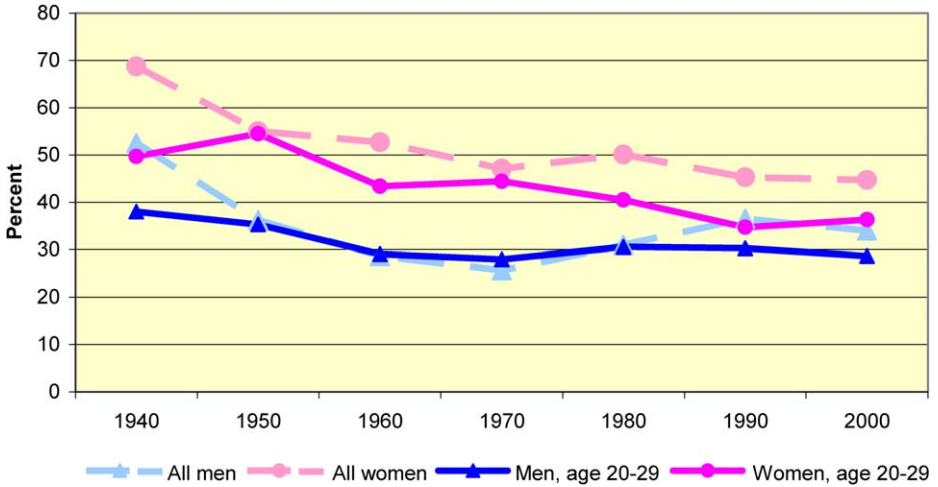


Figure 1. Percent college educated earning less than average teacher, by gender and age, 1940–2000.

systematically hire the best available teachers, any shift in supply would tend to move average quality in the same direction.

Figure 1 traces shows the proportion of 20–29 year old US college graduate non-teachers who earn less than the average 20–29 year old teacher by gender for the decennial censuses from 1940 to 2000.³ Over this period the earnings of young female and male teachers both declined relative to those for other occupations. However, there are substantial gender differences in the time path of relative salaries. For males, relative salaries fell between 1940 and 1960 but remained roughly constant afterward. For females by comparison, relative salaries started out high – above the median for college educated females – but fell throughout the period. The changes are easiest to see for young teachers and college graduates, where the adjustment has been larger, but they also hold for teachers of all ages [see Hanushek and Rivkin (1997)]. In other words, growth in late career salaries has not offset the decline in salaries for younger teachers.

Discussions of the education industry cost structure and women’s employment and earnings point to specific factors that have contributed to the decline in relative earnings of teachers and quite likely the quality of instruction as well. Perhaps most important

³ Note that salaries for teachers include all earnings, regardless of source. Thus, any summer or school year earnings outside of teaching are included. No adjustments are made, however, for any differences in the length of the school day or in the days worked during the year. Nor is any calculation of employer paid fringe benefits made. A clear discussion of the importance of each of these along with interpretation of the overall salary differences can be found in Podgursky (2003). For the time series comparisons, these omitted elements of compensation are most relevant if there have been relative changes in the importance of them between teachers and nonteachers over time. We currently have little data on any such changes.

is the cost pressure placed on schools and other slow growth industries by productivity improvements elsewhere in the economy [see Baumol and Bowen (1965), Baumol (1967)]. In contrast to other industries, education has experienced little technological change, driving up the price of teachers in real terms [Lakdawalla (2001, 2002)]. Notice that real wages tend to rise even if districts do not absorb fully the increased price of skilled labor, in which case the relative quality of new teachers is likely to decline over time.

Because almost all teachers are college graduates and most elementary and secondary school teachers are women, any factors that affect the earnings of highly-skilled workers or women invariably affect the price of teacher quality. Many highlight the adverse impact of the recent expansion in job opportunities for women on the supply of teachers [Flyer and Rosen (1997), Corcoran, Evans and Schwab (2004b), Bacolod (2003), Hoxby and Leigh (2004)]. The aforementioned effects of productivity growth elsewhere in the economy and expansion of international trade in ways that favor skilled workers almost certainly amplify the adverse effects on the supply of teachers. On the other hand, the rapid rise in college enrollment and female employment almost certainly offset at least a portion of the negative effects on the supply of teachers. Nevertheless, as a whole these developments appear to have imposed severe cost pressures on schools, and schools appeared to have responded by raising salaries less than the full increase in the wage growth for college educated females.

The decline in the relative earnings of teachers has likely led to a fall in average teacher quality of incoming teachers over this period. But, as Ballou and Podgursky (1997) point out, the short term implications of a change in relative earnings are not clear cut, because salary affects both the supply of new teachers and retention of currently employed teachers.

The extent of any teacher quality decline remains unclear and depends in large part on the correlation between teaching skill and the skills rewarded in the nonteacher labor market. In a simple unidimensional skill framework in which nonpecuniary factors play no role, the substantial decline in relative salary would be expected to lead to a large fall in teacher quality. However, a more complex and realistic framework in which the skill set of teachers differs from that of other professionals suggests the possibility of a more muted response to the salary changes. For example, if teaching places greater emphasis on a set of communication and interpersonal relation skills than the general labor market, the salaries relative to all college graduates may not provide a particularly good index of teacher quality. These concerns about the congruence of skills in different sectors point to a priority area for further research. The discussion in the following sections offers some insights into possible separation of the various markets, but that evidence also remains indirect.

Another important determinant of the elasticity of teacher quality with respect to salary is the responsiveness of current and prospective teachers to salary changes. There is reason to believe that teachers may be less responsive than other professionals. Specifically, the “family friendly” nature of teacher employment (with, for example, hours and vacations coinciding with those of kids) or intrinsic rewards from teaching may have

limited substitutes, making the decisions to enter or remain in teaching less sensitive to salary [see, for example, Scafidi, Sjoquist and Stinebrickner (2002)].

2. Distribution of teachers

One approach for disentangling the implications of the aggregate salary movements on quality has been to identify impacts on the distribution of observable teacher characteristics as proxies for quality. Investigations of salary effects on teacher characteristics take many forms and include both intertemporal evidence and cross-sectional evidence derived from different schooling systems and teacher labor markets.

A substantial body of research examines the effects of salary and nonpecuniary factors on the flows into and out of teaching and implicitly the supply of teachers with particular characteristics. This research, extended in a variety of dimensions, typically appears in two forms. The first analyzes the relationship between a specific teacher characteristic (TC) on the one hand and pay (P), benefits (B), or proxies for working conditions (WC) on the other. Examples include the determinants of the share of teachers with full certification, particular levels of experience, education, or teacher test scores,

$$TC = f(P, B, WC). \quad (1)$$

A second set of studies examines the determinants of teacher transitions, where transition probabilities are a function of the pecuniary and nonpecuniary factors described in equation (1), proxies for quality, and importantly the interactions of these two. Studies of shortages also fall into this category. Four types of teacher characteristics have received considerable attention: (1) experience; (2) measured achievement or skill; (3) specialty or subject area; and (4) credentials and teacher certification.

As is the case in other occupations, transition probabilities are quite high early in the career, decline with experience, and then increase as teachers move closer to retirement [e.g., Hanushek, Kain and Rivkin (2004)]. Evidence indicates that nonpecuniary characteristics likely related to working conditions have much stronger effects than pay on teacher transitions.⁴ Moreover, it appears that opportunity costs in terms of foregone earnings in other occupations are much less important than the complementarity of family considerations and school working conditions [e.g., Scafidi, Sjoquist and Stinebrickner (2002), Podgursky, Monroe and Watson (2004)] in determining the probability of exiting teaching. This is consistent with the view that salary plays a larger role in the decision to become a teacher than the choice of schools or exit from teaching. Finally, studies of teacher exits find that salaries and outside opportunities have differing impacts on teachers depending on experience; see, for example, Murnane and Olsen

⁴ Greenberg and McCall (1974), Murnane (1981), Hanushek, Kain and Rivkin (2004), Lankford, Loeb and Wyckoff (2002), Boyd et al. (2002, 2005) provide evidence on determinants of teacher transitions.

(1989, 1990), Dolton and Van der Klaauw (1995, 1999), Brewer (1996), Stinebrickner (1999, 2001a, 2001b), Gritz and Theobald (1996), Murnane et al. (1991), Scafidi, Sjoquist and Stinebrickner (2002).⁵ It appears that district personnel policies also affect teacher flows [cf. Murnane (1981)]. Therefore this evidence captures the reduced form relationship between characteristics and transition probabilities, and inferences about supply responses rely upon specific assumptions about the demand side of the market.

Scores on licensing, college entrance, and other examinations provide objective skill measures, and a number of studies investigate the relationship between scores on a particular test on the one hand and salaries and other school or labor market characteristics on the other [Murnane et al. (1991), Hanushek and Pace (1995), Podgursky, Monroe and Watson (2004)]. The majority of this work considers entry into the teaching profession.

The change in the character of entering teachers over time has also been addressed [Bacolod (2003), Corcoran, Evans and Schwab (2004a, 2004b)]. The impact of salary changes and of changes in other occupational opportunities for women, discussed above, is clearly seen from data splicing together performance on standardized tests over time. Bacolod (2003) combines information from the various National Longitudinal Surveys (Young Men, Young Women and Youth). Corcoran, Evans and Schwab (2004a, 2004b) extend the samples of teachers to other data sets, thus expanding the periods that can be investigated, and also concentrate on individuals who actually enter teaching.

Bacolod (2003) shows that the standardized test scores of people entering teaching as opposed to other professions have fallen over time – dramatically in the case of females. Specifically, recent birth cohorts who score near the top of IQ or AFQT tests are much less likely to want to be teachers than those in earlier birth cohorts.⁶ This drop is especially dramatic for women, but also holds for men and is consistent with the aggregate salary trends. Corcoran, Evans and Schwab (2004b) find that the relative fall in mean performance of female teachers, while significant, is much less than the fall at the top of the distribution.

The consideration of preparation has focused on the varying opportunity costs of teachers with different specialties. One of the first such studies considered how the uniform pay structure in teaching leads to shortages in specific areas, such as mathematics and science teachers who have better outside earnings opportunities [Kershaw and McKean (1962)]. That study highlighted the differential effects of policies and institutions on teachers with different characteristics. Following on Kershaw and McKean (1962), Rumberger (1987) examines how salaries affect the supply of science and math teachers.

Finally, considerable attention (although limited analysis) has been devoted to the possibility that school characteristics affect the ability of schools to hire fully credentialed teachers. In general this analysis simply reports gross correlations of lower

⁵ Note that these conclusions are frequently implicit from an analysis of hazard functions for exiting teaching.

⁶ This evidence splices together information from different surveys. By relying on relative performance measures, however, differences in tests are minimized.

proportions of uncertified teachers in central city and lower SES schools. Nonetheless, these casual observations almost surely do describe the reality – even if they do not fully identify the underlying impacts of individual, district, and state policy choices on the outcomes.

These studies provide information on the determinants of teacher transitions and the distributions teachers along a number of dimensions. The importance of the findings depends crucially on the relevance of the identified characteristics for determining student performance and other outcomes, i.e. the relationship with actual effectiveness in the classroom. This issue is the subject of the next section.

3. Teacher characteristics and student achievement

One general approach to understanding more about the extent to which specific teacher characteristics capture differences in instructional effectiveness is the estimation of the effects of specific characteristics on achievement and other student outcomes. We begin by describing the basic framework within which much of this research sits and then discuss the findings.

3.1. Basic structure

A large number of investigations of teacher quality focus on the effects of specific teacher characteristics on outcomes, controlling for student differences. These studies take a variety of forms. Here we provide an overview of the range of approaches that have been used. We critique the underlying modeling and interpretation in the subsequent sections.

A basic framework for the study of teacher effects begins with a model of achievement such as

$$O_g = f(F^{(g)}, P^{(g)}, C^{(g)}, T^{(g)}, S^{(g)}, \alpha), \quad (2)$$

where O_g is the outcome for a student in grade g ; F , P , C , T and S represent vectors of family, peer, community, teacher and school inputs, respectively; α is ability; and the superscript g indicates all of the inputs are cumulative from birth through grade g . Simply put, student achievement at any point in time represents the cumulative outcome of a wide variety of inputs.

This model, which is frequently referred to as an educational production function, has been applied often. Its history is generally traced back to the “Coleman Report” [Coleman et al. (1966)], an early study conducted under the auspices of the United States government. Since 1966, over 400 such studies have been published in journals and books. Empirical research pursuing this type of analysis typically collects data on the relevant inputs into performance from either administrative records or surveys.

The numerous current and past factors that affect achievement at any point in time seriously complicate efforts to estimate the effects of specific characteristics. Perhaps

most important is the extent to which any observed association between a school or teacher variable and student outcomes capture a causal relationship. For example, if children in higher income families attend schools with smaller classes on average than children in lower income families, the finding that smaller classes raise achievement may be driven in part by a failure to account fully for the direct effect of family income on student performance.

Teacher choice of schools can also complicate the estimation of teacher effects. As noted above, experienced teachers frequently have an option to move across districts and to choose the school within the district in which they are teaching, and they tend to take advantage of this [Greenberg and McCall (1974), Murnane (1981)]. Hanushek, Kain and Rivkin (2004) further show that teachers switching schools or districts tend to move systematically to places where student achievement is higher. This movement suggests the possibility of a simultaneous equations bias – that higher student achievement causes more experienced teachers or at least that causation runs both ways.

Another potential source of omitted variables bias is variation in state policy that might be correlated with the teacher characteristics. States, for example, determine the requirements to be a certified teacher, set the rules of collective bargaining on teacher contracts, and determine the financial structure including providing varying amounts of support for local schools depending upon their circumstances and tax base. States also specify the specific curriculum and outcome standards, establish testing requirements, and regulate a wide range of matters of educational process including various class size requirements, the rules for placement into special education classes, and disciplinary procedures. Because these policies vary widely across states, their omission could lead to bias coefficients in analyses that use data on a number of states. On the other hand, this concern is not relevant for cross-sectional analyses conducted within a single state where the policy environment is constant.⁷

More generally, value added models use prior achievement to mitigate problems of omitted variables bias. These models can take several forms depending upon assumptions regarding the depreciation of knowledge over time, and the most flexible form includes prior achievement in grade g^* as an additional explanatory variable:

$$O_g = f'(O_{g^*}, F^{(g)}, P^{(g)}, C^{(g)}, T^{(g)}, S^{(g)}, \alpha). \quad (3)$$

The precise estimation approach, and the resulting interpretation of any results, depends fundamentally on a series of assumptions about the structure of achievement and the underlying data generation process [see Hanushek (1979), Rivkin, Hanushek and Kain (2005), Todd and Wolpin (2003), Rivkin (2005)].

Though the use of such value added models mitigates problems resulting from the lack of historical information, it does not protect against the confounding influences of

⁷ In some other estimation, say, related to overall spending or class sizes, aggregation of data becomes an additional issue, but this is relatively unimportant for the teacher characteristics considered here, because those analyses have uniformly been conducted at lower levels of aggregation (the school district down to the classroom). See Hanushek, Rivkin and Taylor (1996).

contemporaneous factors related to the variables of interest and not captured by prior achievement. Given the limitations of most data, available variables may not account for all relevant variables. This has led to the use of panel data methods, instrumental variables, and other approaches described below.

A remaining limitation of virtually all education production function studies is the use of a small number of observed characteristics to capture school and teacher quality. Although this parametric approach lends itself to standard regression techniques, it provides limited information on the variation in teacher quality, in part because most studies use administrative or survey data that typically contain a very limited set of characteristics. The most commonly available characteristics, teacher education and experience, are clearly important variables to consider, because they almost always enter into the determination of teacher pay. Yet, as described below, they explain little of the actual variation in teacher effectiveness, and even more detailed information about college quality, scores on standardized examinations or other information continues to leave much unexplained. Moreover, whenever separate surveys are designed to provide a richer set of characteristics, the specific items are seldom replicated in other surveys, thus providing little ability to ascertain the generalizability of any findings.

3.2. Evidence on measurable characteristics

Investigations of measurable teacher characteristics invariably begin with education and experience. In the United States and many other countries, these account for much of the salary variation within school districts. Because of their administrative use, these variables are frequently available for researchers. A smaller number of studies use other characteristics including teacher test scores, college quality, salary and teacher certification.

The empirical analyses take many forms. A vast majority investigate variable effects on student achievement as measured by some form of standardized test, while the others estimate effects on school attainment, future earnings and other outcomes. The studies cover a range of grade levels, types of schools, areas of the United States and other countries, and they produce a divergent set of results on the key variables of interest.

3.2.1. Teacher experience and education

As noted, the most frequently studied aspects of teachers include their education and experience levels, the items that generally enter into pay determination. The simplest summary of their impact on student achievement from available analyses comes from aggregating the results across studies. Table 1, taken from Hanushek (1997, 2003), describes the estimated parameters from studies through 1994 in the United States.⁸

⁸ While more studies have appeared since then, they are small in numbers relative to the stock in 1994, and they show no discernibly different pattern of results from those in Table 1. For a description of the studies, a discussion of inclusion criteria, and the bibliography of included work, see Hanushek (1997).

Table 1
Percentage distribution of estimated effect of key teacher resources on student performance

Resources	Number of estimates	Statistically significant		Statistically insignificant
		Positive	Negative	
<i>All estimates</i>				
Teacher education	170	9%	5%	86%
Teacher experience	206	29	5	66
<i>High-quality estimates^a</i>				
Teacher education	34	0	9	91
Teacher experience	37	41	3	56

Source: Hanushek (1997, 2003).

^aHigh-quality estimates come from value-added estimation [equation (3)] where the sample is drawn for individual students from a single state.

Perhaps most remarkable is the finding that a master's degree has no systematic relationship to teacher quality as measured by student outcomes. This immediately raises a number of issues for policy, because advanced degrees invariably lead to higher teacher salaries and because advanced degrees are required for full certification in a number of states. Indeed, over half of current teachers in the US have a master's degree or more.

Teacher experience has a more positive relationship with student achievement, but still the overall picture is not that strong. While a majority of the studies finds a positive effect, only a minority of all estimates provides statistically significant results. Even the subset of studies that use a value added approach and information from a single state produce a highly variable set of results (see bottom panel in Table 1). If anything, the 37 value-added estimates within individual states suggest more strongly that experience has an impact, although still only 41% of the estimates are statistically significant. It is quite likely that a number of these studies lack the statistical power necessary to identify precisely the experience effects.

An important consideration in the case of experience is the possibility of a highly nonlinear relationship between the quality of instruction and experience. Murnane and Phillips (1981b) investigates the impact of experience with spline functions and find nonlinearities, although the actual estimates differ sharply across data samples. Rivkin, Hanushek and Kain (2005) also pursue a nonparametric investigation of experience and find that experience effects are concentrated in the first few years of teaching. Specifically, teachers in their first and, to a somewhat lesser extent, their second year tend to perform significantly worse in the classroom. Using a different estimation methodology, Hanushek et al. (2005) pinpoint the experience gains as arising during the first year of teaching, with essentially flat impacts of experience subsequently. Consequently, misspecification of the relationship between outcomes and experience likely contributed to the failure to find a systematic link between quality and experience.

Because of the high turnover rate early in the career, estimated returns to experience typically combine the acquisition of skills on the job with any nonrandom transitions out of teaching. Rivkin, Hanushek and Kain (2005) estimate experience coefficients identified by variation both across and within teachers with coefficients identified solely by within teacher changes in experience. The estimated experience effects are quite similar, indicating that the dominant effect is learning by doing in the first year in the classroom.

Similar investigations of teacher education and experience have been conducted in a wide range of developed and developing countries [Hanushek (2003)]. As a broad statement, the results are qualitatively similar except there is perhaps slightly stronger support for a positive impact of these in developing countries. At the same time, the additional support is slight with the majority of studies still not finding significant impacts. Moreover, these studies seldom provide truly adequate controls for the omitted variables problems discussed here.

3.2.2. *Teacher salary*

Instead of concentrating on the prior characteristics of teachers that enter into salary decisions, it is of course possible to analyze whether or not salary directly relates to student performance. Unfortunately such studies are frequently muddled. The majority of analyses relate the salary levels of teachers to the achievement of student. Yet, the salary level for any individual teacher is a composite of pay for specific characteristics (experience, education and other attributes as identified above) and, whenever the analysis crosses school districts, differences in the salary schedule. In other words, it has elements of movements along the salary schedule and shifts in the entire schedule.

The econometric evidence, presented in Table 2, again shows no strong evidence that salaries are a good measure of teacher quality. Overall, the studies show that salaries are more likely to be positively related to student achievement than negatively. Nonetheless, only a minority is statistically significant.

Many of the studies of teacher salaries are subject to the prior mentioned quality problems – lack of historical information and missing measures of state policy. The state policy concerns are especially important because states intervene in wage determination in a variety of ways that also are likely to influence school outcomes. The bottom portion of the table provides information on the more refined set of value-added, single state estimates. For this very small set of estimates, most are statistically insignificant. The estimates that are significant all come from a set of studies considering just single districts, so they provide estimates just about moves along the schedule and not what might happen with shifts in the entire schedule.

A series of other issues complicate efforts to identify the link between salaries outcomes. Perhaps most important is the possibility that nominal salaries in part reflect compensating differentials – for cost-of-living differences, for the desirability of partic-

Table 2
Percentage distribution of estimated effect of teacher salaries on student performance

Resources	Number of estimates	Statistically significant		Statistically insignificant
		Positive	Negative	
<i>All estimates</i>				
Teacher salary	118	20%	7%	73%
Teacher test scores	41	37	10	53
<i>High-quality estimates^a</i>				
Teacher salary	17	18	0	82
Teacher test scores	9	22	11	67

Source: Hanushek (1997, 2003).

^aHigh quality estimates come from value-added estimation [equation (3)] where the sample is drawn for individual students from a single state.

ular schools and their working conditions, or for such other things as urban crime.⁹ Most of the studies considering compensating differentials do not directly relate job-related characteristics and salaries to student outcomes but simply show that salaries vary with such characteristics. [An exception is Loeb and Page (2000) who argue on the basis of state panel data that compensating differentials have masked the effects of salaries in many prior studies of educational outcomes.¹⁰]

A second vexing issue is the importance of both past and current salaries in the distribution of the current stock of teachers. Salary influences entry into the profession, choice of first job, and movements among jobs, but tenure, lack of transferability of experience credit, and other factors almost certainly reduce the sensitivity of teacher transitions to salary as experience rises. Because virtually all analyses of salary effects compare current salaries with the effectiveness of the existing stock of teachers, this stock/flow amalgamation raises questions about the findings. An exception is Hanushek et al. (2005) who use a sample of district switchers to identify the relationship between salary and the quality of instruction. They do not find that higher salaries attract significantly more effective teachers, though the very small number of district switchers leads to imprecise estimates.

⁹ See, for example, Antos and Rosen (1975), Levinson (1988), Eberts and Stone (1985), Kenny (1980), Toder (1972), Hanushek and Luque (2000), Chambers and Fowler (1995), Fowler and Monk (2001) and Hanushek, Kain and Rivkin (2004).

¹⁰ Their study, relying on interstate variations in school completion and teacher pay, faces an analytical tradeoff between using aggregate state data subject to potential missing policy information and providing some control for state amenity differences.

3.2.3. *Teacher tests*

One measured characteristic – teacher scores on achievement tests – has received considerable attention, because it has more frequently been significantly correlated with student outcomes than the other characteristics previously discussed. [Table 2](#) displays the results of these studies. Several points are important. First, while the evidence is stronger than that for other explicit teacher characteristics, it is far from overwhelming. Second, the tests employed in these various analyses differ in focus and content, so the evidence mixes together a variety of things. At the very least, it is difficult to transfer this evidence to any policy discussions that call for testing teachers – because that would require a specific kind of test that may or may not relate to the evidence. Third, even when significant, teacher tests capture just a small portion of the overall variation in teacher effectiveness (see below).

The open research questions on both changes over time in the quality of instruction and the distribution among districts relate directly to the nature of tested knowledge and how it influences achievement. For example, [Wayne and Youngs \(2003\)](#) suggest that achievement does not uniformly matter but may relate to specific subjects (e.g., more important in secondary school mathematics instruction than in primary school reading). Additionally, as the investigations of time patterns cited suggest, the changes in teacher scores have not been uniform but instead have related more to the thickness of the upper tails of the distribution than to the mean. The existing research gives no hints of whether there is any nonlinear impact of knowledge in different ranges.

3.2.4. *Teacher certification*

The most pervasive policy action of states aimed at teacher quality is setting certification requirements. Although there is substantial variation across states in what is required for certification, the underlying theme is to set minimum requirements in an effort to ensure that no students are subjected to bad teaching. The problem is that, though certification requirements may prevent some poorly prepared teachers from entering the profession, they may also exclude others who would be quite effective in the classroom. Not only may some potentially good teachers be unable to pass the examinations, the certification requirements may discourage others from even attempting to enter the teaching profession; see, for example, [Murnane et al. \(1991\)](#). The nature of this tradeoff depends in large part on the objectives and skills of administrators who make teacher personnel decisions.¹¹

The literature provides mixed evidence on the effects of certification on teacher quality. Extensive literature has been accumulating on the importance of teacher certification and credentials, although it has proved quite controversial. Much of the work is based on specifications that are susceptible to substantial biases from other determinants of

¹¹ We thank Dale Ballou for providing a clear description of this tradeoff.

achievement, though a few recent papers provide more persuasive empirical specifications. Wayne and Youngs (2003) document the limitations of most studies on certification while reviewing some of the components of certification. Elements of the debate over the effectiveness of teacher certification can be traced through National Commission on Teaching and America's Future (1996), Abell Foundation (2001), Walsh (2002), Goldhaber and Brewer (2000, 2001), Darling-Hammond, Berry and Thoreson (2001). Goldhaber and Brewer (2000) find, for example, that teachers with subject-matter certification in mathematics perform better than other teachers, while teachers with emergency certification perform no worse than teachers with standard certification, although Darling-Hammond, Berry and Thoreson (2001) dispute the interpretation. Jepsen and Rivkin (2002) find small certification effects on teacher value added to mathematics and reading achievement once the nonlinearities in the return to experience are adequately controlled.

Two elements of this line of research merit particular attention. First, most states require teachers to meet certification requirements either upon hiring or within a short period of time. The studies that investigate teacher certification rely upon observations of existing school systems, where the lack of a teaching certificate generally implies a special situation. For example, urban school systems with heavily disadvantaged populations frequently find it hard to attract sufficient numbers of fully certified teachers and thus resort to hiring noncertified teachers. A very different situation is the development of specialized recruitment programs that are designed to bring people into the teaching profession for short periods of time. For example, the Teach for America program actively recruits top graduates of some of the best undergraduate schools to teach in difficult urban schools for a two year period [Raymond, Fletcher and Luque (2001), Raymond and Fletcher (2002), Decker, Mayer and Glazerman (2004)]. In these cases, not having a teacher certificate is intertwined with having attended a high-quality college or university. The nature of these hires is seldom explicitly described, but it clearly complicates the interpretation of the estimated effects. None of the studies of certification is clear about the nature of the selection process and, thus, about the generalizations that can be drawn from the findings.

Second, teacher certification varies dramatically across states. Simply identifying whether or not a teacher is certified will mean very different things depending on the state. Moreover, a variety of states have gone into alternative entry systems, and many will award a teacher certificate based on different criteria from those entering through traditional training institutions. Thus, even within a state, a teaching certificate may not indicate the completion of a given set of requirements.

4. Outcome-based measures of quality

An alternative approach to the examination of teacher quality concentrates on pure outcome-based measures of teacher effectiveness. The general idea is to investigate "total teacher effects" by looking at differences in growth rates of student achievement

across teachers. A good teacher would be one who consistently obtained high learning growth from students, while a poor teacher would be one who consistently produced low learning growth. In its simplest form, we could think of separating teacher effects from other inputs as in equation (4):

$$O_g - O_{g^*} = f''(F^{(g-g^*)}, P^{(g-g^*)}, C^{(g-g^*)}, T^{(g-g^*)}, S^{(g-g^*)}, \alpha) + t_j, \quad (4)$$

where t_j is the influence of having teacher j [conditional upon the other inputs, $f''(\cdot)$].

Equation (4) obviously places some structure on the achievement process, but the approach is appealing for several reasons. First, it does not require the choice of specific teacher characteristics, a choice that data limitations often constrain. Second, and related, it does not require knowledge of how different characteristics might interact in producing achievement. (Most prior work on specific characteristics assumes that the different observed characteristics enter linearly and additively in determining classroom effectiveness.) Third, it gives a benchmark for the importance of variations in teacher quality against which any consideration of specific skills or types of policy interventions can be compared.

A variety of studies have pursued this general approach over the past four decades; see Hanushek (1971, 1992), Armor et al. (1976), Murnane (1975), Murnane and Phillips (1981a), Aaronson, Barrow and Sander (2003), Rockoff (2004), Rivkin, Hanushek and Kain (2005) and Hanushek et al. (2005). Careful consideration of such work reveals the difficulties that must be overcome in order to estimate the variation of overall teacher effects.¹²

The major threats to the semiparametric estimation of the variance of teacher quality result from the nonrandom sorting of families among schools, the nonrandom sorting of students among classrooms, and test measurement issues.¹³ In addition to problems introduced by random measurement error, most achievement tests are not designed to provide valid rankings of the effectiveness of teachers with very different mixes of students in terms of academic preparation. For example, a test that concentrates on rudimentary material will do a poor job identifying differences in teacher quality among teachers whose students could answer the vast majority of questions on the basis of knowledge acquired prior to the current school year. Moreover, the average achievement gain could be higher for a poor teacher with initially low achieving students than for an excellent teacher with initially high achieving students if the test does not cover most of the material taught by the high-quality teacher.

Much of the early work was based on a single cross-section of teachers. In this framework, the observed student characteristics must control for all student heterogeneity. Moreover, the between teacher variance in achievement will conflate actual differences

¹² A similar study for developing countries (specifically Brazil) finds very consistent findings [Harbison and Hanushek (1992)].

¹³ The discussion of measurement error in school accountability measures is related. Kane and Staiger (2002a, 2002b) point out that aggregate school measurement error will introduce variability in apparent school performance over time.

and measurement error requiring the estimation of the teacher fixed effect error variance.¹⁴ In other words, the estimated teacher effect (\hat{t}) equals the true teacher effect (t) plus error.

The availability of multiple years of information for teachers permits the identification of the variance in teacher quality on the basis of the persistence of teacher fixed effects across years [Hanushek (1992), Hanushek et al. (2005)]. This eliminates the influences of random measurement error and year to year differences in student characteristics within classrooms. Specifically, if the measurement error in estimated teacher quality is uncorrelated across years, the expected value of the correlation of teacher by year fixed effects for years t_1 and t_2 is

$$E(r_{12}) = \frac{\text{var}(t)}{\text{var}(\hat{t})} \quad (5)$$

and the variance in true teacher quality (t) can be estimated directly (as long as it is constant across years).

Of course actual teacher effectiveness may change from year to year, and this approach classifies all nonpersistent outcomes as noise. This is particularly problematic in specifications that focus on within school and year variation. These estimated fixed effects are quite sensitive to teacher turnover, because turnover can dramatically change a teacher's place in the quality distribution in her school even when her effectiveness in the classroom is unchanged. Therefore, by focusing solely on the persistent quality differences [Equation (5)], some true systematic differences in teachers are masked by a varying comparison group and are treated as random noise. On the other hand, any persistent differences in classroom composition even within schools continue to bias the variance estimates.

Efforts to eliminate the confounding influences of student heterogeneity take a number of forms. Both Aaronson, Barrow and Sander (2003) and Hanushek et al. (2005) focus on within school variation in some specifications, eliminating both the actual variance in teacher quality between schools and any unobserved student, community, and school differences including the impacts of principals and other administrators. Controlling for differences in the quality of school administration is crucial given the important role attributed to principals and superintendents and the failure of observable characteristics to explain much of the variation in administrator quality.¹⁵ This approach mitigates most of the problem introduced by the nonrandom sorting of students among schools, and the inclusion of observed student and peer characteristics further reduces the effects of confounding factors. Hanushek et al. (2005) also transform the test score gain measure such that teachers are measured on the basis of the performance of their students relative to other students at a similar place in the initial test score distribution.

¹⁴ Aaronson, Barrow and Sander (2003) and Rockoff (2004) use different information from the teacher fixed effect regressions to construct estimates of the error component of the estimated between teacher variance.

¹⁵ See Broad Foundation and Thomas B. Fordham Institute (2003) for a discussion of administrator credentials.

Using a very different approach, [Rockoff \(2004\)](#) simultaneously estimates both student and teacher fixed effects on the level of achievement. This controls for all time invariant student differences in the level of achievement but does not account for systematic changes as students progress through school. In particular, knowledge acquired in a given year likely affects achievement in subsequent years, raising serious questions about the validity of this approach.

Regardless of the approach, the direct estimates of the teacher quality variance remain subject to biases resulting from unobserved student differences across classrooms. In order to control fully for student heterogeneity and avoid problems introduced by measurement error, [Rivkin, Hanushek and Kain \(2005\)](#) aggregate across teachers in a grade, remove student and school by grade fixed effects, and focus on the link between teacher turnover and variation in student achievement. This approach produces a lower bound estimate of the variation in teacher quality that almost certainly underestimates the true variance by a substantial amount. Not only does it ignore all between school variation in teacher quality, but violations of the maintained assumptions (about the stability of teacher effects and about the distribution of teacher quality) and measurement error both attenuate the estimated variance.

The magnitude of estimated differences in teacher quality is impressive. [Hanushek \(1992\)](#) shows that teachers near the top of the quality distribution can get an entire year's worth of additional learning out of their students compared to those near the bottom.¹⁶ That is, a good teacher will get a gain of 1.5 grade level equivalents while a bad teacher will get 0.5 year for a single academic year.

The more conservative lower bound estimators used by [[Rivkin, Hanushek and Kain \(2005\)](#)] also generate sizable estimates of the teacher quality variance: moving from an average teacher to one at the 85th percentile of teacher quality (i.e., moving up one standard deviation in teacher quality) increases student achievement gains by more than 4 percentile ranks in the given year. With their data, this is roughly equivalent to the effects of a ten student (approximately 50%) decrease in class size. As noted above, this method almost certainly understates the true variance in the quality of instruction. The within school estimators of the teacher quality variance reported in [Hanushek et al. \(2005\)](#) are roughly 50 percent larger. Importantly, the results for specifications that focus solely on within school differences do not differ markedly from those that also include teacher quality differences among schools, indicating that most of the variation in the quality of instruction occurs within schools.

The pattern of findings in the Project STAR study is also consistent with existence of substantial within school differences in teacher quality. Project STAR is the widely cited study of class size that involved random assignment of students to classes with varying numbers of students [[Word et al. \(1990\)](#)].¹⁷ Average differences by class size

¹⁶ These estimates consider value-added models with family and parental models. The sample includes only low-income minority students, whose average achievement in primary school is below the national average. The comparisons given compare teachers at the 5th percentile with those at the 95th percentile.

¹⁷ Students were assigned to three separate treatment groups: regular-sized classes (22–25 students), regular-sized classes with an aide (22–25 students) and small classes (12–17).

were the focus of the experiment, but the student results actually differed widely by specific classroom. In only 40 out of 79 schools did the kindergarten performance in the small classroom exceed that in the regular classrooms (with and without aides). This is significantly greater than random (26 out of 79), but much smaller than might be expected to result from simple random test error given the large difference in class size among classrooms. The most straightforward interpretation of this heterogeneity is that variations in teacher quality are very important relative to the effects of smaller classes.¹⁸

These estimates of teacher quality can also be related to the popular argument that family background is overwhelmingly important and that schools cannot be expected to make up for bad preparation from home. This perspective emanates from work that treats schools as monolithic institutions or equates quality with expenditure. The existence of substantial within school variation in teacher quality documented in Rivkin, Hanushek and Kain (2005) points to the fact that high quality teachers can offset a substantial portion of disadvantage related to family economic and social circumstances.

The discussion to this point treats teacher quality as common to all students in a classroom, but evidence suggests that teachers may be more effective with some students than with others. Specifically, both Dee (2004) using the random assignment data from the Tennessee STAR experiment and Hanushek et al. (2005) find strong evidence that teachers are more effective with students whose race matches their own. Similar variations across student ability dimensions do not, however, show such variations – suggesting that a good teacher is generally good for all students.

5. Markets for teacher quality

Output-based quality measures can also be used to trace patterns of teacher movements by classroom effectiveness rather than by proxies for quality as is the case in the work discussed in Section 2. Hanushek et al. (2005) utilize the matched panel data for teachers and students for a single large metropolitan district in Texas to describe the distribution of teacher quality by transition status.

Figure 2 plots the distributions of estimated teacher fixed effects by transition status based just upon within school variations in teacher performance [Hanushek et al. (2005)]. Neither these distributions nor comparisons of average quality across transition categories indicate that the average quality of teachers who leave inner city schools either for other districts or for employment outside of the Texas public schools exceeds the average quality of those who remain. This contrasts sharply with the popular belief that inner city districts disproportionately lose their better teachers to other school districts or other occupations. The inner city districts do have higher teacher turnover, but this evidence suggests that it is not concentrated among higher-quality teachers.

¹⁸ A discussion of the experiment and overall results can be found in Word et al. (1990). Hanushek (1999) analyzes the basic experimental results and identifies the variation across classrooms.

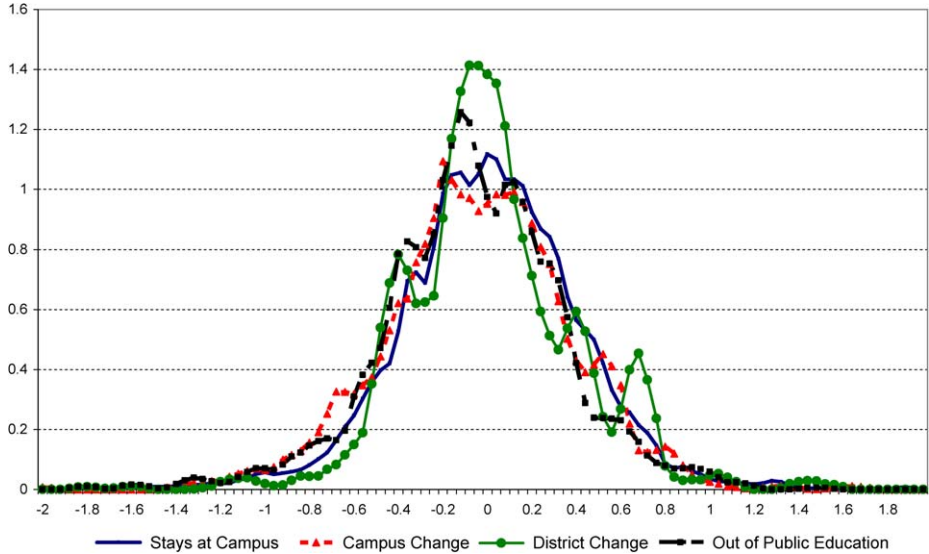


Figure 2. Kernel density estimates of teacher quality distribution: standardized average gains compared to other teachers at the same campus by teacher move status.

These data also permit an exploratory analysis of the market for quality and the competition across districts. Specifically, a number of the teachers in the large urban district decide to move to suburban districts. With the available data, the specifics of market interactions are not observed, only the results. It is not known where teachers applied for jobs or what districts were advertising for teachers. Nonetheless, if the simple assumption that higher salary and student demographic characteristics found to attract teachers deepens the applicant pool holds, the relationship between quality on the one hand and salary and school characteristics on the other provides information about district demand for quality. The preliminary results show little systematic evidence that districts prefer teachers who were more effective. Rather, the evidence suggests that higher salaries and lower minority enrollments enable districts to hire teachers with master's degrees, a characteristic with virtually no value in predicting quality. Importantly, this finding is consistent with both the inability to form an informative estimate of teacher effectiveness and with a lack of district focus on quality. Note, however, that the small sample size and use of estimates of teacher quality led to quite noisy estimates.

The possibility of obtaining outcome based quality measures from a wider range of local labor markets offers the prospect of understanding better the choices of teachers and of districts. It would, for example, be useful to investigate how the competitiveness of different areas in terms of alternative school districts affects the hiring patterns.

6. Policy connections

The research into teacher quality is scrutinized intensely because it has a direct relationship to current policy debates. Policy makers face conflicting suggestions about how to proceed. It is useful to relate the evidence on teacher quality to some of the central debates and to consider where the evidence is strong and weak.

Perhaps the key issue that pervades discussions is the tension between expanded state or even federal regulation of teacher labor markets versus decentralization of authority to schools and local education authorities. Another way to frame this discussion is as a debate over whether to tighten or to loosen licensing requirements for teachers.

The available evidence indicates clearly that legislating “good teachers” has been extraordinarily difficult. The idea behind most certification requirements is that they ensure that nobody gets a really terrible teacher. In other words, the general idea is that we can put a floor on quality. But doing this requires knowledge of characteristics that systematically affect performance. The prior evidence does not indicate that we can do this with any certainty.

Two caveats are, however, important. First, the existing research has not been very precise about the characteristics of certification requirements. The requirements in the US vary significantly by state, but the typical analysis has not investigated the components of certification in any detail. Second, and related, much of the attention to certification has centered on calls to expand current certification in significant ways. For example, certification for secondary school mathematics and science might require a college major in the subject (as opposed to a degree in mathematics education *per se*). Certification might also require advanced degrees in a combination of child psychology, pedagogy and the like. These details have not been adequately addressed in existing research.

Tightening up on requirements essentially makes it more costly to enter teaching, and thus one would expect it to the lower supply of teachers. This would imply that the cost of teachers of any given quality would rise. Nonetheless, virtually nothing is known about the magnitude or importance of such feedback effects on the teacher labor market.

The other side, loosening up, begins with the observation that existing evidence shows substantial variation in teacher quality, even among teachers with similar education and experience. This variation likely results from several factors: differences in teacher skill and effort; inadequate personnel practices (particularly the retention process but also the hiring process) in many schools and districts; and differences in the number and quality of teachers willing to work by subject and working conditions.

This policy position would allow more flexibility on who could enter the teaching profession but then would focus more on the overall incentive structure including retention, promotion, and pay decisions. The key ambiguities here center on the ability to identify teacher quality with sufficient precision to be useful in formulating policies and the ability to craft incentives that lead to higher quality.

Schools could utilize two basic methods for measuring teacher quality, one based on evaluations of overall effectiveness and the other based on statistical estimation of teacher value added. The former measure is clearly more comprehensive and nuanced, but it requires that administrators can both formulate a useful measure of teacher effectiveness and actually use that measure to rate teachers. There is evidence that principals can identify high-quality teachers in terms of value added to student learning. Early research on this by Murnane (1975) and Armor et al. (1976) showed that the normal evaluations of principals were highly correlated with the value added of teachers, even though the principal did not have the test and value added information available. This research has, nonetheless, not been replicated using different samples or different estimation approaches for finding the value added of teachers. In addition, the lack of success of merit pay programs suggests that it might be quite difficult for principals to actually apply these ratings in a high stakes environment.

In terms of the statistical evaluation approach, the State of Tennessee formalized the estimation of value added for teachers using annual state tests that linked pupil results with their teachers [Sanders and Horn (1994, 1995), Sanders, Saxton and Horn (1997)].¹⁹ This approach, while mandated for the state, has not been directly linked to incentives for teachers in Tennessee, though other states have linked student performance with teacher compensation. Concerns have arisen about flaws in the structure of some state accountability systems, but little or no evidence exists regarding the impact of these systems on the quality of classroom instruction.

Most policy evaluations also take the existing training of teachers as given without considering alternatives.²⁰ For example, loosening up on the certification requirements for entry and relying more on subsequent evaluation of performance should, in theory, lead principals and school decision makers to pay more attention to teacher performance. This new role could well imply that they pay more attention to the pre-service and in-service training that teachers receive and this in turn could put pressure on education schools to alter their programs.

There remains limited evidence on the effect of incentive systems more generally on the quality of instruction. Some evidence has accumulated about merit pay plans, and this does not indicate that merit pay as applied to schools has been very effective [Cohen and Murnane (1986)]. There is reason to believe that these experiments are, however, too limited in the magnitude and character of the incentive scheme [cf. Hanushek et al.

¹⁹ For a discussion of the specific approach along with an analysis of its sensitivity, see Ballou, Sanders and Wright (2004). The issues of error variance in teacher quality estimates are also relevant [Kane and Staiger (2002b)].

²⁰ There have been a variety of experiments in different states with alternative routes to teaching that do not involve traditional certification. The existing evidence on their success or failure is limited, but one program that has been carefully studied, the Teach for America program, shows generally positive results [see Raymond, Fletcher and Luque (2001), Decker, Mayer and Glazerman (2004)]. This program concentrates on getting graduates from very selective universities to commit to teaching for a limited amount of time and does not require the commitment to formal teacher training that normal certification requires.

(1994)]. Newer evidence from direct experiments provides stronger results with incentive pay comparing favorably to other school policies [Lavy (2002)]. For consideration of the available evidence on teacher merit pay, see Karnes and Black (1986), Cohen and Murnane (1985, 1986), Ballou and Podgursky (1993, 1997), Cohn (1996) and Brickley and Zimmerman (2001).²¹

Credible research into training versus selection issues as related to certification policies, merit pay, and so forth clearly requires longitudinal observations that link teachers, programs, and student performance. Until recently, there has been little possibility of such work, although recent developments of large, longitudinal databases from administrative records indicate that this may soon change.

7. Research agenda

The range of research needs and productive areas of inquiry can largely be seen by retracing the open questions of the previous sections. The most obvious complication to research arises from the fact that observed schooling situations represent the outcomes of several interrelated choices – those of parents, teachers, administrators and policy-makers. This complexity makes it difficult to separate the various influences reliably. Thus, for example, judging variations in teacher quality require distinguishing teacher effects from elements of students and parents themselves.

Attention to these issues of selection, omitted variables bias, and causation was not a central element of the early work on teacher quality but has come to the forefront in recent research. As has developed in related work in public economics and in labor economics, there are a variety of ways of potentially disentangling the effects of various programs and elements. While this is not the place to go through these approaches, it is clear that refinement of research in these directions is an important part of any research agenda.²²

Another area at the top of any agenda has to be developing a better understanding of how the market for teacher quality works. This research is clearly dependent upon developing reliable measures of teacher quality in a variety of different institutional circumstances. Suffice it to say that, even though a majority of discussion of teachers concentrates directly on teacher quality, most of the research about teacher markets lacks any direct investigation of teacher quality differences.

One area, however, warrants special attention. The discussion to this point has been virtually silent on the issue of cost. Policy decisions clearly require combining information about benefits with that about costs. Yet, almost nothing has been done to measure

²¹ One important issue in the evaluation of merit pay schemes is the expectations for where results should show up. With an evaluation over a short period of time, the results would indicate whether a merit pay scheme affects the amount of additional effort that is induced from teachers. Over a longer period of time, however, evaluation would point to the impact on selection into teaching and retention of teachers.

²² See also the related discussions in Hanushek, this volume, and Glewwe and Kremer, this volume.

the cost of teacher quality. For example, the costs of various training programs (pre-service and in-service) focused on improving teacher quality can be estimated, but they are never related to variations in teacher quality that are achieved. Similarly, discussions of salary policies tend not to be related to any measures of teacher quality. Attention to cost issues is a neglected area that sorely needs further work.

A second area of considerable neglect has been the interaction of teacher unions with teacher quality. Although it is widely believed that teacher unions create rigidities in hiring systems, little specific analysis identifies the magnitude or impact of these.²³ The discussion of retention and selection of teachers, for example, suggests that more focused policies might improve teacher quality, but these policies would appear to conflict with many union objectives and contract restrictions. Such an analysis, which necessarily gets into questions of political economy, is closely related to issues of policies related to incentives.

Along the same lines, much of the current policy discussion of accountability in schools and of choice in schools has a direct bearing on teacher quality concerns. Indeed, most people would see the potential effects of these policies as coming through their impacts on teacher quality. But, again, little is known about the potential interactions of these institutional structures and teacher quality.

Finally, recent advances in the economic analysis of contracting has obvious application to schools. The range of questions involving partial observability of performance, principal-agent problems, and the like are frequently motivated by suggestions of school reward structures [see, for example, Baker (1992, 2002), Lazear (1995)].

8. Conclusions

The growth in interest in questions of teacher quality is being met by an explosion of new data and analytical possibilities. This is married with increased interest in new strategies to separate true causal effects from associations due to selection and omitted variables. It seems reasonable then to presume that many of the open issues in the discussion here will soon be addressed if not resolved.

Acknowledgement

This research has been supported by a grant from the Packard Humanities Institute.

²³ One study finds that unions increase costs and lower productivity of schools, but it does not directly relate to issues of teacher quality [Hoxby (1996)].

References

- Aaronson, D., Barrow, L., Sander, W. (2003). "Teachers and student achievement in the Chicago public high schools". WP 2002-28. Federal Reserve Bank of Chicago (June).
- Abell Foundation (2001). *Teacher Certification Reconsidered: Stumbling for Quality*. Abell Foundation, Baltimore, MD.
- Antos, J.R., Rosen, S. (1975). "Discrimination in the market for teachers". *Journal of Econometrics* 2 (May), 123–150.
- Armor, D.J., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., Zellman, G. (1976). *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools*. RAND Corp., Santa Monica, CA.
- Bacolod, M.P. (2003). "Do alternative opportunities matter? The role of female labor markets in the decline of teacher supply and teacher quality, 1940–1990". Department of Economics, University of California, Irvine (September).
- Baker, G.P. (1992). "Incentive contracts and performance measurement". *Journal of Political Economy* 100 (3), 598–614.
- Baker, G.P. (2002). "Distortion and risk in optimal incentive contracts". *Journal of Human Resources* 37 (4), 728–751.
- Ballou, D., Podgursky, M. (1993). "Teachers' attitudes toward merit pay: Examining conventional wisdom". *Industrial and Labor Relations Review* 47 (1), 50–61.
- Ballou, D., Podgursky, M. (1997). *Teacher Pay and Teacher Quality*. W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.
- Ballou, D., Sanders, W., Wright, P. (2004). "Controlling for student background in value-added assessment of teachers". *Journal of Educational and Behavioral Statistics* 29 (1), 37–65.
- Baumol, W.J. (1967). "Macroeconomics of unbalanced growth: The anatomy of urban crisis". *American Economic Review* 57 (3), 415–426.
- Baumol, W.J., Bowen, W.G. (1965). "On the performing arts: The anatomy of their economic problems". *American Economic Review* 55 (May), 495–502.
- Boyd, D., Lankford, H., Loeb, S., Wyckoff, J. (2002). "Do high-stakes tests affect teachers' exit and transfer decisions? The case of the 4th grade test in New York State". Mimeo. Stanford Graduate School of Education.
- Boyd, D., Lankford, H., Loeb, S., Wyckoff, J. (2005). "The draw of home: How teachers' preferences for proximity disadvantage urban schools". *Journal of Policy Analysis and Management* 24 (1), 113–132.
- Brewer, D.J. (1996). "Career paths and quit decisions: Evidence from teaching". *Journal of Labor Economics* 14 (2), 313–339.
- Brickley, J.A., Zimmerman, J.L. (2001). "Changing incentives in a multitask environment: Evidence from a top-tier business school". *Journal of Corporate Finance* 7, 367–396.
- Broad Foundation and Thomas B. Fordham Institute (2003). *Better Leaders for America's Schools: A Manifesto*. Broad Foundation and Thomas B. Fordham Institute, Washington, DC.
- Chambers, J., Fowler Jr., W.J. (1995). *Public School Teacher Cost Differences Across the United States*. National Center for Education Statistics, Washington, DC.
- Cohen, D.K., Murnane, R.J. (1985). "The merits of merit pay". *Public Interest* 80 (Summer), 3–30.
- 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 (1), 1–17.
- Cohn, E. (1996). "Methods of teacher remuneration: Merit pay and career ladders". In: Becker, W.E., Baumol, W.J. (Eds.), *Assessing Educational Practices: The Contribution of Economics*. MIT Press, Cambridge, MA, pp. 209–238.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., York, R.L. (1966). *Equality of Educational Opportunity*. U.S. Government Printing Office, Washington, DC.
- Corcoran, S.P., Evans, W.N., Schwab, R.M. (2004a). "Changing labor-market opportunities for women and the quality of teachers, 1957–2000". *American Economic Review* 94 (2), 230–235.

- Corcoran, S.P., Evans, W.N., Schwab, R.M. (2004b). "Women, the labor market, and the declining relative quality of teachers". *Journal of Policy Analysis and Management* 23 (3), 449–470.
- Darling-Hammond, L., Berry, B., Thoreson, A. (2001). "Does teacher certification matter? Evaluating the evidence". *Educational Evaluation and Policy Analysis* 23 (1), 57–77.
- Decker, P.T., Mayer, D.P., Glazerman, S. (2004). "The effects of teach for America on students: Findings from a national evaluation". Discussion Paper 1285-04. Institute for Research on Poverty, University of Wisconsin, Madison (July).
- Dee, T.S. (2004). "Teachers, race, and student achievement in a randomized experiment". *Review of Economics and Statistics* 86 (1), 195–210.
- Dolton, P.J., Van der Klaauw, W. (1995). "Leaving teaching in the UK: A duration analysis". *The Economic Journal* 105 (March), 431–444.
- Dolton, P.J., Van der Klaauw, W. (1999). "The turnover of teachers: A competing risks explanation". *Review of Economics and Statistics* 81 (3), 543–552.
- Eberts, R.W., Stone, J.A. (1985). "Wages, fringe benefits, and working conditions: An analysis of compensating differentials". *Southern Economic Journal* 52 (1), 74–79.
- Flyer, F., Rosen, S. (1997). "The new economics of teachers and education". *Journal of Labor Economics* 15 (1), 104–139.
- Fowler Jr., W.J., Monk, D.H. (2001). *A Primer on Making Cost Adjustments in Education*. National Center for Education Statistics, Washington, DC.
- 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 (2), 129–145.
- Goldhaber, D.D., Brewer, D.J. (2001). "Evaluating the evidence on teacher certification: A rejoinder". *Educational Evaluation and Policy Analysis* 23 (1), 79–86.
- Greenberg, D., McCall, J. (1974). "Teacher mobility and allocation". *Journal of Human Resources* 9 (4), 480–502.
- Gritz, M.R., Theobald, N.D. (1996). "The effects of school district spending priorities on length of stay in teaching". *Journal of Human Resources* 31 (3), 477–512.
- Hanushek, E.A. (1971). "Teacher characteristics and gains in student achievement: Estimation using micro data". *American Economic Review* 60 (2), 280–288.
- Hanushek, E.A. (1979). "Conceptual and empirical issues in the estimation of educational production functions". *Journal of Human Resources* 14 (3), 351–388.
- Hanushek, E.A. (1992). "The trade-off between child quantity and quality". *Journal of Political Economy* 100 (1), 84–117.
- Hanushek, E.A. (1997). "Assessing the effects of school resources on student performance: An update". *Educational Evaluation and Policy Analysis* 19 (2), 141–164.
- Hanushek, E.A. (1999). "Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects". *Educational Evaluation and Policy Analysis* 21 (2), 143–163.
- Hanushek, E.A. (2003). "The failure of input-based schooling policies". *Economic Journal* 113 (485), F64–F98.
- Hanushek, E.A., Kain, J.F., O'Brien, D.M., Rivkin, S.G. (2005). "The market for teacher quality". Working Paper 11154. National Bureau of Economic Research, Cambridge, MA (February).
- Hanushek, E.A., Kain, J.F., Rivkin, S.G. (2004). "Why public schools lose teachers". *Journal of Human Resources* 39 (2), 326–354.
- Hanushek, E.A., Luque, J.A. (2000). "Smaller classes, lower salaries? The effects of class size on teacher labor markets". In: Laine, S.W.M., Ward, J.G. (Eds.), *Using what We Know: A Review of the Research on Implementing Class-Size Reduction Initiatives for State and Local Policymakers*. North Central Regional Educational Laboratory, Oak Brook, IL, pp. 35–51.
- Hanushek, E.A., Pace, R.R. (1995). "Who chooses to teach (and why)?" *Economics of Education Review* 14 (2), 101–117.
- Hanushek, E.A., Rivkin, S.G. (1997). "Understanding the twentieth-century growth in U.S. school spending". *Journal of Human Resources* 32 (1), 35–68.

- Hanushek, E.A., Rivkin, S.G., Taylor, L.L. (1996). "Aggregation and the estimated effects of school resources". *Review of Economics and Statistics* 78 (4), 611–627.
- Hanushek, E.A., et al. (1994). *Making Schools Work: Improving Performance and Controlling Costs*. Brookings Institution Press, Washington, DC.
- Harbison, R.W., Hanushek, E.A. (1992). *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. Oxford University Press, New York.
- Hoxby, C.M. (1996). "How teachers' unions affect education production". *Quarterly Journal of Economics* 111 (3), 671–718.
- Hoxby, C.M., Leigh, A. (2004). "Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States". *American Economic Review* 94 (2), 236–240.
- Jepsen, C., Rivkin, S.G. (2002). "What is the trade-off between smaller classes and teacher quality?". National Bureau of Economic Research.
- Kane, T.J., Staiger, D.O. (2002a). "The promise and pitfalls of using imprecise school accountability measures". *Journal of Economic Perspectives* 16 (4), 91–114.
- Kane, T.J., Staiger, D.O. (2002b). "Volatility in school test scores: Implications for test-based accountability systems". In: Ravitch, D. (Ed.), *Brookings Papers on Education Policy 2002*. Brookings Institution Press, Washington, DC, pp. 235–269.
- Karnes, E.L., Black, D.D. (1986). *Teacher Evaluation and Merit Pay: An Annotated Bibliography*. Greenwood Press, New York.
- Kenny, L.W. (1980). "Compensating differentials in teachers' salaries". *Journal of Urban Economics* 7 (March), 198–207.
- Kershaw, J.A., McKean, R.N. (1962). *Teacher Shortages and Salary Schedules*. McGraw-Hill, New York.
- Lakdawalla, D. (2001). "The declining quality of teachers". Working Paper 8263. National Bureau of Economic Research, Cambridge, MA (April).
- Lakdawalla, D. (2002). "Quantity over quality". *Education Next* 2 (3), 67–72.
- Lankford, H., Loeb, S., Wyckoff, J. (2002). "Teacher sorting and the plight of urban schools: A descriptive analysis". *Educational Evaluation and Policy Analysis* 24 (1), 37–62.
- Lavy, V. (2002). "Evaluating the effect of teachers' group performance incentives on pupil achievement". *Journal of Political Economy* 110 (6, December), 1286–1317.
- Lazear, E.P. (1995). *Personnel Economics*. MIT Press, Cambridge, MA.
- Levinson, A.M. (1988). "Reexamining teacher preferences and compensating wages". *Economics of Education Review* 7 (3), 357–364.
- Loeb, S., Page, M.E. (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 (3), 393–408.
- Murnane, R.J. (1975). *Impact of School Resources on the Learning of Inner City Children*. Ballinger, Cambridge, MA.
- Murnane, R.J. (1981). "Teacher mobility revisited". *Journal of Human Resources* 16 (1), 3–19.
- Murnane, R.J., Olsen, R. (1989). "The effects of salaries and opportunity costs on length of stay in teaching: Evidence from Michigan". *Review of Economics and Statistics* 71 (2), 347–352.
- Murnane, R.J., Olsen, R. (1990). "The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina". *Journal of Human Resources* 25 (1), 106–124.
- Murnane, R.J., Phillips, B. (1981a). "What do effective teachers of inner-city children have in common?". *Social Science Research* 10 (1), 83–100.
- Murnane, R.J., Phillips, B.R. (1981b). "Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance". *Economics of Education Review* 1 (4), 453–465.
- Murnane, R.J., Singer, J.D., Willett, J.B., Kemple, J.J., Olsen, R.J. (1991). *Who Will Teach? Policies that Matter*. Harvard University Press, Cambridge, MA.
- National Commission on Teaching and America's Future (1996). *What Matters Most: Teaching for America's Future*. NCTAF, New York.
- Podgursky, M. (2003). "Fringe benefits". *Education Next* 3 (3), 71–76.

- Podgursky, M., Monroe, R., Watson, D. (2004). "The academic quality of public school teachers: An analysis of entry and exit behavior". *Economics of Education Review* 23 (5), 507–518.
- Raymond, M.E., Fletcher, S. (2002). "Teach for America". *Education Next* 2 (1), 62–68.
- Raymond, M.E., Fletcher, S., Luque, J.A. (2001). *Teach for America: An Evaluation of Teacher Differences and Student Outcomes in Houston, Texas*. CREDO, Hoover Institution, Stanford, CA.
- Rivkin, S.G. (2005). "Cumulative nature of learning and specification bias in education research". Mimeo. Department of Economics, Amherst College, Amherst, MA.
- Rivkin, S.G., Hanushek, E.A., Kain, J.F. (2005). "Teachers, schools, and academic achievement". *Econometrica* 73 (2), 417–458.
- Rockoff, J.E. (2004). "The impact of individual teachers on student achievement: Evidence from panel data". *American Economic Review* 94 (2), 247–252.
- Rumberger, R.W. (1987). "The impact of salary differentials on teacher shortages and turnover: The case of mathematics and science teachers". *Economics of Education Review* 6 (4), 389–399.
- Sanders, W.L., Horn, S.P. (1994). "The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment". *Journal of Personnel Evaluation in Education* 8, 299–311.
- Sanders, W.L., Horn, S.P. (1995). "The Tennessee value-added assessment system (TVAA): Mixed model methodology in educational assessment". In: Shinkfield, A.J., Stufflebeam, D.L. (Eds.), *Teacher Evaluation: Guide to Effective Practice*. Kluwer Academic, Boston, pp. 337–376.
- Sanders, W.L., Saxton, A.M., Horn, S.P. (1997). "The Tennessee value-added assessment system: A quantitative, outcomes-based approach to educational assessment". In: *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* Corwin Press, Thousand Oaks, CA, pp. 137–162.
- Scafidi, B., Sjoquist, D., Stinebrickner, T.R. (2002). "Where do teachers go?". Mimeo. Georgia State University (October).
- Stinebrickner, T.R. (1999). "Estimation of a duration model in the presence of missing data". *Review of Economics and Statistics* 81 (3), 529–542.
- Stinebrickner, T.R. (2001a). "Compensation policies and teacher decisions". *International Economic Review* 42 (3), 751–779.
- Stinebrickner, T.R. (2001b). "A dynamic model of teacher labor supply". *Journal of Labor Economics* 19 (1), 196–230.
- Todd, P.E., Wolpin, K.I. (2003). "On the specification and estimation of the production function for cognitive achievement". *Economic Journal* 113 (485).
- Toder, E.J. (1972). "The supply of public school teachers to an urban metropolitan area: A possible source of discrimination in education". *Review of Economics and Statistics* 54 (4), 439–443.
- Walsh, K. (2002). "Positive spin: The evidence for traditional teacher certification, reexamined". *Education Next* 2 (1), 79–84.
- Wayne, A.J., Youngs, P. (2003). "Teacher characteristics and student achievement gains: A review". *Review of Educational Research* 73 (1), 89–122.
- Word, E., Johnston, J., Bain, H.P., Fulton, B.D., Zaharies, J.B., Lintz, M.N., Achilles, C.M., Folger, J., Breda, C. (1990). *Student/Teacher Achievement Ratio (STAR), Tennessee's K-3 Class Size Study: Final Summary Report, 1985–1990*. Tennessee State Department of Education, Nashville, TN.