How to Improve the Supply of High-Quality Teachers

ERIC A. HANUSHEK *and*
STEVEN G. RIVKIN

When considering schools, one must pay attention to teachers. After all, teachers are the largest single budget item of schools, and many believe that they are the most important determinant of school quality. Yet research does not find a systematic link between teacher characteristics and student outcomes, leading to doubts about many current policy thrusts that are keyed to measurable attributes of teachers and their background.

The relevant research follows four 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. Going beyond that, a second level of studies relates pay and other characteristics of teaching jobs to the characteristics of teachers in different schools and districts and teacher turnover. A third line of research, following naturally from these, relates teacher characteristics to student performance. The failure to find a strong relationship between the contributions of teachers to student achievement and other outcomes, on the one hand, and teacher education, experience, and salaries, on the other, is inconsistent with the popular view of teachers as a key determinant of the quality of education. Finally, the fourth line of research appears to have solved this conundrum by demonstrating both the large impact of teachers on student learning and the lack of explanatory power of traditional quality measures.

The central focus of this paper is to relate these various bodies of research to a set of teacher quality policy initiatives. These proposals can be divided into three broad areas that are not mutually exclusive. First, because salaries

of teachers have fallen relative to other jobs, some argue that an obvious move is simply to restore teacher salaries to their previous position in the earnings distribution to attract better teachers into the profession. Second, states should adopt more stringent qualifications for teachers such as mandatory master's degrees to improve quality. Salary increases are often, but by no means always, recommended along with more stringent qualifications to offset any possible negative impacts on teacher supply. Finally, an alternative set of policy proposals has taken a different tack. These typically advocate less strict instead of stricter requirements in combination with incentives for higher teacher performance and improved school personnel practices.

Research on Teacher Quality

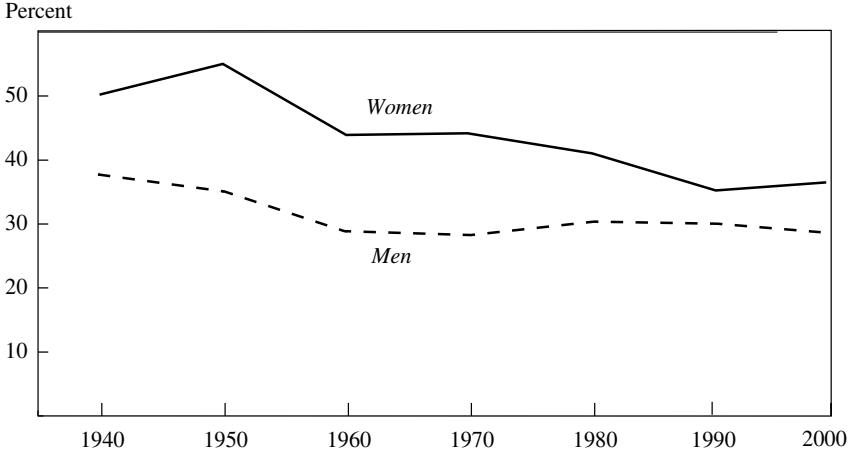
While the evidence related to teacher quality is widely scattered, common themes do emerge. A key distinction is whether or not the investigations are related directly to student outcomes or simply rely upon a presumed relationship.

Aggregate Salary Trends

A starting point in the consideration of teacher quality is the evolution of teacher salaries over time. Figure 1 traces the wages of teachers age twenty to twenty-nine compared with those of other young college graduates between 1940 and 2000. The calculations, done separately by gender, give the proportion of nonteachers with a bachelor's degree or more who earn less than the average teacher.[1]

Over the entire time period since World War II, salaries of young female and male teachers have fallen relative to those for other occupations. However, we have shown that substantial gender differences are evident in the time path of relative salaries.[2] For males, relative salaries fell between 1940 and 1960 but have remained roughly constant afterward. For females, relative salaries started out high—above the median for college-educated females—but then continuously fell. The changes are easiest to see for young teachers and college graduates, for whom the adjustment has been larger, but they also hold for teachers of all ages. In other words, the growth in late-career salaries has not offset the decline in salaries for younger teachers.

Figure 1. Percent College Educated Earning Less Than Average Teacher, by Gender, Age Twenty to Twenty-Nine, 1940–2000



Source: Eric A. Hanushek and Steven G. Rivkin, "Understanding the Twentieth Century Growth in U. S. School Spending," *Journal of Human Resources*, vol. 32, no. 1 (1997), pp. 35–68, updated to 2000.

Others have attempted to go deeper into the structure of teacher supply responses. Frederick Flyer and Sherwin Rosen describe a more formal model of changing female opportunities and its impact on the teaching profession.[3] Darius Lakdawalla extends this to concentrate on the role of productivity changes in competing industries.[4] Technological change, expanded opportunities for women, growth in international trade, and other factors that have increased the demand for highly skilled workers have placed upward pressure on teacher salaries.

The decline in the relative earnings of teachers has likely led to a fall in average teacher quality over this period, though Dale Ballou and Michael Podgursky discuss why the shorter-term implications of a change in relative earnings are less clear-cut.[5] However, the extent of that quality decline is 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.

The Supply of Teachers and Teacher Characteristics

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. Joseph A. Kershaw and Roland N. McKean in one of the first such studies considered how the uniform pay structure in teaching led to shortages in specific areas, such as mathematics and science, where teachers had better outside earnings opportunities.[6] This 1962 study underscored the well-recognized fact that various differences among teachers were important and that policies and institutions had differential effects on teachers with different characteristics.

This research, extended in a variety of dimensions, has a common form:

$$TC = f(P, B, WC), \tag{1}$$

where TC is a specific teacher characteristic, P is pay, B is benefits, and WC is working conditions.

This general form has been considered in a wide variety of circumstances, including entry into teaching, mobility of teachers across schools and districts, and exit from teaching. Four general types of teacher characteristics have received considerable attention: (1) experience, (2) measured achievement or skill, (3) specialty or subject area, and (4) credentials and certification.

The analysis of experience comes in two separate lines of research. First, from consideration of teacher mobility, a general finding is that teachers with more seniority tend to make moves related to the characteristics of the students—such as race, income, or achievement.[7] These movements may be conditioned by district policies, but the underlying analyses identify how attributes of schools affect the characteristics of teachers who are attracted to them.[8] Moreover, Eric A. Hanushek, John F. Kain, and Steve G. Rivkin demonstrate that nonpecuniary factors may have a stronger influence than salaries in determining the location of teachers with differing seniority.[9] Because teachers, particular women teachers, are unresponsive to salary differences, it would take very large bonuses to neutralize the larger teacher

turnout found in the most disadvantaged schools. Second, a large number of studies of exit from teaching find that salaries and outside opportunities have differing impacts on teachers depending on experience.[10] These studies also find varying responses to salaries and other conditions, including suggestions that opportunity costs are much less important than the complementarity of family considerations and school working conditions.[11]

In a different dimension, salaries and other attributes of teaching careers have been seen to influence the average skill level of teachers as measured by achievement scores. The majority of this work has considered entry into the teaching profession.[12] These studies also suggest that the stringency of requirements for teacher certification influences training and career choices.

The consideration of preparation has focused on the varying opportunity costs of teachers with different specialties. For example, following on Kershaw and McKean's analysis, Russell W. Rumberger examines how salaries affect the supply of science and math teachers.[13]

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 proportions of uncertified teachers in central city and lower socioeconomic status 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.

Two important conclusions flow from this literature. First, much of the policy discussion about teacher quality relates to these studies. They indicate how various policies and other factors influence the characteristics of teachers. Second, the importance of these findings depends crucially on the relevance of the identified characteristics for determining student performance and other outcomes; that is, the relationship with effectiveness of teaching.

Teacher Characteristics and Student Achievement

A general approach to the identification of the determinants of teacher quality is to estimate the relationship between student achievement and other outcomes, on the one hand, and specific measures of teacher quality, on the other. While important issues surround how best to conduct this analysis, this approach fits naturally into most policy discussions.

BASIC STRUCTURE. A large number of investigations of student achievement have focused on how various teacher characteristics influence outcomes. These studies take a variety of forms.

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 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, an early study conducted under the auspices of the U.S. government.[14] Since 1966, more than four hundred 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.

However, empirical studies clearly do not analyze this complete model, because complete school and family histories are rarely available. Instead, virtually all work considers the influence of contemporaneous school and teacher inputs and a limited number of family characteristics. This is a serious drawback, because the omission of current and historically important factors that are correlated with the contemporaneous inputs raises a number of questions about how to interpret the results. Perhaps most important is the extent to which any observed association between a school or teacher variable and a student outcome captures 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 fully account for the direct effect of family income on student performance.

An alternative approach attempts to isolate the effects of inputs over a specified period of time. One form of such a value-added model analyzes achievement growth between grade g^* and grade g:

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

The lagged test score O_{g*} may also be included on the right-hand side under somewhat different assumptions and estimation complications. 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.[15] Regardless of the specifics, the fundamental purpose of this approach is the identification of the effects of specific inputs during the specified grades. Generally, when the effects of teachers and schools are approximately additive, the prior achievement measure will summarize the historical inputs.

Though the use of such value-added models mitigates problems resulting from the lack of historical information, an important 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. While this fits into general econometric schemes, it proves to be very restrictive, because most studies use administrative or survey data that typically contain a limited number of characteristics. The most commonly available characteristics, teacher education and experience, are important variables to consider, because they almost always enter into the determination of teacher pay. Thus it is plausible to think that they should be directly related to productivity. And, because of their administrative uses, these data are frequently available for researchers. Yet they explain little of the variation in teacher effectiveness, and even more detailed information about college quality, scores on standardized examinations, or other information leaves 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 factor.

The specific estimation of such models takes many forms. A vast majority of studies investigate student achievement as measured by some form of standardized test, although the investigations cover a range of grade levels, types of schools, areas of the United States and other countries, and other outcomes including years of schooling and future earnings. It is not possible to describe all of the studies, but a summary provides clear evidence on the relationship of key teacher characteristics and achievement.

TEACHER EXPERIENCE AND EDUCATION. 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 aggre-

Table 1. Distribution of Estimated Effect of Key Teacher Resources on Student Performance

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

Source: Eric A. Hanushek, "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis*, vol. 19, no. 2 (1997), pp. 141–64; and Eric A. Hanushek, "The Failure of Input-Based Schooling Policies," *Economic Journal,* vol. 113 (2003), pp. F64–F98.

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

gating the results across studies. Table 1 describes the estimated parameters from studies through 1994 in the United States.[16]

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. More than half of current teachers in the United States 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 find a positive effect, only a minority of all estimates provide statistically significant results.

The interpretation of these results is, however, complicated by concerns about the underlying estimation. The major concern about the quality of the underlying studies involves estimation bias arising from missing data. Many of the studies previously tabulated look at the level of student performance and not at a value-added version. When looking at the level of achievement, the data requirements are huge—encompassing the student's entire past history as relevant to education. Virtually none of the level studies meets such a requirement, but instead virtually all concentrate only on contemporaneous measures of teachers and schools. To the extent that past school inputs are correlated with current ones, which they almost necessarily will be, one is left with biased estimates of the parameters of interest.

A second concern is neglect of state policy variations that might be correlated with the measures of teacher quality. Even though most policy is

made by separate states, variations in state policies are seldom captured in the modeling. 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, omission of them from the modeling is likely to lead to substantial bias in the estimation of the impact of teacher characteristics and other school inputs. Analyses conducted within a single state—where the policy environment is constant—eliminate the largest sources of such bias.[17]

The magnitudes of these problems can be examined by concentrating on the subset of studies that use a value-added approach and information from a single state. Specifically, value-added studies largely eliminate issues of missing historical information. Further, studies drawing their data from within single states lessen any biases from omitted measures of the state policy environment. Table 1 displays the results for high-quality studies; that is, the pay parameters for just the value-added studies conducted within a single state.

The results for the high-quality set of studies underscore the ineffectiveness of further teacher education. The estimates from the thirty-four value-added estimates within individual states are evenly split between positive and negative point estimates, but none of the positive estimates is statistically significant.

The distribution of estimates for teacher experience is largely unaffected by going to the high-quality studies. If anything, the thirty-seven value-added estimates within individual states suggest more strongly that experience has an impact, although still only 41 percent of the estimates are statistically significant.

The stability of the experience estimates across different samples and estimation forms is especially interesting. David Greenberg and John McCall as well as Richard J. Murnane found that 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.[18] Hanushek, Kain, and Rivkin further show that teachers

switching schools or districts tend to move systematically to places where student achievement is higher.[19] 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. However, the value-added models condition on the initial achievement level of students, avoiding the largest complication that schools with high-achieving students might simply be attracting experienced teachers.

The summary table aggregates the estimates of experience across all studies relying on the crude division between positive and negative effects. No consideration is given to the magnitude of estimated effects. More important, because various analyses consider different functional forms, some of the variation in results could result from different estimation approaches and from data samples that are more or less concentrated in different parts of the distribution. For example, Richard J. Murnane and Barbara R. Phillips investigate the impact of experience with spline functions and find nonlinearities, although the estimates differ sharply across data samples.[20] Steve G. Rivkin, Eric A. Hanushek, and John F. Kain also pursue a nonparametric investigation of experience and find that experience effects are concentrated in the first few years of teaching.[21] Holding constant overall teacher quality, teachers in their first and to a somewhat lesser extent their second year tend to perform significantly worse in the classroom.

An alternative interpretation of teacher experience is that it is not that teaching skills improve over time but that teacher experience is an index of selection into teaching. Because a large number of teachers exit from teaching within the first few years, those with more experience are those who remain in teaching for that period of time—and those with few years of experience are a mixture of teachers who will stay for a long time and greater proportions of those who will leave early. Rivkin, Hanushek, and Kain investigate this within their rich longitudinal data set by separating experience effects for teachers who subsequently leave from those who continue teaching.[22] They find that the dominant effect is learning to teach better in the first few years. Following the initial period, however, little additional improvement is evident at least in terms of measured achievement.

TEACHER SALARY. Instead of concentrating on the prior characteristics of teachers that enter into salary decisions, whether or not salary directly relates to student performance can be analyzed. The interpretation is nonetheless frequently muddled. The majority of analyses relate the salary

Table 2. Distribution of Estimated Effect of Teacher Salaries on Student Performance

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

Source: Eric A. Hanushek, "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis*, vol. 19, no. 2 (1997), pp. 141–64; and Eric A. Hanushek, "The Failure of Input-Based Schooling Policies," *Economic Journal*, vol. 113 (2003), pp. F64–F98.

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

levels of teachers to the achievement of students. Yet the salary level for any individual teacher is a composite of pay for specific characteristics (experience, education, and other attributes) and, whenever the analysis crosses individual 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 indication 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 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. Table 2 also 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 only about moves along the schedule and not what might happen with shifts in the entire schedule.

The analysis of salaries as a measure of teacher quality introduces another range of issues. A variety of analyses have highlighted the possibility that nominal salaries in part reflect compensating differentials—for cost-of-living differences, for the desirability of particular schools and their working conditions, or for such other things as urban crime.[23] Most of the

studies considering compensating differentials do not directly associate job-related characteristics and salaries to student outcomes, but instead stop at showing that salaries vary with such characteristics. (An exception is found in the work of Susanna Loeb and Marianne E. Page, 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.)[24] Nonetheless the main thrust is important both for the interpretation of existing studies and for the design of new analyses of the relationship between teacher quality and pay.

TEACHER TESTS. One measured characteristic—teacher scores on achievement tests—has received considerable attention, because it has more frequently been correlated with student outcomes than teacher experience and education or teacher salary. 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.

TEACHER CERTIFICATION. The most pervasive policy action of states aimed at teacher quality is setting certification requirements. While states vary in what is required for certification, the underlying theme is an attempt 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 effective in the classroom. Not only may some potentially good teachers be unable to pass the examinations, but the certification requirements also may discourage others from even attempting to enter the teaching profession.[25] The nature of this trade-off depends in large part on the objectives and skills of administrators who make teacher personnel decisions.[26]

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 controversial. Much of the work is based on specifications that are susceptible to substantial biases from other determinants of achievement, though

a few recent papers provide more persuasive empirical specifications. Andrew J. Wayne and Peter Youngs document the limitations of most studies on certification while reviewing some of the components of certification.[27] A heated debate has taken place on the nature of certification itself.[28]

The overall weight of the evidence suggests that existing credentialing systems do not distinguish well between good and bad teachers. Because many people teach even though they do not have standard credentials, teachers with and without certification can be compared—and little evidence shows that existing hurdles provide much information about performance in the classroom.

Total Teacher Effects

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

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

where t_j is the influence of having teacher j [conditional upon the other inputs, $f'(\cdot)$]. This formulation places some structure on the achievement process, but it avoids the need to explicitly measure the characteristics related to good teaching.

This 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 to the first, 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.

This estimation, frequently referred to as analysis of teacher value-added, is closely related to the work of William L. Sanders and Sandra P. Horn.[29] While the estimation approaches differ, the objective is the same—extracting what individual teachers add to student learning.

A variety of studies have pursued this general approach over the past three decades.[30] Careful consideration of such work reveals the difficulties that must be overcome to estimate the variation of overall teacher effects.[31] For example, teacher effects, school effects, and classroom peer effects are not separately identified if the estimates come from a single cross section of teachers. Hanushek, however, demonstrates the consistency of individual teacher effects across grades and school years, thus indicating that the estimated differences relate directly to teacher quality and not the specific mix of students and the interaction of teacher and students.[32] Rivkin, Hanushek, and Kain go even further and remove separate school and grade fixed effects and observe the consistency of teacher effects across different cohorts, thus isolating the impact of teachers.[33]

The magnitude of estimated differences in teacher quality is impressive. Hanushek 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 with those near the bottom.[34] 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.

A second set of estimates comes from recent work on students in Texas by Rivkin, Hanushek, and Kain.[35] The analysis follows several entire cohorts of students and permits multiple observations of different classes with a given teacher. The authors look at just the variations in performance from differences in teacher quality within a typical school because of the difficulties involved in separating differences in teacher quality from other factors that differ among schools. The variation in teacher quality is large. Moving from an average teacher to one at the 85th percentile of teacher quality (that is, moving up 1.0 standard deviation in teacher quality) implies that the teacher's students would move up more than 4 percentile rankings in the given year. This is roughly equivalent to the effects of a ten-student (roughly 50 percent) decrease in class size. (For a variety of reasons, these are lower-bound estimates of variations in teacher quality. Any variations in quality across schools would add to this. Moreover, the estimates rely on a series of conservative assumptions, which all tend to lead to understatement of the systematic teacher differences.)

Another indication of magnitude is found in Tennessee's Project STAR (Student/Teacher Achievement Ratio) results. The average difference in performance of students in small kindergartens has been the focus of all attention, but the results differed widely by classroom. In only forty out of seventy-nine schools did the kindergarten performance in the small classroom exceed that in the regular classrooms (with and without aides). The most straightforward interpretation of this heterogeneity is that variations in teacher quality are important relative to the effects of smaller classes.[36]

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. The Rivkin, Hanushek, and Kain estimates of teacher performance suggest that having five years of good teachers in a row (1.0 standard deviation above average, or at the 85th quality percentile) could overcome the average seventh-grade mathematics achievement gap between lower-income kids (those on the free or reduced-price lunch program) and those from higher-income families.[37] In other words, high-quality teachers can make up for the typical deficits seen in the preparation of kids from disadvantaged backgrounds.

These family background deficits do not seem to be disappearing, however, because the current school system does not ensure any streaks of such high-quality teachers—particularly for disadvantaged students. In fact, the typical student now is as likely to get a run of bad teachers—with the symmetric achievement losses—as a run of good teachers. A crucial question for policy is the extent to which schools use additional resources or better working conditions to attract higher-quality teachers. Given the lack of simple measures of quality, this question is difficult to answer. However, recent work focusing on a single metropolitan area and measuring teacher quality by the contribution to student learning suggests that, while a statistically significant relationship exists between a suburban district's salary and the quality of teacher hired, the magnitude is small.[38] This is consistent both with the notion that sorting teachers based on characteristics and even an interview is difficult and with the belief that school personnel practices fail to hire, develop, and retain high-quality teachers.

Policy Implications

The review of work on teacher quality provides a number of insights into current policy proposals.

First, attention to teacher quality is warranted, because it is an important determinant of student outcomes. A string of good teachers can overcome the deficits of home environment. Moreover, a string of good teachers can push students with good preparation even further.

Second, legislating "good teachers" is extraordinarily difficult, if not impossible. The currently available data provide little reason to believe that enough is known about good teachers to set appropriate training and hiring standards. The idea behind most certification requirements is to guarantee that nobody gets a terrible teacher. In other words, the general idea is that a floor on quality can be set. But doing this requires knowledge of characteristics that systematically affect performance. The prior evidence does not indicate that this can be done with any certainty.

Third, the current screens used to make judgments on teacher quality are imprecise. Credentials, degrees, and teacher test scores are not consistently or strongly correlated with teaching skill. Thus cracking down on these—such as requirements that only fully certified teachers can enter the classroom—may have little impact on student performance, even if it can be achieved.

Some supporters of the current credentialing movement would nonetheless not concede that basic point. In fact, improving the teaching force is motivation for a variety of policies to tighten up credentials; that is, to make the standards higher and more rigorous. For example, some propose ensuring that all certified teachers have a master's degree. But, because past evidence shows that many of the master's degrees currently obtained are not useful, these proposals offer no strong expectation that quality would improve. The trick for those advocating tightened certification requirements is coming up with standards that are meaningfully correlated with teaching quality, even in the absence of strong evidence.

Fourth, the currently proposed screens are likely to be costly. Most of the current proposals call for tightening up on the entry requirements of teachers. This tightening up would come from increased course requirements in undergraduate school, from new requirements for master's degrees, and from heightened test score requirements for entry into teacher training or for certification. Each of these makes entry into teaching relatively more costly than today. Other things equal, this would reduce the supply of potential teachers and exacerbate any current problems of teacher supply. Salary increases could offset any reduction in supply. The magnitude of the needed increase would depend on the responsiveness of prospective teachers to

salary changes—something about which only rudimentary knowledge is now available.

Fifth, while teacher salaries have slipped relative to other opportunities for college graduates, simply raising all salaries would be expensive and inefficient. Specifically, overall increases in salaries for teachers would have their largest potential impacts on attracting a new group of people into the profession and on retaining existing teachers who would otherwise leave the profession. But as Ballou and Podgursky point out, there is no reason to believe that this will necessarily increase the quality of teachers in the short term.[39] Retaining teachers would be beneficial if high-quality teachers were the ones retained—but the existing data on teacher labor markets do not indicate that this would be the case. Recent work by Eric A. Hanushek and others finds that the average effectiveness of those who exit a large urban district is below the average of those who remain, raising doubts that reducing turnover given the current composition of entering teachers should receive high priority.[40]

Higher 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 from Ballou and Podgursky, while not definitive, suggests that schools are not effective at choosing the best teachers.[41]

A central feature of the strengthened regulation and higher salary approach with few exceptions is the lack of focus on the performance of teachers in the classroom and a de-emphasis on having administrators make personnel decisions. The alternative policy approach is to focus much more on student performance while freeing up the supply of potential teachers. Instead of tightening up on the requirements to enter teaching, the idea is to loosen up on requirements and focus much more on potential and actual effectiveness in the classroom.

A variety of experiments with alternative routes to teaching do not involve traditional certification. The existing evidence on their success or failure is limited, but one careful study of the performance of the Teach for America program by Margaret E. Raymond, Stephen Fletcher, and Javier A. Luque shows generally positive results.[42]

The simple position taken here is: If one is concerned about student performance, one should gear policy to student performance. Existing evidence on schools highlights the substantial variation in teacher quality that exists today, even among teachers with similar education and experience. This

variation appears to result from several factors: differences in 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. The final source of variation may justify substantial differences and flexibility in pay schedules, and more should be learned about the consequences of differentiated pay. However, the variation in skill and effort raises the most difficult set of issues for policymakers, because regulations—including but not limited to certification requirements—are not likely to get at the crux of the issue.

Perhaps the largest problem with the current organization of schools is that nobody's job or career is closely related to student performance. This is not to say that teachers or other school personnel are currently misbehaving. We believe that most teachers and administrators are very hard working and that the vast majority are trying to do the best they can. It is simply a statement that they are responding to the incentives that they currently face, and these incentives do not weight student achievement heavily.

Specifically, policy advice that the evidence strongly suggests is that principals and superintendents must make decisions about teachers based on the evaluation of potential and actual effectiveness in raising student performance rather than a set of prior attributes. While room for improvement in hiring certainly exists, it will always be an imperfect process. The other aspects of personnel management, including mentoring and support, tenure review, and the management of experienced teachers, leave tremendous room for improvement.[43] Such proposals are similar to those developed in more detail in the volume edited by Marci Kanstoroom and Chester E. Finn Jr.[44] They also encapsulate the current experiments being fostered in the Teacher Advancement Program of the Milken Family Foundation.[45]

Existing research demonstrates that principals do know who the better teachers are.[46] While the evidence is not as complete as one might like, the ability to identify teachers at the top and bottom of the quality distribution almost certainly goes further than this, particularly if good tests of student achievement are administered regularly. Unfortunately, little use is made of any such information in the current system, and educators have little experience with the range of possible approaches.

Some evidence has accumulated about merit pay plans, which has not suggested that merit pay as applied to schools has been effective.[47] There is reason to believe that these experiments are, however, too limited.[48] The

historical experiments in merit pay have been extensive in number but limited in the magnitude and character of the incentive scheme.[49]

The measurement of teacher or administrator performance from test score data is a complicated and often opaque process, and test scores are only one out of a number of important student outcomes. Nonetheless much more needs to be learned about the effective use of test scores specifically and outcome information more generally in the evaluation of teacher and administrator performance.

One of the obvious implications is that principals and superintendents must be held accountable for the impact of their hiring, retention, and other management decisions on student achievement. Such structures are not common in education, so little is available to build upon in the actual structuring of such a notion. Moreover, making such active decisions is often difficult and uncomfortable, and the path of least resistance is to grant tenure to virtually all teachers and to refrain from intervening except in extreme cases.

A variety of institutional structures may provide appropriate incentives, and schools across the nation are experimenting with many organizational arrangements including charter schools, school report cards, merit schools, school vouchers, and public school choice. The best way to structure incentives is currently unknown, and describing and evaluating alternative incentive schemes is beyond the scope of this paper.[50] The absence of much variety in incentive systems that have been implemented implies that very little experience or evidence has accumulated. Nonetheless, the experiments now under way offer some hope for learning.

Comment by Richard Rothstein

Two important themes in Eric A. Hanushek and Steven G. Rivkin's paper deserve further examination. First, the authors argue that excellent teachers can make a huge difference in raising student achievement. If disadvantaged students could have a string of excellent teachers, these students' achievement would rise to the level of middle-class students, eliminating the test score gap. Second, they say that identifying excellent teachers for hire and retention is extraordinarily difficult. Econometric analysis has found little correlation between students' achievement and their teachers'

characteristics, such as certification, teachers' test scores, their verbal ability, or their education beyond a bachelor's degree. Because using any of these imperfect characteristics as screens for teacher hiring narrows the pool of potential teachers, schools should be permitted to hire any college graduates whom administrators believe are likely to raise student achievement. And schools should retain only those teachers who do raise student achievement.

The Role of Teacher Quality in Closing the Achievement Gap

No dispute can be raised about the important impact of teachers on student achievement. But presenting the impact as this paper does ("a string of good teachers can overcome the deficits of home environment") can be misleading and dangerous from a policy perspective.

Hanushek and Rivkin base their conclusion on analyses they have performed using student test score data from Texas, showing that low-income students can get middle-class test scores if they have five consecutive years of teachers who are "at the 85th percentile of teacher quality"; that is, more effective than 85 percent of all teachers, or a standard deviation above the mean of effectiveness. This is similar to analyses by William L. Sanders, using data from Tennessee, finding that students gain 50 percentile points from having three consecutive years of teachers who are more effective than 80 percent of all teachers.[51] Sanders calls this consequence of good teaching "awesome." His work has been widely cited to support the claim that good teaching can overcome learning impediments stemming from low socioeconomic status.

Neither Hanushek and Rivkin nor Sanders claims to identify the characteristics of such good teachers other than by a circular description—good teachers can raise student achievement, and teachers are defined as good if they raise student achievement. Yet even if good teachers could be defined more usefully (I understand that Sanders is attempting to do this in Tennessee), it might do little to eliminate the socioeconomic test score gap.

Partly it is a matter of logic. Researchers have known at least since the publication of the Coleman Report in 1966 (and subsequent reanalyses of the data) that both schools and families contribute to student achievement. Most researchers conclude that families contribute considerably more than schools but that does not negate the proposition that, in theory, stronger schools could overcome negative family influences.

Yet Hanushek and Rivkin (along with many others who cite their work) employ their findings to minimize the importance of family and social factors in student achievement. For while they are correct that stronger schools could offset, at least in part, negative family influences, stronger families could also overcome weak schools. Hanushek and Rivkin show that low-income children who have five years of 85th percentile teachers can achieve at middle-class levels. Presumably, low-income children who have 50th percentile teachers but who are the beneficiaries of social policies that raise their family characteristics by a standard deviation would also achieve at middle-class levels.

(Because families have been a more important influence than schools, the achievement gain realized by a standard deviation improvement in teachers could probably also be realized by less than a standard deviation improvement in family characteristics.)

So from a logical viewpoint, improvement in either families or schools (or in a combination of both) could be levers to close the achievement gap. It is curious, therefore, that Hanushek and Rivkin, and Sanders, identify schools alone as the appropriate lever. Why do they suggest, and permit their work to be interpreted as finding, that the achievement gap is primarily a problem of faulty schools and their poor-quality teachers?

In an earlier paper, Hanushek wrote that he frames the issue in this way simply because it is more practical: "While family inputs to education are indeed extremely important, the differential impacts of schools and teachers receive more attention when viewed from a policy viewpoint. This reflects simply that the characteristics of schools are generally more easily manipulated than what goes on in the family."[52]

If this practical conclusion is warranted, then focusing attention on schools alone makes sense. But the practical conclusion is flawed, not in the least because, after so many years of failing to close the achievement gap almost exclusively with school reform, an acknowledgment should be made that characteristics of schools are harder to manipulate than they appear.

Hanushek and Rivkin's illustration of the importance of teacher quality, invoking the provision of an 85th percentile teacher to low-achieving students, is misleading. This policy—providing high-quality teachers to low-achieving students—is not readily achievable if only schools set their goals correctly and were held accountable. Unsophisticated readers will not appreciate what a huge and unattainable improvement a full standard deviation (providing students with 85th percentile teachers) represents. In no

arena can policy reasonably aim for a full standard deviation of gain. Especially in the case of schools, where, as Hanushek and Rivkin note, they cannot identify the characteristics of good teachers (they know they are good only because their results are good), it is inconceivable that any policy could raise teacher quality by a full standard deviation. Even if the recommendation of eliminating hiring screens were adopted, it is hard to imagine that school principals and district offices could become so brilliantly insightful in their evaluation of candidates to be able to identify and hire only those who are as effective as the top 15 percent of the current pool.

As in any policy area, a more reasonable goal might be to begin by trying to improve teacher quality by a tenth of a standard deviation (teachers who are more effective than about 54 percent of all teachers) or even two-tenths (more effective than about 58 percent). Once one accepts that improving teacher quality substantially (but considerably less than a full standard deviation) will not overcome the deficits of home environment, then it follows that other policies are also needed.

The full policy implications of Hanushek and Rivkin's claim need only be elaborated to see how fanciful are the policy possibilities that flow from their paper. To eliminate the achievement gap by means they favor (identifying and then assigning 85th percentile—or higher—teachers to low-income children) would also require ensuring that middle-class children be taught by only 50th percentile—or lower—teachers. If all teachers could be improved so that their skills are comparable to those of 85th percentile teachers today, then presumably the achievement of middle-class children would also rise, maintaining the gap but at a higher level. (Possibly, the gap would narrow because excellent teachers could make more of a difference for low-income than for middle-class children, but the gap would still persist. Nobody claims that better teachers could make no difference for middle-class children.) What kind of policy environment would be required to permit schools to give poor children only superb teachers and middle-class children only mediocre ones? It is hard to imagine. But a more modest goal would be within the realm of possibility. Salary differentials might work to entice, say, 54th or even 58th percentile teachers to schools with disadvantaged children.

Even in this case, however, the Hanushek and Rivkin policy assumes that high-quality teachers can readily be identified. One barrier to the identification of such teachers is the enormity of data required to ensure that

teachers who appear to achieve good results are not simply the beneficiaries of sampling error. Even where multiyear standardized test data seem to exist, the amount of consistent data for teachers and students needed to reliably assess a teacher without such error is unrealistically great. (Hanushek and Rivkin derive their conclusion about teacher effects from a database covering five years.) In attempting to identify excellent teachers, Sanders must compensate for missing data for many students of some teachers. He tries to handle the problem by substituting average results for missing data. A consequence of this choice (no better choice is available) is that good teachers appear less exemplary and poor teachers appear to be better if they teach in schools with more transient student populations.[53]

An additional impediment to the identification of 85th percentile teachers, given no attention in the papers of Hanushek and Rivkin, or in the work of Sanders, is the lack of knowledge regarding whether any substantial intersubject correlation exists in teaching effectiveness. That a string of 85th percentile teachers can raise low-income children to middle-class achievement levels has been detected only in the case of math, even though these data all come from multisubject elementary school teachers. The same teachers who achieve great results in math may not also achieve them in reading. Or in social studies. Or in art, or in music, or in social skills.

An elementary school administrator seeking to hire a good teacher must seek candidates capable of delivering a balanced curriculum. Even in the unlikely event that principals had good quantitative evidence on teacher effectiveness, they would still have to determine whether a candidate who would likely be at the 85th percentile of effectiveness in math and the 55th percentile in reading was superior to a candidate who would likely be at the 70th percentile of effectiveness in both.

To date, no publications have dealt with this problem of intersubject correlations in teacher value-added, although Dale Ballou has observed that "discrepancies will . . . arise across subjects." Ballou adds that

> for reasons probably due to the home environment, more of the variation in student reading performance is independent of school quality than is the case in math performance. As a result, it is harder to detect particularly strong (or weak) performance by reading instructors than by math teachers.[54]

But in elementary schools, reading instructors and math teachers are the same people. For this reason, as well as from conversations I have had with scholars who have worked with Sanders's data, I conclude that the inter-

subject correlations for teaching effectiveness are likely to be small. Hanushek and Rivkin do not investigate this problem.

Was Hanushek correct when he asserted that teacher quality is "more easily manipulated than what goes on in the family"? Surely a full standard deviation improvement in family characteristics is not conceivable. But more modest goals might be reasonable and more easily achievable (or as easily achievable) as teacher quality improvement.

For example, mobility is one of the most serious impediments to higher achievement of disadvantaged children in urban areas. Even an 85th percentile teacher cannot do much for children who enroll in one school after another for relatively brief periods. Because of the need to constantly regroup classes, review material for newcomers, and take time to identify the individual learning problems of newly arrived pupils, even stable students' achievement suffers in schools characterized by high mobility.

Mobility has many causes, but one is the lack of affordable housing in urban areas where real estate values have skyrocketed relative to working-class wages. Housing policy is not something that education researchers think much about, but realistic programs to stabilize low-income housing might have a big impact on student achievement, perhaps comparable to a tenth of a standard deviation improvement in teacher quality.

Many such social policies, some of which are relatively easy to implement, belie the notion that "schools are generally more easily manipulated" than student characteristics. Health problems that are unknown in middle-class children exist in epidemic proportions in low-income communities and contribute to low achievement. Untreated dental cavities, uncorrected vision problems, lead poisoning, environmentally provoked asthma are all easily addressed by policies that are more proven than policies to improve teacher quality. But these health problems will not be addressed if researchers and policymakers continue to focus exclusively on school reform as the way to close the test score gap. School reform (and improvement in teacher quality in particular) is certainly needed, but it cannot do the job alone and may not be the most easily manipulable lever.

One advantage of directing attention to social reforms that might enhance student achievement is that the policy environment in which the gap could be narrowed in this way is more politically realistic. It is not feasible to think the gap could be abolished by assigning only superb teachers to low-income schools and only mediocre teachers to middle-class schools. If the characteristics of superb teachers could be identified, parents of chil-

dren in all schools would demand them. But no such dynamic operates in social policy. If, for example, dental clinics were placed in low-income schools and student achievement then improved (because children without toothaches can concentrate more on their studies), middle-class communities would not then demand that the achievement gap be maintained by providing improved dental care to middle-class children whose care was already adequate.

Nobody would oppose improving teacher quality across the board, simply because the gap would be maintained as middle-class children also benefited from the improvement. Equity and quality improvement are separable challenges. But if closing the gap is the objective, then a focus on policies to improve family characteristics may be more productive than an exclusive focus on teachers and schools.

Identifying and Hiring Excellent Teachers

At the end of their paper, Hanushek and Rivkin note that "the measurement of teacher or administrator performance from test score data is a complicated and often opaque process, and test scores are only one out of a number of important student outcomes." Regrettably, they do not develop this point. Were they to do so, they might look on professional educators' attachment to the certification process with more sympathy.

Taking seriously the notion that "test scores are only one out of a number of important student outcomes" makes it hard to imagine the alternative to certification put forward not only by Hanushek and Rivkin but, as they note, also by Dale Ballou and Michael Podgursky and by Chester E. Finn Jr. and Marci Kanstoroom. If teachers are to be hired and then retained irrespective of certification but on the basis of their students' achievement, a satisfactory method must be devised for measuring this achievement. None now exists. Basic skills achievement in reading and math can be and is measured, but many equally important outcomes, such as good citizenship (not the same as knowledge of history facts), collaborative behavior habits (more highly valued than academic skills by most employers), good judgment (New York courts, for example, say the aim of schools should be to develop intelligent voters and jurors), self-discipline, ethical values, artistic sensibility, a love of learning, and so on, are not now measured. No techniques have been developed for measuring these outcomes, so no alternative is left but to hire teachers who have the characteristics that administrators believe

are likely to generate balanced outcomes, without quantitative evidence to support this belief.

While advocates of eliminating teacher hiring screens typically acknowledge, as Hanushek and Rivkin do, that satisfactory standardized test scores in reading and math are only one of a number of important goals of schools, the implications of this point have not been developed for their recommendations for teacher hiring. One suspects, perhaps unfairly, that this is because their recommendations lend themselves only to test scores, which are easily quantifiable and observable.

If, however, teachers must continue to be hired based on their possession of characteristics deemed likely to generate a balanced set of outcomes, what is the best way to identify whether candidates possess these characteristics? Should principals or other administrators make this judgment based on interviews and college transcripts, or should they rely, in addition, on certification that ensures that the candidate has completed a training program in an accredited teacher-training institution?

Chester E. Finn Jr. has invoked an analogy that is useful for resolving this question.[55] He acknowledges that some professions, medicine in particular, hire only those who have completed professional training programs and whose skills are certified. Few people would advocate hiring college graduates without such certification as physicians, retaining them in the profession only if their patients had a satisfactory survival rate.

But journalism, Finn says, is a different matter. Prospective reporters can, if they wish, attend a professional journalism school before seeking employment. If editors consider this training useful, candidates who have invested in professional training will have enhanced their chances of being hired. But editors are not precluded from hiring candidates without such training and many reporters are hired without professional degrees. Editors retain reporters based on the quality of their production, not on their possession of prehire credentials.

Teaching, Finn says, is more of an art like journalism and less of a science like medicine. Principals should have the same freedom to hire teachers without certification and retain them based on their output, as editors have freedom to hire reporters.

Analogies can go only so far, but this is a useful one. Yet Finn errs when he places education so heavily in the tradition of journalism and so little in the tradition of medicine. In truth, teaching is somewhat like medicine and somewhat like journalism. Medicine is less a science and more of an art than

Finn acknowledges. Teaching is less scientific but still requires substantial specialized knowledge about curriculum, pedagogy, and child and adolescent development. This specialized knowledge is especially critical when good teaching is defined as the production of a balanced set of outcomes, not test scores alone. And it is even more critical when good teachers are being especially recruited to instruct low-income children who come to school less prepared to learn.

I have recently interviewed principals and superintendents around the country about their teacher hiring practices. These administrators were not chosen randomly; I selected those who have better records of minority and majority student achievement and who have reputations of being unusually insightful and effective. These were administrators who, based on student test scores and reputation, were probably more likely than others to have large numbers of teachers who were in the top quintile of teacher effectiveness. I asked the administrators what qualities they seek in prospective teachers and whether they would prefer to choose from a candidate pool that included college graduates who had not received a teacher education degree and certification.

Almost invariably the administrators insisted that they would not willingly abandon the teacher certification process. They wanted to be assured that prospective teachers had a deep knowledge of the science of teaching as well as competence in subject matter. They wanted to know that candidates had extensive student fieldwork experience, not only a full semester of internship (student teaching) but frequent classroom observations, case studies, and practice lessons that characterize the better teacher education programs. Teaching, these reputedly excellent administrators almost uniformly insisted, has become too complex to be undertaken by those whose only qualifications were enthusiasm and academic proficiency.

Consider one principal with whom I met, in a relatively high-poverty elementary school in a Seattle suburb near the headquarters of the Microsoft Corporation. Because Microsoft gives its employees released time to contribute to their communities, this principal has had many volunteers in her school who were employed by the corporation, and she has observed them carefully. Some, especially in the high-tech downturn, wanted to teach full time. But, the principal insisted, both she and the volunteers recognized that they needed professional education training before they were ready to teach. Regardless of their good intentions and what the principal described as their "big hearts," she would not entrust children to these college gradu-

ates unless they returned with a much greater base of knowledge about the science of education.

On the whole, these administrators estimated that, if freed to hire candidates without traditional training and certification, they might choose up to 10 percent of new hires from such an unconventional pool, but no more, and with some trepidation.

This does not mean that the current teacher education and certification process is satisfactory. Many of the principals I interviewed said they have learned from experience to hire only graduates from some teacher education programs, and not others. This makes teaching distinct from medicine. The accreditation process for medical colleges is today sufficiently satisfactory that a doctor who graduated from any domestic accredited college would be certified. The same cannot be said for colleges of education.

But the solution is not to abolish certification requirements. Instead it is to upgrade teacher-training institutions so that they do, in every case, graduate only those prospective teachers in whom the care of children could be entrusted. The National Council for the Accreditation of Teacher Education (NCATE) has recently revised its standards in ways that are reasonably likely to improve the proficiency of prospective teachers. The new NCATE standards require that graduating teachers be evaluated not only by the courses they have taken but also by the quality of student work produced in classes where they have interned and given demonstration lessons. The standards require professors to engage in extensive observation and criticism of student teachers' classroom practice.

It is too soon to say if the NCATE standards will succeed in raising the quality of new teachers. But no reasonable alternative exists to pursuing such reforms. If a cub reporter blunders, no great harm is done. Editors can print a correction and, if the blunders are serious, dismiss the novice. But the harm done by putting children in the care of unqualified teachers for a year or more cannot similarly be corrected. In this respect, teaching is a lot more like medicine, and journalism-like proposals for reform ignore the enormity of a teacher's responsibility.

Beyond improvement in teacher education, more attention should also be paid to the identification and selection of high-quality principals. Because good principals can identify good teachers—not only by quantifiable measures such as teacher or student test scores but also by the application of judgment about teachers' ability to balance a wide range of curricular and child development goals—then hiring and retaining good teachers also

requires hiring and retaining principals who exercise good judgment about teachers. This, in turn, requires doing a good job of identifying and selecting superintendents who exercise similarly good judgment about principals. In the selection of superintendents and principals as well as of teachers, reliance on quantitative measures such as student test scores can be only a part, perhaps only a small part, of the process.

Comment by Michael Podgursky

Eric A. Hanushek and Steven G. Rivkin provide a good review of the status of recent research on teachers and student achievement, no small part of which has been produced by the authors themselves. They highlight an anomaly that has emerged in this research. On the one hand, measurable teacher characteristics such as teacher credentials or experience seem to have little detectable effect on student achievement. Even when the effects of measurable teacher characteristics pass conventional thresholds of statistical significance, the effect sizes are small. On the other hand, a growing body of research based on large longitudinal student achievement data files in states or large school districts suggests substantial differences in classroom effectiveness of teachers. In studies in which these individual teacher effects can be recovered from the estimates and correlated with measurable teacher characteristics, the association between the two is extremely weak. For example, a 2003 paper by Daniel Aronson, Lisa Barrow, and William Sander reports that over 90 percent of the variation in these teacher effects are unexplained by any combination of measured teacher characteristics of the teachers themselves, for example, experience, certification, master's degree, or college attended.[56]

Given such findings, Hanushek and Rivkin are understandably skeptical about proposals to focus educational resources on traditional teacher credentials. They advance the reasonable argument that if the goal of public policy is to raise student achievement, then the focus of policy ought to be on student achievement and not on teacher credentials that have little or no demonstrated link to student achievement.

From an economic point of view, Hanushek and Rivkin's position is uncontroversial and surely represents the consensus of economists who are familiar with research in this area. However, it also highlights a major divide between economists and much of the education policy community on the

question of teacher quality and licensing. Well-intentioned school reformers, aware of the research showing large teacher effects, are eager to tighten up regulation in the labor market as a means to improve teacher quality. Economists by training are wary of occupational licensing and are concerned that the costs of more restrictive licensing are likely to exceed the benefits.

Take the case of teacher exams. Hanushek and Rivkin point out that the evidence linking licensing exam scores and student achievement gains at best suggests only a modest relationship. Thus, even if such a policy had no cost, the benefit of raising cut scores for passing licensing exams in terms of student achievement gains would be very small. However, raising cut scores does have a cost, a point often overlooked in education policy discussions. The cost of such a policy is that it reduces the applicant pool. Local school administrators have other, and typically superior, information on the quality of applicants (for example, direct observation of classroom performance). If teacher quality is idiosyncratic, as the research suggests, then state regulators should give local school administrators leeway to audition as many candidates as possible. Restricting the pool of candidates whom a district can consider imposes a cost on the district. Moreover, this cost falls disproportionately on poor school districts that already have fewer applicants per vacancy than their suburban counterparts. Raising licensing bars risks forcing these districts to hire any certified applicant who walks through the door.

Relative Pay Trends for Female Teachers

Hanushek and Rivkin reproduce in figure 1 a chart they have presented elsewhere showing the decline in the pay of female teachers relative to other female college graduates. This leads them to the rather pessimistic conclusion that, in the long run, the relative quality of female teachers must fall as well.

Simply comparing mean annual earnings of female teachers with nonteachers may paint an overly gloomy portrait. First, as the authors note, the mix of teaching jobs held by female nonteacher college graduates has changed dramatically over the long time period covered in their chart. In 1960, 58 percent of female college graduates who were not teachers were employed as secretaries or in other jobs classified as "clerical" by the U.S. Bureau of the Census, and only 13 percent were in managerial jobs. By

1990 the clerical share had fallen to 30 percent while the managerial share rose to 35 percent, along with the shares of lawyers, doctors, and other previously male-dominated professions. No doubt 2000 census data will reveal a continuation of that trend. Thus, while relative pay was changing, so was the nature of the nonteaching jobs. As the share of managers and other professions increases, so, in all likelihood, do annual hours of work, job stress, travel, and other less attractive job characteristics, for which the higher pay compensates. If the disamenities of the nonteaching jobs have increased over time, then the relative decline in teacher pay has been overstated, particularly for college-educated women with young children.

In addition, the analysis takes no account of the massive increase in the supply of female college graduates over this period. Between 1960 and 2000, the number of bachelor's degrees awarded annually grew by 109 percent for males, but by 413 percent for females. There are now 178,000 more bachelor's degrees awarded annually to women than to men. Among adults age twenty-five or older, the share of men with a bachelor's degree rose threefold, from 9.7 to 27.8 percent over this period, while for women the share grew nearly fivefold, from 5.8 to 23.6 percent.

Teaching is a job that is attractive to women, particularly women with children, and the supply of college-educated women has increased far faster than the demand for teachers. Between 1960 and 2000 the number of public school teaching jobs increased by 110 percent while the number of women with bachelor's degrees grew by 630 percent. So while other employment opportunities have opened up for women, a far larger supply of high- and low-ability women with college degrees is available for all occupations—including teaching. The current experience with alternative certification programs in many states suggests that many high-ability female college graduates are willing to become teachers at the current level of wages and benefits, if entry barriers in the form of preservice seat time in education school pedagogy courses are relaxed. In any event, an analysis of long-term historical trends in the teaching labor market must consider not only relative pay, but also the large increase in the supply of female college graduates.

Only High-Quality Teacher Studies Count

The education research community recognizes that in evaluating the effect of teachers on student achievement a rigorous study design requires

either random assignment of students to teachers or prior controls for student achievement and student socioeconomic status (SES). A 2003 survey of the research on teacher quality sets just such a standard.[57] The reason for this is clear. If the contribution of a teacher to student learning is to be estimated on the basis of a spring test score, then the level of achievement of students when they first enter the classroom in the fall must be taken into account. Even within the same school building, teacher A may have been assigned some particularly difficult students whereas teacher B may have a group with above-average achievement for the school. Even if these two teachers produce identical gains in student achievement, an examination of spring test scores alone would suggest that teacher A was less productive simply based on the students he or she was assigned in the fall.

The ideal way to evaluate teacher productivity would be to randomly assign students to teachers and then test them in the spring (although, even with random assignment, data on prior student achievement would be highly desirable). Random assignment is the scientific gold standard for policy evaluation research. Unfortunately, at present no studies of teachers meet this standard, although several are under way. In coming years a body of research on teachers will emerge that uses randomized study designs, thanks in no small part to standards set by the new Institute for Education Sciences.

However, as Hanushek and Rivkin and other researchers have shown, much can be learned from rigorous analysis of nonexperimental data, particularly from the massive longitudinal student data files being developed in states and school districts that test students annually. These longitudinal databases permit the estimation of value-added models discussed by the authors.

I think it is important that the authors broke out high-quality, value-added estimates in their survey of the literature. In my opinion, studies of teacher effects that do not provide random assignment of students to teachers or prior controls for student achievement and student socioeconomic status should be off the table in terms of research syntheses. The problem is that family background has a powerful effect on student achievement and SES controls in most data sets (for example, free and reduced-price lunch status) are very poor proxies for home educational resources. If teachers in low SES schools tend to have particular characteristics (for example, substandard certification or fewer master's degrees), then evaluations that fail to control for prior student achievement and SES are going to be overestimating the

effect of these teacher characteristics on student achievement. Because all such studies are biased in the same direction, meta-analyzing fifty or five hundred of them moves educational researchers no closer to discerning the true effect. Only high-quality studies deserve attention.

Families versus Teachers

The authors claim that, based on the current longitudinal research, a string of good teachers can offset the effect of poverty for a low-income student.

> The [Steven G.] Rivkin, [Eric A.] Hanushek, and [John F.] Kain estimates of teacher performance suggest that having five years of good teachers in a row (1.0 standard deviation above average, or at the 85th quality percentile) could overcome the average seventh-grade mathematics achievement gap between lower-income kids … and those from higher-income families.

The education policy community has an understandable desire to seize upon research that holds out the promise that good schools and teachers can compensate for unequal family resources. However, these types of calculations are likely to overstate what can realistically be expected from policy. For example, the statistical methodology in Rivkin, Hanushek, and Kain does not permit identification of the effectiveness of any particular teacher, nor does it permit any direct analysis of the stability of these teacher effects over time. My reading of this emerging literature is that some time may pass before such an experiment can be run. One problem is that it is not known ex ante which teachers, if any, are consistently in the 85th percentile and above for five consecutive years. As this type of research matures, educational researchers may be in a position to identify consistently high-performing teachers and run such an experiment (which would then run for five years). However, at this point, the efficacy of such an intervention remains speculative.

Notes

1. 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 Michael

Podgursky, "Fringe Benefits," *Education Next,* vol. 3, no. 3 (2003), pp. 71–76. For the time series comparisons, these omitted elements of compensation would be most relevant if there have been relative changes in the importance of them between teachers and nonteachers over time. Little data currently are available on any such changes.

2. See Eric A. Hanushek and Steven G. Rivkin, "Understanding the Twentieth-Century Growth in U.S. School Spending," *Journal of Human Resources,* vol. 32, no. 1 (1997), pp. 35–68.

3. Fredrick Flyer and Sherwin Rosen, "The New Economics of Teachers and Education," *Journal of Labor Economics,* vol. 15, no. 1, part 2 (1997), pp. 104–39.

4. Darius Lakdawalla, "The Declining Quality of Teachers," Working Paper W8263 (Cambridge, Mass.: National Bureau of Economic Research, 2001); and Darius Lakdawalla, "Quantity over Quality," *Education Next,* vol. 2, no. 3 (2002), pp. 67–72.

5. Dale Ballou and Michael Podgursky, *Teacher Pay and Teacher Quality* (Kalamazoo, Mich.: W. E. Upjohn Institute for Employment Research, 1997).

6. Joseph A. Kershaw and Roland N. McKean, *Teacher Shortages and Salary Schedules* (McGraw-Hill, 1962).

7. See David Greenberg and John McCall, "Teacher Mobility and Allocation," *Journal of Human Resources*, vol. 9, no. 4 (1974), pp. 480–502; Richard J. Murnane, "Teacher Mobility Revisited," *Journal of Human Resources*, vol. 16, no. 1 (1981), pp. 3–19; Eric A. Hanushek, John F. Kain, and Steve G. Rivkin, "Why Public Schools Lose Teachers," *Journal of Human Resources* (forthcoming Spring 2004); Hamilton Lankford, Susanna Loeb, and James Wyckoff, "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis," *Educational Evaluation and Policy Analysis,* vol. 24, no. 1 (2002), pp. 37–62; and Don Boyd and others, "Do High-Stakes Tests Affect Teachers' Exit and Transfer Decisions? The Case of the Fourth-Grade Test in New York State," mimeo (Stanford Graduate School of Education, 2002).

8. Murnane, "Teacher Mobility Revisited."

9. Hanushek, Kain, and Rivkin, "Why Public Schools Lose Teachers"; and Eric A. Hanushek, John F. Kain, and Steven G. Rivkin, "The Revolving Door, " *Education Next,* vol. 4, no. 1 (2004), pp. 77–82.

10. See, for example, Richard J. Murnane and Randall Olsen, "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from Michigan," *Review of Economics and Statistics*, vol. 71, no. 2 (1989), pp. 347–52; Richard J. Murnane and Randall Olsen, "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from North Carolina," *Journal of Human Resources*, vol. 25, no. 1 (1990), pp. 106–24; Peter J. Dolton and Wilbert van der Klaauw, "Leaving Teaching in the UK: A Duration Analysis," *Economic Journal*, vol. 105 (1995), pp. 431–44; Peter J. Dolton and Wilbert van der Klaauw, "The Turnover of Teachers: A Competing Risks Explanation," *Review of Economics and Statistics*, vol. 81, no. 3 (1999), pp. 543–52; Dominic J. Brewer, "Career Paths and Quit Decisions: Evidence from Teaching," *Journal of Labor Economics*, vol. 14, no. 2 (1996), pp. 313–39; Todd R. Stinebrickner, "Estimation of a Duration Model in the Presence of Missing Data," *Review of Economics and Statistics*, vol. 81, no. 3 (1999), pp. 529–42; Todd R. Stinebrickner, "Compensation Policies and Teacher Decisions," *International Economic Review*, vol. 42, no. 3 (2001), pp. 751–79; Todd R. Stinebrickner, "A Dynamic Model of Teacher Labor Supply," *Journal of Labor Economics*, vol. 19, no. 1 (2001), pp. 196–230; R. Mark Gritz and Neil D. Theobald, "The Effects of School District Spending Priorities on Length of Stay in Teaching," *Journal of Human Resources*, vol. 31, no. 3 (1996), pp. 477–512; Richard J. Murnane and others, *Who Will Teach?* (Harvard University Press, 1991); and Benjamin Scafidi, David Sjoquist, and Todd R Stinebrickner, "Where Do Teachers Go?" mimeo (Georgia State University, 2002).

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

11. Scafidi, Sjoquist, and Stinebrickner, "Where Do Teachers Go?"

12. See, for example, Murnane and others, *Who Will Teach?*; and Eric A. Hanushek and Richard R. Pace, "Who Chooses to Teach (and Why)?" *Economics of Education Review,* vol. 14, no. 2 (1995), pp. 101–17.

13. Kershaw and McKean, *Teacher Shortages and Salary Schedules*; and Russell W. Rumberger, "The Impact of Salary Differentials on Teacher Shortages and Turnover: The Case of Mathematics and Science Teachers," *Economics of Education Review,* vol. 6, no. 4 (1987), pp. 389–99.

14. James S. Coleman and others, *Equality of Educational Opportunity* (Government Printing Office, 1966).

15. See Eric A. Hanushek, "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources*, vol. 14, no. 3 (1979), pp. 351–88; Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," Working Paper W6691 (Cambridge, Mass.: National Bureau of Economic Research, 2000, revised); and Petra E. Todd and Kenneth I. Wolpin, "On the Specification and Estimation of the Production Function for Cognitive Achievement," *Economic Journal*, vol. 113, no. 485 (2003).

16. While more studies have appeared since then, they are small in number 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 Eric A. Hanushek, "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis,* vol. 19, no. 2 (1997), pp. 141–64; and Eric A. Hanushek, "The Failure of Input-Based Schooling Policies," *Economic Journal,* vol. 113 (2003), pp. F64–F98.

17. 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 Eric A. Hanushek, Steven G. Rivkin, and Lori L. Taylor, "Aggregation and the Estimated Effects of School Resources," *Review of Economics and Statistics,* vol. 78, no. 4 (1996), pp. 611–27.

18. Greenberg and McCall, "Teacher Mobility and Allocation"; and Murnane, "Teacher Mobility Revisited."

19. Hanushek, Kain, and Rivkin, "Why Public Schools Lose Teachers."

20. Richard J. Murnane and Barbara R. Phillips, "Learning by Doing, Vintage, and Selection: Three Pieces of the Puzzle Relating Teaching Experience and Teaching Performance," *Economics of Education Review*, vol. 1 no. 4 (1981), pp. 453–65.

21. Rivkin, Hanushek, and Kain, "Teachers, Schools, and Academic Achievement."

22. Rivkin, Hanushek, and Kain, "Teachers, Schools, and Academic Achievement."

23. See, for example, Joseph R. Antos and Sherwin Rosen, "Discrimination in the Market for Teachers," *Journal of Econometrics*, vol. 2 (1975), pp. 123–50; Arik M. Levinson, "Reexamining Teacher Preferences and Compensating Wages," *Economics of Education Review*, vol. 7, no. 3 (1988), pp. 357–64; Randall W. Eberts and Joe A. Stone, "Wages, Benefits, and Working Conditions: An Analysis of Compensating Differentials," *Southern Economic Journal*, vol. 52, no. 1 (1985), pp. 74–79; Lawrence W. Kenny, "Compensating Differentials in Teachers' Salaries," *Journal of Urban Economics*, vol. 7 (1980), pp. 198–207; Eric J. Toder, "The Supply of Public School Teachers to an Urban Metropolitan Area: A Possible Source of Discrimination in Education," *Review of Economics and Statistics*, vol. 54, no. 4 (1972), pp.

439–43; Eric A. Hanushek and Javier A. Luque, "Smaller Classes, Lower Salaries? The Effects of Class Size on Teacher Labor Markets," in Sabrina W. M. Laine and James G. Ward, eds., *Using What We Know: A Review of the Research on Implementing Class-Size Reduction Initiatives for State and Local Policymakers* (Oak Brook, Ill.: North Central Regional Educational Laboratory, 2000); Jay Chambers and William J. Fowler Jr., *Public School Teacher Cost Differences across the United States* (Washington: National Center for Education Statistics, 1995); William J. Fowler Jr. and David H. Monk, *A Primer on Making Cost Adjustments in Education* (Washington: National Center for Education Statistics, 2001); and Hanushek, Kain, and Rivkin, "Why Public Schools Lose Teachers."

24. Their study, relying on interstate variations in school completion and teacher pay, faces an analytical trade-off between using aggregate state data subject to potential missing policy information and providing some control for state amenity differences. Susanna Loeb and Marianne E. Page, "Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Nonpecuniary Variation," *Review of Economics and Statistics,* vol. 82, no. 3 (2000), pp. 393–408.

25. See, for example, Murnane and others, *Who Will Teach?*

26. We thank Dale Ballou for providing a clear description of this trade-off.

27. Andrew J. Wayne and Peter Youngs, "Teacher Characteristics and Student Achievement Gains: A Review," *Review of Educational Research,* vol. 73, no. 1 (2003), pp. 89–122.

28. Elements of the debate over the effectiveness of teacher certification can be traced through National Commission on Teaching and America's Future, *What Matters Most: Teaching for America's Future* (New York: 1996); Abell Foundation, *Teacher Certification Reconsidered: Stumbling for Quality* (Baltimore, Md.: 2001); Kate Walsh, "Positive Spin: The Evidence for Traditional Teacher Certification, Reexamined," *Education Next*, vol. 2, no. 1 (2002), pp. 79–84; Dan D. Goldhaber and Dominic J. Brewer, "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement," *Educational Evaluation and Policy Analysis*, vol. 22, no. 2 (2000), pp. 129–45; Dan D. Goldhaber and Dominic J. Brewer, "Evaluating the Evidence on Teacher Certification: A Rejoinder," *Educational Evaluation and Policy Analysis,* vol. 23, no. 1 (2001), pp. 79–86; and Linda Darling-Hammond, Barnett Berry, and Amy Thoreson, "Does Teacher Certification Matter? Evaluating the Evidence," *Educational Evaluation and Policy Analysis,* vol. 23, no. 1 (2001), pp. 57–77. Goldhaber and Brewer, in "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement," 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, in "Does Teacher Certification Matter? Evaluating the Evidence," dispute the interpretation. Christopher Jepsen and Steve G. Rivkin, "What Is the Trade-off between Smaller Classes and Teacher Quality?" Working Paper W9205 (Cambridge, Mass.: National Bureau of Economic Research, 2002), find very small certification effects on teacher value added to mathematics and reading achievement once the nonlinearities in the return to experience are adequately controlled.

29. Their work has directly entered into school decisionmaking through Tennessee law. See William L. Sanders and Sandra P. Horn, "The Tennessee Value-Added Assessment System (TVAA): Mixed Model Methodology in Educational Assessment," in Anthony J. Shinkfield and Daniel L. Stufflebeam, eds., *Teacher Evaluation: Guide to Effective Practice* (Boston: Kluwer Academic Publishers, 1995), pp. 337–76; and William L. Sanders and Sandra P. Horn, "Research Findings from the Tennessee Value-Added Assessment System (TVASS) Database: Implications for Educational Evaluation and Research," *Journal of Personnel Evaluation in Education,* vol. 12 (1998), pp. 247–56.

30. See Eric A. Hanushek, "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," *American Economic Review,* vol. 60, no. 2 (1971), pp. 280–88; Eric A. Hanushek, "The Trade-off between Child Quantity and Quality," *Journal of Political Economy,* vol. 100, no. 1 (February 1992), pp. 84–117; David J. Armor and others, *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools* (Santa Monica, Calif.: RAND Corporation, 1976); Richard J. Murnane, *Impact of School Resources on the Learning of Inner-City Children* (Cambridge, Mass.: Ballinger, 1975); Richard J. Murnane and Barbara Phillips, "What Do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research*, vol. 10, no. 1 (1981), pp. 83–100; and Rivkin, Hanushek, and Kain, "Teachers, Schools, and Academic Achievement."

31. A similar study for developing countries (specifically Brazil) contains consistent findings. See Ralph W. Harbison and Eric A. Hanushek, *Educational Performance of the Poor: Lessons from Rural Northeast Brazil* (New York: Oxford University Press, 1992).

32. Hanushek, "The Trade-off between Child Quantity and Quality."

33. Rivkin, Hanushek, and Kain, "Teachers, Schools, and Academic Achievement."

34. 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 are between teachers at the 5th percentile and those at the 95th percentile. Hanushek, "The Trade-off between Child Quantity and Quality."

35. Rivkin, Hanushek, and Kain, "Teachers, Schools, and Academic Achievement."

36. A discussion of the experiment and overall results can be found in Elizabeth Word and others, *Student/Teacher Achievement Ratio (STAR), Tennessee's K–3 Class-Size Study: Final Summary Report, 1985–1990* (Nashville, Tenn.: Tennessee State Department of Education, 1990). Eric A. Hanushek, in "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class-Size Effects," *Educational Evaluation and Policy Analysis*, vol. 21, no. 2 (1999), pp. 143–63, analyzes the basic experimental results and identifies the variation across classrooms.

37. Rivkin, Hanuschek, and Kain, "Teachers, Schools, and Academic Achievement."

38. Eric A. Hanushek and others, "The Market for Teacher Quality," paper presented at the annual meeting of the American Economic Association, Washington, D.C., 2003.

39. Ballou and Podgursky, *Teacher Pay and Teacher Quality.*

40. Hanushek and others, "The Market for Teacher Quality."

41. Dale Ballou, "Do Public Schools Hire the Best Applicants?" *Quarterly Journal of Economics*, vol. 111, no. 1 (1996), pp. 97–133; and Ballou and Podgursky, *Teacher Pay and Teacher Quality*.

42. See Margaret E. Raymond, Stephen Fletcher, and Javier A. Luque, *Teach for America: An Evaluation of Teacher Differences and Student Outcomes in Houston, Texas* (Stanford University: CREDO, 2001); and Margaret E. Raymond and Stephen Fletcher, "Teach for America," *Education Next,* vol. 2, no. 1 (2002), pp. 62–68.

43. The issues of hiring and retaining district administrators are similar to those for teachers. Little evidence exists that the current requirements for administrator certification are closely related to the effectiveness of administrators. One relevant study is Ronald G. Ehrenberg, Randy A. Ehrenberg, and Richard P. Chaykowski, "Are School Superintendents Rewarded for Performance?" in David H. Monk and Julie Underwood, eds., *Microlevel School Finance: Issues and Implications for Policy* (Cambridge, Mass.: Ballinger, 1988). A recent policy statement also makes recommendations for administrator policy that parallel the lines of thought about teachers that are contained here. See Broad Foundation and Thomas B. Fordham Institute, *Better Leaders for America's Schools: A Manifesto* (Washington: 2003).

44. Marci Kanstoroom and Chester E. Finn Jr., eds., *Better Teachers, Better Schools* (Washington: Thomas B. Fordham Foundation, 1999).

45. Lowell Milken, *Growth of the Teacher Advancement Program: Teaching as the Opportunity 2002* (Santa Monica, Calif.: Milken Family Foundation, 2002).

46. See Armor and others, *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools;* and Murnane, *Impact of School Resources on the Learning of Inner-City Children*. They identify total teacher effects as discussed above and relate them to principals' evaluations.

47. David K. Cohen and Richard J. Murnane, "Merit Pay and the Evaluation Problem: Understanding Why Most Merit Pay Plans Fail and a Few Survive," *Harvard Educational Review,* vol. 56, no. 1 (1986), pp. 1–17.

48. Eric A. Hanushek and others, *Making Schools Work: Improving Performance and Controlling Costs* (Brookings, 1994).

49. For consideration of the available evidence on teacher merit pay, see Elizabeth Lueder Karnes and Donald D. Black, *Teacher Evaluation and Merit Pay: An Annotated Bibliography* (New York: Greenwood Press, 1986); David K. Cohen and Richard J. Murnane, "The Merits of Merit Pay," *Public Interest,* vol. 80 (1985), pp. 3–30; Cohen and Murnane, "Merit Pay and the Evaluation Problem"; Dale Ballou and Michael Podgursky, "Teachers' Attitudes toward Merit Pay: Examining Conventional Wisdom," *Industrial and Labor Relations Review*, vol. 47, no. 1 (1993), pp. 50–61; Ballou and Podgursky, *Teacher Pay and Teacher Quality*; Elchanan Cohn, "Methods of Teacher Remuneration: Merit Pay and Career Ladders," in William E. Becker and William J. Baumol, eds., *Assessing Educational Practices: The Contribution of Economics* (MIT Press, 1996); and James A. Brickley and Jerold L. Zimmerman, "Changing Incentives in a Multitask Environment: Evidence from a Top-Tier Business School," *Journal of Corporate Finance,* vol. 7 (2001), pp. 367–96.

50. See Hanushek and others, *Making Schools Work*.

51. William L. Sanders and June C. Rivers, *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement* (University of Tennessee, Value-Added Research and Assessment Center, November 1996).

52. Hanushek, "The Trade-off between Child Quantity and Quality," p. 106.

53. I am indebted to Haggai Kupermintz of the University of Colorado for sharing this observation.

54. Dale Ballou, "Sizing Up Test Scores," *Education Next* (Summer 2002), p. 12.

55. Education Commission of the States, "Spring Steering Committee Meeting: Linda Darling-Hammond–Chester Finn Jr. Quality Teaching Debate" (March 26, 2000) (www.ecs.org/ [October 21, 2003]).

56. Daniel Aaronson, Lisa Barrow, and William Sander, "Teachers and Student Achievement in Chicago Public Schools," mimeo (Federal Reserve Bank of Chicago, 2003).

57. Wayne and Youngs, "Teacher Characteristics and Student Achievement Gains."