LACEA 2001
MONTEVIDEO URUGUAY

Eric Hanushek

Can Governments Legislate Higher Teacher Quality?

Can government legislate higher teacher quality?

Eric A. Hanushek
Steven G. Rivkin

September 2001

Paper prepared for
Latin American and Caribbean Economic Association (LACEA)
Montevideo 2001 Conference
October 18-20, 2001

Abstract

This paper discusses different perspectives from which to assess teacher quality and develops a conceptual framework focusing on outcome based measures of instructional performance. Results from a series of analyses on academic achievement in Texas public schools are used to describe both the variation in teacher effectiveness and the extent to which that variation is captured by observable measures of teachers and classrooms. These empirical findings form the basis of a discussion of alternative policy approaches to raising quality, particularly for economically disadvantaged students.

CONTENTS

Can government legislate higher teacher quality?

By Eric A. Hanushek and Steven G. Rivkin*

Much of the discussion of school policy involves teachers in one way or another, often focusing on the need for improved hiring, retention and training. Current teacher shortages in particular subjects such as mathematics and special education, often concentrated in poor, inner city or rural communities, are projected to be much more widespread in the coming years. The combination of retirements and the projected future high demand for college educated workers in other fields adds to the sense of urgency regarding the need to improve the quality of classroom instruction as measured by its contribution to student achievement and other outcomes. In the face of these concerns, a number of policy documents have garnered the attention of educators, administrators and politicians (e.g., Darling-Hammond (1997), Finn, Kanstoroom, and Petrilli (1999; Finn and Kanstoroom (1999)). These varied approaches often appeal to core constituencies sympathetic to particular ideological perspectives. However, the actual effectiveness of specific recommendations and expenditures rests on the validity of assumptions about the market for teachers and the determinants of instructional quality including the institutional structure of schools, teacher training, experience and class size.

In this paper we consider the evidence on teacher quality and implications for education policy. We begin with an overview of the estimation and interpretation of educational production function models that are designed to identify the importance of teachers. Next we present estimates of these models based on unique matched panel data for the state of Texas that permit us to identify the variation in teacher effectiveness and specific determinants of teacher

---

* Stanford University, National Bureau of Economic Research, and University of Texas at Dallas; Amherst College and University of Texas at Dallas, respectively. This research has been supported by the Packard Humanities Institute.

quality. In the final section of the paper we consider a number of policy proposals designed to improve teacher quality in light of the empirical evidence.

Measuring Teacher Quality and Its Determinants

Students and parents refer often to differences in teacher quality and act to ensure placement in classes with specific teachers. Yet this emphasis on teachers is largely at odds with common interpretations of empirical analyses. Specifically, there has been no consensus on the importance of specific teacher factors, leading to the common conclusion that the existing empirical evidence does not find a strong role for teachers in the determination of student outcomes. It may be that parents and students overstate the importance of teachers, but an alternative explanation supported by recent research is that measurable characteristics such as teacher experience, education, and even scores on standardized tests explain little of the true variation in teacher effectiveness.

In this section we begin with a comprehensive model of student achievement and move to identify the importance of teacher quality. Importantly, we do not focus solely on measurable characteristics of teachers or schools as is typically done in this literature. Our objective is to obtain bounds on the magnitude of teacher effects, regardless of our ability to identify and measure any specific components. This restricted formulation has less information than one with a well-specified set of underlying observable characteristics, but both past work and our own summarized below have been unable to provide a very rich representation of the important attributes of teachers and schools.

A. Previous investigations—Basic Formulation

The implied model that lies behind most empirical studies of student achievement, while often not specifically described, has the general form:

$$(1) \quad A_{ijgs} = \sum_{\tau=0}^{g} X_{i\tau} + \sum_{\tau=0}^{g} C_{js\tau} + \sum_{\tau=0}^{g} S_{s\tau} + \sum_{\tau=0}^{g} \gamma_i + \sum_{\tau=0}^{g} \varepsilon_{ij\tau s}$$

Equation (1) describes student achievement (A_{ijgs}) for individual i in classroom j and grade g of school s. Achievement is a function of the entire histories of family background (**X**); of specific experiences in classrooms (**C**); of school quality (**S**); of inherent student abilities[1] (γ); and of random errors (ε).[2] For notational ease, each of the inputs is measured in achievement units such that the coefficient is 1. The classroom effects include specific teachers for each grade.[3] "Inherent abilities" refer to cognitive skills and other motivation and personality traits that effect achievement but that do not change during the school years.

In large part due to data limitations, the models actually estimated typically differ significantly from equation (1). By far the most common approach involves cross-sectional estimation that relates achievement at a single point in time to contemporaneous measures of observed family and school inputs such as:

$$(2) \quad A_{ijgs} = X_{ig}\alpha_X + C_{jg}\alpha_C + S_{sg}\alpha_S + \eta_{ijgs}$$

It is easy to see from equation 1 that the error term in this cross-sectional equation is quite complicated:

[1]The isolation of inherent student abilities for these purposes does not rely on any presumption about their source (genetic, environmental, or an interaction of these). Any fixed learning differences that are formed before school entry will be incorporated in this term.
[2] For expositional purposes we leave out preschool experiences and begin consideration of achievement with kindergarten (g=0). This choice is not meant to downplay the importance of preschool experience but only to recognize that those experiences do not affect the analytical development here.
[3]The subsequent development also allows for a classroom-school interaction that captures not only curricular and management differences that effect specific classrooms (grades) differently within schools but also possible nonlinearities in school experiences.

(3) $$\eta_{ijgs} = \sum_{\tau=0}^{g-1} X_{i\tau} + \sum_{\tau=0}^{g-1} C_{j\tau} + \sum_{\tau=0}^{g-1} S_{s\tau} + \sum_{\tau=0}^{g-1} \gamma_i + \sum_{\tau=0}^{g} \varepsilon_{ij\tau s}$$

The underlying parameters (the α's) of the empirical specifications are meant to describe the causal relationship between observed family and school variables and outcomes. As is well known, however, the key to interpretation of estimates from such models rests upon the correlation of the measured inputs (X_g, C_g, S_g) with the composite error term, η_g.[4] Since it is implausible that the contemporaneous measures of the various inputs are uncorrelated with their historical values, the α's confound the effects of the specific measured elements with all historical factors with which they are correlated. Thus, as a simple example, to the extent that some families seek out particularly good teachers and schools, current measures of C and S are likely correlated with historical family elements as well as historical teacher and school inputs.

The reality is even more complicated than the structure described. The composite error term described in equation 3, as written, implies that the true contemporaneous inputs are perfectly measured; e.g., that the specific measures of teachers accurately capture all of the relevant dimensions of the classroom (C) for grade g. Any imperfectly measured dimensions, however, introduce bias into the estimation unless they are uncorrelated with the included inputs. Past analysis, however, has demonstrated the difficulty in accurately measuring variations in teacher and school quality as well as family background. In sum, a lack of information on specific historical and current inputs makes identification and estimation of general achievement models virtually impossible.

B. Value-Added Formulations

Consider, however, focusing on achievement growth or the rate of learning during a specified time period. The advantage of the growth formulation is that it eliminates a variety of

---

[4]Where there is no confusion, extraneous subscripts are dropped for expositional purposes.

confounding influences including the prior, and often unobserved, history of parental and school inputs. This formulation, frequently referred to as a value-added model, explicitly controls for variations in initial conditions when looking at how schools influence performance during, say, a given school year.[5]

The precise estimation approach found in the literature does vary. At times, initial achievement is added to the right hand side of a regression equation, possibly with corrections for measurement error. At other times, simple differences or growth rates in scores are analyzed. See Hanushek (1979) for a discussion of the value- added model.

To see the implications, consider writing equation 1 for achievement in grade g-1 (which will involve cumulative inputs through that year) and taking the difference in achievement for grade g and g-1. Putting the model in terms of achievement gains eliminates the cumulative impacts through g-1 and puts everything in terms of contemporaneous values of the inputs, such as:

$$
\begin{aligned}
(4) \quad \Delta A_{ijgs} &= A_{ijgs} - A_{ijgs-1} \\
&= X_{ig}\beta_X + C_{jgs}\beta_C + S_{gs}\beta_S + \gamma_i + \varepsilon_{ijgs}
\end{aligned}
$$

Formulations similar to equation 4 have been estimated in a variety of circumstances where two years of individual data are available (see, for example, Murnane (1975) or Summers and Wolfe (1977)). Yet despite controls for initial achievement conditions (A_{g-1}), the inability to measure accurately all factors still raises the possibility of inconsistent estimates of school and teacher effects.

[5]One consequence is that the parameter estimates capture effects only for the specific period, ignoring any continuing impacts of inputs at an earlier age. See Krueger (1999) for a discussion of this issue. However, without detailed information and knowledge of the full cumulative achievement production process, it is virtually impossible to isolate the cumulative effects of specific school factors.

Two issues arise immediately in the general context of estimating equation 4. First, the formulation recognizes that students differ in ability (γ), but such differences are seldom included in any way in the estimation. Second, while not generally mentioned in such estimation, a variety of selection effects can enter. Parents make housing decisions that consider the quality of school districts and schools. And both parents and school officials enter into placement decisions for individual classrooms. Both of these factors raise the potential of biased estimation, since they suggest reasons why the unmeasured ability and selection influences (captured by $\gamma+\varepsilon$ in equation 4) may be correlated with the observed factors of interest, X and S.

We introduce three elements to the estimation in order to deal with these factors. First, with longer panel data on individuals, it is possible to remove a student fixed effects from equation 4. In essence, we allow each student to have a specific expected rate of achievement growth. By doing this, we explicitly estimate a term that includes γ. (Note, however, that we do not separately identify innate ability differences because the student fixed effect incorporates any family factors, peer influences, or school effects that are constant over the period of the panel. It is not possible to separate the separate elements of time invariant differences in achievement growth).

Second, in order to understand the effect of various components of teachers and schools, it is possible to concentrate on variations within schools. The selection of residential location by parents can be accommodated by eliminating the influence of school choice per se and by seeing how teacher characteristics and other things influence patterns of achievement growth within each school. Of course, while looking at variations within schools, the estimation will necessarily involve patterns over time. Therefore, the actual application will have to be sensitive to factors that might change over time within the school and that might be correlated with either parental selection or other things that could influence achievement. These issues are discussed below within the context of the specific estimation.

Third, there is extensive anecdotal evidence suggesting widespread placement decisions and resource decisions within schools and grades. For example, principals may attempt to match particular students with particular teachers. They might also adjust class sizes within grades to provide for the best match of students and resources, as with compensatory use of small classes. Such within grade selection clearly can distort estimation of the achievement parameters by mixing the effect of various factors with the placement of students. The within grade selection into specific classrooms and instructional settings can be dealt with by aggregating teachers and resources across a grade within a school (which is equivalent to instrumental variable estimation using an indicator for school and grade as the instrument).

The combined application of these suggests a modified version of equation 4 such as:

$$(5) \quad \Delta A_{igs}^{c} = X_{ig}^{c}\beta_X + S_{gs}^{c}\beta_S + \gamma_i^c + \delta_{gs} + \varepsilon_{igs}^{c}$$

In this formulation, we begin by adding an identifier for student cohort (c) since estimation of the model requires multiple cohorts. There is an explicit consideration of both an individual intercept, or fixed effect (γ_i^c), and a school-by-grade fixed effect (δ_{gs}) that is constant across the period of observation. The school resources (S_{gs}^c) include those factors that vary both across time and across grades. Notice that the specific classroom of the student (j) indicated previously is no longer relevant (because of aggregation to the grade level in each school) and is left off.

Estimation of this formulation extends previous work on value-added models by incorporating explicitly differences in individual factors that affect achievement growth, including importantly any ability differences. By estimating the model on within school variation (which follows from allowing for different school intercepts), the most significant sorting of students across schools is eliminated. Finally, by instrumenting by grade within school, any classroom assignment rules that might affect the estimation are eliminated.[6]

The most common approach to estimation of value added models like equation 4 (and level models such as equation 2) is to include specific descriptors of the family and classroom factors (X and S). For example, families are generally proxied by measures of socio-economic status such as income or parental education; teachers are often measured by amount of education and experience or by salary levels; classroom characteristics commonly involve class size. We pursue this approach in our initial empirical work, because this is especially relevant to policy discussions involving specific determinants of teacher quality. For example, it relates directly to various proposals about certification issues to the extent that characteristics of certification are incorporated in the measured effects of S.

However, prior work has tended to show that such measures capture little systematic variation in student performance. Therefore, we also consider an alternative estimation strategy that does not focus on measurable characteristics of teachers. Rather it estimates the variance in teacher quality based on differences in student performance across teachers. This approach builds on the work by Hanushek (1971), Armor et al. (1976), and Murnane and Phillips (1981)) that estimated teacher fixed effects based on between classroom differences in student achievement gains. However, the aforementioned non-random nature of classroom assignment could confound the effect of teachers with differences in student and other school characteristics. Hanushek (1992) provides some suggestive evidence that teachers are the primary component by showing that classroom gains for individual teachers tend to be highly

[6] The estimation assumes that the separate effects are additive and that a "good match" of students and teachers is not important. In other words, a good teacher has a similar effect on student gain regardless of whether the student is strong or weak.

correlated across time (for different groups of students). Yet some ambiguity remains because of the small number of observations of multiple teachers.

We extend this prior work on covariance models of achievement growth in order to isolate better the variation in teacher effectiveness. Panel data spanning several grades permit observations of how individual students achieve with different teachers. At the same time, data on multiple cohorts permit observations of how an individual teacher does with different groups of students. These two vantage points on schools provide the possibility of separating teacher effects from those of student abilities and selection and other factors that determine achievement in any classroom.

The following section describes the data used in the estimation of the achievement models. We employ these data in two parallel analyses: the first considers traditional measures of schools and teachers (as in equation 5), while the second analyses variations in quality within the general fixed effect framework.

The UTD Texas Schools Project Database

The data used in this paper come from the data development activity of the UTD Texas Schools Project.[7] Its extensive data on student performance are compiled for all public school students in Texas, allowing us to include the universe of students in the analyses. The data also contain a limited number of individual and family characteristics including gender, race, ethnicity, and eligibility for a free or reduced price lunch. We use 4th, 5th, and 6th grade data for three cohorts of students: 4th graders in 1993, 1994, and 1995, and also add third grade data

[7] The UTD Texas Schools Project has been developed and directed by John Kain. Working with the Texas Education Agency (TEA), this project has combined a number of different data sources to compile an extensive data set on schools, teachers, and students. Demographic information on students and teachers is taken from the PEIMS (Public Education Information Management System), which is TEA's statewide educational database. Test score results are stored in a separate database maintained by TEA and must be merged with the student data on the basis of unique student IDs. Further descriptions of the database can be found in Rivkin, Hanushek, and Kain (2001).

for the youngest cohort. Each cohort contributes two years of test score gains.[8] Students who switch public schools within the state of Texas can be followed just as those who remain in the same school or district, a characteristic we use in our analysis.

The Texas Assessment of Academic Skills (TAAS), which is administered each spring, is a criterion-referenced test used to evaluate student mastery of grade-specific subject matter. We focus on test results for mathematics, the subject most closely linked with future labor market outcomes. We transform all test results into standardized scores with a mean of zero and variance equal to one. The bottom one percent of test scores and the top and bottom one percent of test score gains are trimmed from the sample in order to reduce measurement error. Participants in bilingual or special education programs are also excluded from the sample because of the difficulty in measuring school and teacher characteristics for these students.

Importantly, the student database can be linked to information on teachers and schools through the school IDs. The school data contain detailed information on individual teachers including grade and subject taught, class size, years of experience, highest degree earned, and student population served.[9] While individual student-teacher matches are not possible, students and teachers can be uniquely related to a grade on each campus. For a subset of new teachers just entering the Texas public schools, the standard teacher certification test scores are also available.[10] Each student is assigned the school average teacher test score values, class size, proportion of teachers with a master's degree, and a series of variables capturing the distribution

[8] For the youngest cohort, fourth and fifth grade gains are used in the education production function analysis, while fifth and sixth grade gains are used in the analysis of the variance of teacher quality.
[9] Considerable attention was given to the elimination of measurement error in the school variables. Unlike most educational studies, we concentrate specifically on the actual class size, as reported by regular classroom teachers, instead of the more common pupil-teacher ratios for a school. We have access to longitudinal information on key data and can therefore adjust reports for inconsistencies that occur over time. Morever, average class size for the estimation is the average of teacher reports of class size for that grade, and such averaging further reduces measurement error.
[10] The average test scores are computed over the subset of teachers who have nonmissing test score data. A dummy variable for each test is set equal to one for those students in grades where none of the teachers report a score for that particular test.

of teacher experience for mathematics teachers in regular classrooms for the appropriate grade and school year.[11] The mathematics teacher may actually teach all subjects or just mathematics.

Empirical Models of Achievement

A. The Determinants of Instructional Quality

A variety of concerns have been registered about previous estimation of education production functions. Most of these stem from consideration of the process through which students and teachers are placed into classrooms. The inability to control fully for differences in family background could potentially bias the estimated effects of specific teacher characteristics, as could the failure to account for the way in which schools assign teachers to classrooms. These concerns have been most frequently raised in terms of class size considerations, where schools may assign lower achieving students to smaller classes, but it could also hold for other school resources.

The use of achievement gain rather than level and the removal of the fixed student and school effects in equation 4, which is possible with the stacked panel data, address most concerns that are typically raised. Arguments about simultaneity arising from compensatory resource allocations based on previous student performance are directly eliminated, since the level and expected rate of gain of achievement are explicitly dealt with through the investigation of achievement growth, ΔA, and the estimation of the individual fixed effects. Removal of the school fixed effects also eliminates the confounding influences of fixed school factors that may

[11]As Boozer and Rouse (1995) and others have pointed out, it is important to separate regular and special education students because class size and possibly other characteristics differ dramatically by population served and because special education students are much less likely to take tests. If the proportion of students in special education classes or the gap between regular classroom and special education class size differs across schools, estimates of the effect of class size based on the entire school average will be biased. Our measure of class size is the average class size for regular classrooms in specific grades and subjects. Thus, it is not contaminated by special purpose classes. Separate analysis of special education is found in Hanushek, Kain, and Rivkin (forthcoming).

be correlated with the included variables. In other words, the simple patterns of class size policy adjustments based on past individual student achievement levels or growth that motivate alternative estimation approaches are fully accounted for in this estimation context.

Table 1 reports the regression results for all students combined and separately by student income. In addition to coefficients from fixed effect specifications, the table contains estimates from baseline level and value added models for the sake of comparison with prior analyses. The absolute values of Huber-White t-statistics adjusted for the clustering of students into schools are reported for all coefficients.

Our attention is focused on the estimates that extract student and school fixed effects. Nonetheless, it is useful to first compare these with the more standard estimates of the level of achievement and of value-added models that do not control for individual differences. These less complete models differ in a variety of details, but it is the instability of estimates that deserves the greatest attention.[12] The model of sixth grade achievement, for example, suggests some stronger impacts of teacher experience and teacher test scores and less impact of class size than the more complete models. However, the virtual certainty that such a model confounds true variable effects with the influences of other factors leaves little confidence that any of these factors exhibits a causal impact.

The fixed effect results for the entire sample are reported in Column 3. First we discuss the results for teacher experience. Based on preliminary evidence showing that the relationship between achievement and experience is highly nonlinear, we include the shares of teachers in a number of achievement categories (ten to twenty years of experience is the omitted group)

---

[12] The variation in results across different specifications, and the significant impacts with the "wrong sign" – such as for teachers graduate education in the 6^{th} grade achievement estimates – is quite consistent with the findings of past work (see Hanushek (1997)). Our interpretation here is that these measures are very poor proxies for teacher quality, leaving them subject to a variety of biases in different samples and different estimation approaches but providing little insight into the achievement generation process.

rather than simply average years of experience. The results support the notion that those in their first, second, or third year do worse than more experienced teachers but that additional years of experience have little impact.[13]

Notice that the teacher experience effects conceptually combine two very distinct phenomena. First, new teachers may need to go through an adjustment period where they learn the craft of teaching. Second, a number of teachers discover that they are not well matched for teaching and subsequently leave the profession within the first few years (between entry and the end of two years, 18 percent of teachers will leave the Texas public schools and another 6 percent will switch districts). If those who exit after one year of teaching are on average worse than those who remain in teaching, the coefficient on proportion of teachers in their first year will be negative even in the absence of any learning on the job. More generally, the estimated parameters in Table 1 will combine the effect of on-the-job learning and selective exiting.

Separate estimates (not shown) modify the basic estimation of excluding those who leave teaching or switch districts. While the estimates vary somewhat with the sample and are noisy in the 4th grade, the similarity of the estimates across samples indicates that learning is the dominant element of the experience effect and that the average quality of those who quit teaching after one year is similar to the average quality of those who remain.

The second teacher characteristic considered is post-graduate education. Consistent with previous work, there is little or no evidence that a master's degree raises the quality of teaching. All estimates are small (or negative), and none is significant at the 5 percent level. This result has important implications for policy that will be discussed below.

---

[13] Rivkin, Hanushek, and Kain (2001) further disaggregates the experience effect by grade and finds that this pattern holds in 4th and 5th grade but not in 6th grade. New teachers average student gains that are lower by 0.24 standard deviations in 4th grade and 0.12 in 5th grade, while the effects for first year teachers are slightly more than half as large in each grade. Also, from contrasting models with just school fixed effects and those with school-by-grade fixed effects, there is no evidence that the within school distribution of experience across grades introduces an upward bias to the estimates. Indeed, the inclusion of school-by-grade fixed effects increases slightly the estimated effects of experience.

The final set of teacher characteristics includes scores from three certification tests taken by elementary school teachers: education development (a test of pedagogy); mathematics; and comprehensive knowledge. Teachers may complete one, two or all three tests depending upon the specific certification being sought. Because this program has been in effect only for recent years, only a subset of teachers have been tested though most grades do have at least one teacher who reports a test score for at least one of the tests (see Table Appendix Table a1 for information on the proportions of students with no teachers who report specific test results).

Table 1 shows that none of the tests are significantly related to achievement in the preferred fixed effect specifications. Because of the measurement error introduced by the fact that only a portion of teachers report scores, we also created samples restricted to schools in which all teachers report a given test score. The results (not reported) continue to show no evidence that these tests are significantly related to teacher quality as measured by value added to test score gains.

These results differ from prior work by Ferguson (year). One possibility is that the material examined in these three examinations is unrelated to instructional quality while that contained in the test scores used by Ferguson is strongly related. On the other hand, the very different results in the levels specification suggest that specification error could be a problem in prior work. Regardless, the finding that test scores currently used in the certification process are not significantly related to achievement raises a number of policy questions that we return to below.

Class size is the final determinant of instructional quality included in the analysis. While it is not a characteristic of teachers per se, class size has been and continues to be at the center of education policy discussions. The results reveal small but highly significant relationships between achievement gains and class size. A five student reduction in class size – which amounts to between a 20 and 25 percent reduction depending upon the grade – yields an average movement of less than 1.5 percentile points in the distribution. While slightly larger for disadvantaged students, the overall effect remains small even there. Note, however, that these results are average effects for grades 4-6, while much of the attention to class size issues has revolved around earlier grades. Moreover, our previous work on Texas does suggest that the effects of smaller classes for disadvantaged students are larger in 4th grade and decline with age see Rivkin, Hanushek, and Kain (2001).

Other than initial experience, the results suggest that specific characteristics of teachers do not have a significant impact on achievement. This result, mirroring that found in much of the literature (Hanushek (1997)), has often been interpreted as indicating that teachers are not very important in the determination of academic achievement, particularly in comparison to families. This finding, however, mistakenly confuses the effects of measured attributes of teachers with the true impact of teachers – the subject of the next section.

B. Total Variance of Teacher Quality[14]

The finding that conventional measures of teacher attributes do not capture much of the differences in student performance could indicate either that teachers have little influence on student performance – an implausible finding at least from the viewpoint of most parents – or that the measures do not adequately capture differences among teachers. In order to investigate this issue, we return to the estimation of total teacher effects.

Our approach is designed to provide lower bounds estimates of the variation in teacher quality that exists in Texas elementary schools. In order to avoid the confounding influences introduced by family choices of neighborhoods and schools, we concentrate solely on the variation in teacher performance that is found within schools. Therefore any variation in teacher quality between schools is ignored. Further, we identify the variation in quality through the effect of teacher turnover on student achievement growth in different cohorts. This

[14] This section summarizes the estimation contained in Rivkin, Hanushek, and Kain (2001). For a more complete discussion of the estimation procedures, see that analysis.

identification relies upon some strong assumptions – namely, that the quality of teachers hired one year in a given school is independent of the quality of teachers hired the next year and that a given teacher's performance is constant across all years. If either of these assumptions is untrue, the estimates of the variation in teacher quality are again biased downwards. See Rivkin, Hanushek, and Kain (2001) for a comprehensive description of this methodology.

Table 2 provides a series of alternative estimates of the standard deviation of teacher quality based on models that include a variety of additional controls designed to rule out alternative explanations (other than teacher quality) for the observed differences in student achievement growth. The story is a simple one. The standard deviation of quality measured in terms of standardized growth in math scores *within* schools is at least 0.11. This parameter is estimated quite precisely, and the estimate is not affected much by the form of estimation. We present estimates from simple achievement growth models and models that include student, school, and school-by-grade fixed effects. Further, because these are estimated on variations in achievement growth across grades and cohorts, we explicitly consider the potential effects of changes in leadership (principals and superintendents) that might affect teacher performance. None of these factors has any real impact on the estimates.

The magnitude of the teacher effects is striking. Consider the impact on students of moving up the mathematics teacher quality distribution by one standard deviation (Rivkin, Hanushek, and Kain (2001)). This is roughly equivalent to lining up all teachers from least to most effective and passing one third of the teachers. The impact of moving up one standard deviation in teacher quality raises mathematics test scores by roughly ten times as much as the very expensive policy of reducing average class size by one standard deviation (roughly two to three students in Texas). The comparison of teacher effects and of family differences provides another perspective. The gain from having a very good teacher (one standard deviation better) rather than an average teacher for four to five years in a row is roughly equal to the average mathematics test score differential between low income students eligible for a subsidized lunch and higher income students not so eligible.

These results are consistent with earlier estimates. Looking at the range of quality for teachers within a single large urban district, teachers near the top of the quality distribution can get an entire year's worth of additional learning out of their students compared to those near the bottom (Hanushek (1992)).[15] That is, a good teacher will get a gain of 1.5 grade level equivalents while a bad teacher will get 0.5 grade level equivalents for a single academic year.

This finding that schools can exert an effect similar in size to family income contrasts sharply with research on education that focuses on specific characteristics such as expenditure, class size, or teacher education. Beginning with the Coleman Report in 1966 (Coleman et al. (1966)), the vast bulk of research has been interpreted as saying that families exert a much more important effect than schools, raising doubts about the potential for schools to serve as important mechanisms for equalizing opportunities and raising children out of poverty. The results for Texas should not be interpreted as evidence against the importance of families and extra-school influences. On the other hand, they do support the notion that schools exert a very important influence on academic development, an influence much more similar in magnitude to the popular perception of the importance of schools. Importantly, this conclusion emerges only following the relaxation of the view of schools as monolothic institutions whose quality is determined by the salaries paid to teachers, the gleam of the laboratory, the size of classrooms, and the availability of the latest computers. While these factors do exert both a direct effect on students and an indirect effect via making the school more attractive to prospective teachers, it is only consideration of the substantial variation in teacher quality within school buildings that

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

leads to the finding that the quality of school instruction is a primary determinant of academic achievement.

Traditional Policy Alternatives

Several contrasting sets of policy proposals have been discussed and debated. We begin with an explicit discussion of the two most popular approaches. The most sweeping involves an across-the-board salary increase designed to attract new people into the teaching profession. The alternative involves tightening the requirements for entry into teaching in an effort to ensure a more skilled teaching force. Following the discussion of these, we consider other alternatives that come from the estimation results above.

A. Salary increases and general labor market issues

We begin with overall labor market issues, because any policy initiatives will be applied within this context. Much, but not all, of the policy reform debate about teacher quality concentrates on the school demand side and attempts to regulate who can enter and continue in teachers. Part of this considers what schools must do in hiring and retaining people to be teachers. Another part relates to what kinds of screening and preparation goes on at schools of education. Less attention is given to overall labor market issues and the supply decisions of individuals who are choosing among alternative occupations.

The desire to enter into and remain in teaching depends upon a number of factors. Because most Americans including most women expect to spend a majority of their adult lives in the labor market, it is informative to consider the desirability of teaching relative to other occupations. Perhaps the most dramatic change in the U.S. labor market over the past 50 years has been the expansion of labor force participation among women, particularly into historically male professions. As the male/female wage gap has fallen and opportunities for women have opened, schools have found it more and more difficult to attract and retain the highly qualified women who formed the bulwark of the teaching profession for many years (see, for example, the discussion in Flyer and Rosen (1997)). In a very real sense the labor market progress of women has imposed a substantial cost on schools.

A second pronounced shift in the U.S. labor market during the past 40 years has been the substantial increase in the demand for and wages paid to highly educated workers. These workers, particularly those with technical backgrounds in science or math, have enjoyed dramatic salary increases both in absolute terms and relative to less educated workers. Importantly, this same trend that has raised the value of schooling has also made it more costly for schools to maintain a given quality of teachers (see Hanushek and Rivkin (1997) and Lakdawalla (2001)). More resources must be devoted to schools if they are to compete effectively with other industries that are also trying to hire skilled workers.

A summary of the trends in teacher labor markets for men and women is found in Figure 1. To trace teacher quality changes, we use annual earnings data for teachers taken from the six decennial Censuses of Population between 1940 and 1990.[16] Annual earnings, which includes money teachers receive in teaching and in other occupations, obviously goes beyond comparing pure teacher salaries to salaries in other occupations. We believe that, while more common, using just teaching salaries concentrates on the wrong comparison, because teachers enjoy much longer vacations than most other workers. Overall earnings better reflect the monetary benefits of being a teacher as opposed to having a different primary occupation.[17] Private school and public school teachers are also grouped together, but since a roughly

[16]The category is defined as all teachers below the post-secondary level. As such, it includes some preschool teachers, but these people cannot be separated out before 1980. In 1980, 97 percent of all included teachers were primary and secondary teachers.

[17]Broad occupations clearly differ in a variety of nonmonetary ways including fringe benefits and average length of workday and work year. This analysis assumes that the relative importance of these over time has remained constant between teaching and other occupations. Rothstein and Miles (1995) suggest that benefits for teachers have risen faster than those for the rest of the economy between 1967 and 1991, although, as they point out, such comparisons are difficult to make with a high degree of reliability.

constant 10 percent of students attend private schools throughout the period, it is unlikely that movement in the earnings of private school teachers will have a significant impact on the overall relative wages of teachers.

Teacher earnings are compared to the earnings of those who do not teach. Specifically, the location of average teacher earnings in the distribution of nonteacher earnings is our primary measure of potential teacher quality. The lower is the percentage of nonteachers who earn less than the average teacher, the relatively worse teaching jobs are when compared to alternative occupations. The use of percentile rankings as opposed to a comparison of mean earnings reduces problems associated with the Census top-coding of incomes and lessens the impact of changes in the tails of the nonteacher earnings distribution.

The movements in relative earnings of teachers have been dramatic, but, as shown in Figure 1, differ noticeably between men and women. While the average male teacher earned more than 84 percent of all males in 1940, this fell to 64 percent by 1990. All of this relative fall, however, occurred before 1960, and, following a slight dip in the 1970s, male teachers have been moving up the earnings distribution. The overall decline in the relative position of women teachers has been almost as large, though female teachers are still better positioned in the earnings distribution than male teachers. But the time path of the decline for females has been very different, with the largest falls occurring after 1970, when the average teacher moved a full 10 percentage points down the earnings distribution.

Whether the decline in relative teacher earnings has been good educational policy and should be continued depends upon how much teacher performance and educational output would have been improved by the payment of higher salaries. While it may be natural to expect that a decline in teacher quality follows directly any decline in relative teacher earnings, the true effect depends on a number of unmeasured factors including the substitutability between teaching skill and other activities and the ability of schools to hire and retain effective teachers. On this question the empirical evidence is mixed. Murnane et al. (1991) find that higher salaries tend to retain higher IQ teachers. On the other hand, Ballou (1996) finds little evidence that public schools hire the most skilled teacher applicants, and Hanushek and Pace (1995) find little salary responsiveness of college students training to be teachers.[18] Nevertheless, there seems to be little doubt that a large fall in relative teacher earnings over a long period of time will eventually exert a depressing effect on school quality.

One of the most profound changes in schools over the past few decades may be the changing female labor market, as women, the mainstay of the teaching force, have been bid away by other professions.[19] Moreover, if allowed to continue, the dramatic changes at entry level—where the relative position of male and female teachers is now very similar—suggests even larger changes. The prospects for large numbers of retirements over the next decade coupled with the now present salary disadvantages of teaching could lead to large swings in the quality of the teaching force over a short period of time.

Thus two of the dominant labor market trends over the past 40 years have made it more difficult and costly to hire high quality teachers by reducing the return to teaching relative to engaging in other occupations. The changes have been even more pronounced for Black and Hispanic college graduates who have also experienced expanded opportunities following the passage of civil rights legislation and other changes in social norms. Therefore attempts to hire teachers of color face even greater obstacles in the form of better opportunities elsewhere. Importantly, even if teachers tend to place a higher value on non-pecuniary rewards, the labor

---

[18]The educational production function studies on the relationship between student outcomes and either teacher salaries or the determinants of teacher salaries also find no clear evidence that higher salaries improve outcomes (Hanushek (1986, (1997)) These studies, however, indicate the impact of movements along the salary schedule rather than the implications of shifts in the entire schedule.

[19]The trends in teacher salaries and labor market competition for the most able women may in fact be major components of achievement score declines in the 1960s and 1970s, although explicit evaluation of this relationship is not possible. See Congressional Budget Office (1986, (1987) for more discussion of achievement trends and explanations.

market shifts still affect the number and quality of those willing to teach unless they place no value whatsoever on salary. [3]

The evidence on the decline in relative teacher salaries is frequently used to argue that the only answer to currently perceived school quality is to raise salaries of teachers to be competitive. These arguments simply replay the previous discussion of teacher wages and call for a return to the position of the past. Three aspects of this argument are important to consider. First, total spending rose steadily over this period, averaging 3.5 percent annually in real spending per student over the entire 20th century.[20] Second, it is clear that schools traded-off teacher salaries for other elements of program. As described in Hanushek and Rivkin (1997), spending other than for instructional expenses rose dramatically over the period that teacher salaries were allowed to drift down. Instructional salary expenditures were 57 percent of total current expenditures in 1970 but were just 46 percent in 1990. Third, pupil-teacher ratios fell dramatically over the period, accounting for XX percent growth in instructional spending over the same 1970-90 period. In other words, there appeared to be a substitution of quantity for quality (cf. Lakdawalla (2001)).

The issue of teacher salaries in this context then becomes a question of what else changes. If the argument is for increasing salaries within the structure that has evolved, it would be extraordinarily expensive. An alternative perspective would consider the tradeoffs between teacher salaries and other expenditures including smaller classes. In any case, the restoration of teacher salaries to their same position as in the overall 1970 distribution of college educated workers would necessitate a substantial increase in expenditures. The benefits

---

[3]Discussion of the differential racial response of potential teachers to alternative occupations can be found in Murnane et al. (1991) and Hanushek and Pace (1995).

[20] These calculations deflate spending by the consumer price index. This choice has been controversial, but the discussion has avoided the issues raised above about the possible changes in the costs of college educated labor. See Rothstein and Miles (1995); Hanushek (1997).

of such a salary policy presumably center on the ability to attract higher quality teachers, a recognition that the policies of the past decades have led to poorer than desirable teachers. Yet these benefits will be slow in arriving if the new teacher salaries are applied to all existing teachers. Not only will the increased salaries confer windfall gains on existing teachers but also they will act to slow the turnover of teachers and to limit the number of new teachers that can be hired for some time to come. Below we consider alternative compensation policies that focus any additional resources on new teachers.

B. Raising the requirements for teachers

The educational requirements needed to join accounting and consulting firms, investment banks, law firms, medical practices, private schools and firms in many other occupations have remained largely unchanged over the past 40 years. In contrast, most public schools make it much more difficult to become a teacher now than it was twenty or thirty years ago. At least partly in response to the decline in teacher quality, the movement to professionalize teaching has adopted a number of more stringent requirements for prospective teachers. These include master's degrees, the passage of more stringent certification tests, and more rigorous professional development to name a few. While the desires to weed out poor teachers and to raise the status of teaching relative to other occupations are laudable, the effects on the costs of becoming a teacher and ultimately on the supply of teachers may well offset any benefits. As others have pointed out (e.g., Murnane et al., 1991), if teacher preparation requirements have little payoff outside of teaching but simultaneously reduce the time that can be spent on more general preparation, fewer college students will be inclined to enter into teacher training.

The simple argument is that teacher policy discussions must consider how the overall labor market affects the supply of potential teachers. Surely there are many people who are dedicated to teaching and who feel a calling, and these people will tend to be less affected by outside pressures. But, current concerns about the shortages in specific areas and about quality suggest that there are problems of overall supply. Reforms that make entry into teaching more costly or more difficult are likely to exacerbate such shortages.

With this background, we can now examine some of the more popular policies being discussed (National Commission on Teaching and America's Future (1996), Darling-Hammond (1997)). The impact of these can be seen, for example, in the plans of New York State educators who basically have moved to adopt the National Commission's policies. We focus on two specific areas: those that affect the cost of becoming a teacher and those that affect pay and working conditions.[5] Throughout the discussion we spend little time on the justification for any specific policy, because we assume that those are well known.

C. Affecting the cost of becoming a teacher

(1) *New teachers must complete a master's degree by the third year of teaching*. Past research leads to the conclusion that there is no significant difference in teacher quality between teachers with and without master's degrees. This likely results from that fact that in general teachers treat post-graduate education as a requirement that must be satisfied, and the reward for obtaining a useful post-graduate education pails in comparison to the higher pay and job possibilities following the completion of the degree. On the other hand, by making it more difficult and costly to become a teacher, this requirement will likely reduce supply, particularly of highly skilled college graduates who desire to teach for a short period prior to entering a more permanent occupation. Consequently the more stringent Masters degree requirement may well reduce average quality of new teachers while at the same time raising salaries due to the

[5]These policies are described in "New York's Committment: Teaching to Higher Standards," by the New York State Board of Regents and the New York State Education Department

reduction in supply. The elimination of the transitional certificate entirely would only exacerbate the problems introduced by the more stringent requirements.

(2) *New teachers must pass more difficult certification examinations*. The requirement of mastery of a specific set of skills makes perfect sense. However, teaching requires multi-dimensional skills, and some who may perform poorly on standardized tests may offset that deficiency with other attributes. Past research (e.g., Murnane et al. (1991) also suggests that minority students are much more sensitive to such testing requirements than are white students. By eliminating such teachers from consideration, the state reduces the supply of effective teachers, which may impose particularly hardship on schools serving disadvantaged students who tend to have the most difficulties in attracting and retaining teachers. Such tests may also have the unintended consequence of reducing the diversity of the teaching force. Of course these tests also weed out many poor teachers. Thus the key questions are whether the tests eliminate more bad than competent teachers, what is the appropriate pass threshold, and whether using the tests simply as additional information for local administrators would be a better policy. Of course the answers depend in large part upon the quality of hiring practices in the local jurisdictions.

(3) *Grant scholarships to students in exchange for the obligation to teach a minimum number of years*. The effect of the scholarship on the quantity and quality of those who wish to enter teaching is an empirical question. However, the desirability and ability to enforce such contracts is questionable. Forcing someone who dislikes teaching to complete her obligation is unlikely to yield positive results in the classroom. While such scholarships enable the government to raise the earnings of new teachers without affecting the salaries of current teachers, the costs may not justify the benefits. Alternatives such as two tier contracts that

would pay new entrants (and existing teachers who wish to enter the program) higher salaries but also impose greater accountability may prove more effective at raising school quality. Such programs are more difficult to organize at the state or federal level, however. Another alternative is the provision of in kind benefits such as housing assistance to untenured teachers, although economists generally believe that providing an equivalent amount of cash is generally superior.

(4) *Provide money to improve safety and other working conditions*. These programs depend importantly on whether or not they improve things that teachers value. Such expenditures may be a cost effective way to raise teacher quality, but it requires much more information than currently available to evaluate this. More direct efforts to curb disruptive and dangerous behavior of students may also prove valuable. This includes the ability of teachers to remove disruptive students from classes and the establishment of a clear set of incentives for students.

(5) *Require performance based staff reviews including portfolio evaluations to receive a professional certificate*. The more comprehensive, rigorous, and performance based is the review process the better. There is nonetheless some confusion and difference of opinion on what is to be evaluated. Similar to the overall differences in policy approaches discussed above, performance evaluation can be done on the basis of preparation for teaching or on the basis of student performance. Clearly evaluation based on student performance is much more likely to lead to student performance improvements than the former.

An alternative approach to raising the quality of instruction

The primary thrust of most efforts to improve teacher quality is regulation coupled with higher salaries or some other means to increase the economic return to teaching. The essential feature is that none are directly related to the actual performance of teachers in the classroom. The alternative approach, which we summarize here, is to focus much more on student performance while freeing up the supply of potential teachers.[21]

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 well justify substantial differences and flexibility in pay schedules both by subject and working conditions, and more should be learned about the consequences of differentiated pay. However, it is the variation in skill and effort that raises the most difficult set of issues for policy makers, because regulations -- including but not limited to certification requirements – are not likely to get at the crux of the issue.

The simple position taken here is: *if one is concerned about student performance, one should gear policy to student performance*. 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 is trying to do the best they can. It is simply a statement that they are responding to the incentives that are placed in front of them (just like we all do to the incentives we face), and public schools desperately need a different set of incentives.

[21] There have been a variety of experiments with alternative routes to teaching that 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 shows generally positive results (see Raymond, Fletcher, and Luque (2001)).

Specifically, the evidence strongly suggests to us that 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 there is certainly room for improvement in hiring, it will always be an imperfect process. The other aspects of personnel management, on the other hand, including mentoring and support, tenure review, and the management of experienced teachers leaves tremendous room for improvement.

Existing research demonstrates that principals do in fact know who the better teachers are.[6] 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 are administered regularly. Unfortunately, little use is made of any such information in the current system, and we have little experience with the range of possible approaches.[7] As noted above, 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 measures of student performance. 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. Of course such structures are not common in education, so we have little to build upon in the actual structuring of such a notion. Moreover, making such active decisions is

---

[6]See Armor et al. (1976) and Murnane (1975) who identify total teacher effects as discussed above and relate them to principals' evaluations.

[7]Some evidence has accumulated about merit pay plans, and this has not suggested that merit pay as applied to schools has been very effective (Cohen and Murnane (1986). There is reason to believe that these experiments are, however, too limited (Hanushek and others (1994).

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. This is the challenge. 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. We currently do not know the best way to structure incentives.[8] We have not tried many performance incentive systems, so we have very little experience or evidence with them. Nonetheless, the experiments now going on offer some hope for learning.

The introduction of appropriate incentives and gathering of comprehensive performance measures provide administrators with strong motivation to succeed, but they will be unable to do their job without the flexibility required to manage teachers. Even highly motivated principals will not produce high quality schools if they are hamstrung by a plethora of regulations, a tenure system that protects incompetence, and rigid pay schedules. Teachers undoubtedly respond to incentives, and the addition of a Master Teacher designation is certainly a step in the right direction. Be it pay for performance, flexible promotion rules, the ability to impose sanctions, or more likely some combination of the three, incentives are needed to motivate many teachers. Finally, principals and superintendents cannot be expected to work magic, particularly in economically disadvantaged communities. If the working conditions make it difficult to attract teachers, additional resources may be needed. However, such resources should be contingent upon the enactment of substantial changes in school management.

[8]While describing and evaluating alternative incentive schemes is beyond the scope of this paper, see Hanushek and others (1994).

Table 1. Estimated Effects of Teacher Characteristics on 4th-6th Grade Math Test Score Gain

(absolute value of Huber-White adjusted t statistics in parentheses)

	Achievement level (6^{th} grade)	Achievement gains			
		w/o fixed effects	with student and school fixed effects		
	All students	All students	All students	disadvantaged	not disadvantaged
class size	-0.0026	-0.0036	-0.0063	-0.0048	-0.0078
	(1.68)	(5.31)	(4.44)	(4.23)	(3.25)
% 0 yrs experience	-0.21	-0.11	-0.10	-0.11	-0.09
	(6.12)	(6.14)	(3.28)	(3.33)	(2.24)
% 1 year experience	-0.09	-0.06	-0.08	-0.08	-0.08
	(2.29)	(3.29)	(2.70)	(2.34)	(2.10)
% 2 yrs experience	-0.06	-0.04	-0.07	-0.07	-0.06
	(1.59)	(1.96)	(2.31)	(2.22)	(1.59)
% 3 or 4 yrs experience	-0.05	-0.01	-0.03	-0.05	0.01
	(1.69)	(1.04)	(1.23)	(2.12)	(0.27)
% 5 to 9 yrs experience	0.03	0.00	0.01	-0.01	0.05
	(1.07)	(0.51)	(0.72)	(0.63)	(2.06)
% 20 to 30 yrs experience	-0.02	0.00	0.00	-0.01	0.02
	(0.58)	(0.12)	(0.01)	(0.48)	(0.59)
% 30+ yrs experience	-0.07	-0.01	0.01	-0.02	0.04
	(1.24)	(0.47)	(0.24)	(0.55)	(0.63)
% with graduate degree	-0.06	-0.02	-0.01	-0.01	-0.01
	(3.03)	(2.07)	(0.68)	(0.74)	(0.46)
Test Scores					
Elementary	0.037	0.000	-0.015	0.002	-0.035
Comprehensive	(2.29)	(0.01)	(1.36)	(0.21)	(2.33)
Educational	-0.004	0.011	0.008	0.002	0.017
Development	(0.30)	(2.08)	(0.76)	(0.18)	(1.20)
Mathematics	0.000	-0.005	0.005	0.002	0.009
	(0.04)	(1.14)	(0.59)	(0.87)	(0.24)
observations	303,406	906,206	906,206	517,274	388,932

Note: Regressions also include indicator variables for economically disadvantaged, school transfer, year by grade and for schools in which no teachers in a specific grade and year had taken a particular test. Models without fixed effects include indicator for race and economic disadvantage.

Table 2. Lower bound estimates of the standard deviation of teacher quality from alternative specifications of student gains in mathematics achievement

No fixed effects	Individual and school fixed effects	Individual and school-by-grade fixed effects	No fixed effects	Individual and school fixed effects	Individual and school-by-grade fixed effects
Without controls for administrative changes			With controls for administrative changes		
0.11**	0.13*	0.11**	0.11**	0.13*	0.11**

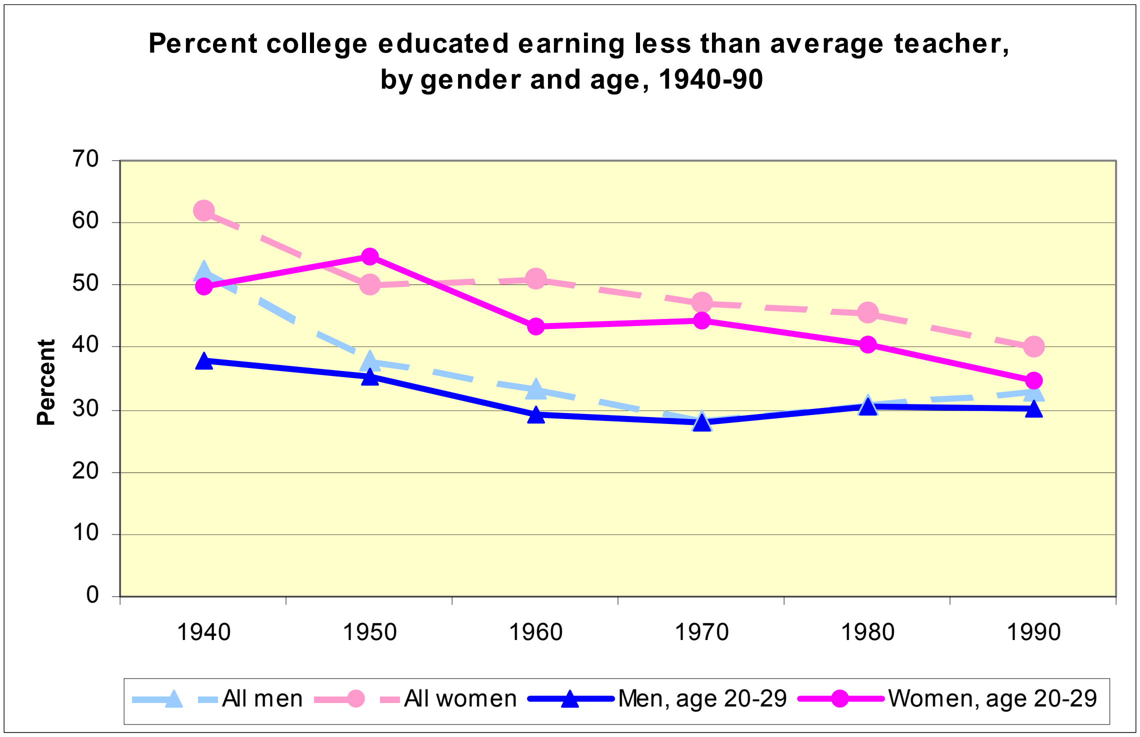
Statistical significance: * $p < .001$; ** $p < .01$.

Notes: Estimation of variance in teacher quality under different specifications is described in Rivkin, Hanushek, and Kain (2001). Estimation is based on the proportion of math teachers who are new each year. Administrative controls identify new principals and teachers in each school and year.

Appendix Table A1. Descriptive statistics on availability of teacher test scores

	% with no teachers who took examination	average % whose teachers took examination | >0
Comprehensive	15.5	23.3
Mathematics	49.4	13.4
Educational Development	19.3	21.6

Figure 1. Teacher Salaries Relative to Salaries for Nonteaching Workers with Bachelor's Degree or More



References

Armor, David, Patricia Conry-Oseguera, Millicent Cox, Niceima King, Lorraine McDonnell, Anthony Pascal, Edward Pauly, and Gail Zellman. 1976. *Analysis of the school preferred reading program in selected Los Angeles minority schools*. Santa Monica, CA: Rand Corp.

Ballou, Dale. 1996. "Do public schools hire the best applicants?" *Quarterly Journal of Economics* 111,no.1 (February):97-133.

Boozer, Michael, and Cecilia Rouse. 1995. "Intraschool variation in class size: Patterns and implications." NBER Working Paper 5144, June 1995

Cohen, David K., and Richard J. Murnane. 1986. "Merit pay and the evaluation problem: Understanding why most merit pay plans fail and a few survive." *Harvard Educational Review* 56,no.1 (February):1-17.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.

Congressional Budget Office. 1986. *Trends in educational achievement*. Washington, D.C.: Congressional Budget Office.

———. 1987. *Educational achievement: Explanations and implications of recent trends*. Washington, DC: Congressional Budget Office.

Darling-Hammond, Linda. 1997. *Doing what matters most: Investing in quality teaching*. New York: National Commission on Teaching and America's Future.

Finn, Chester E. Jr, and Marci Kanstoroom, eds. 1999. *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation.

Finn, Chester E. Jr, Marci Kanstoroom, and Michael J Petrilli, eds. 1999. *The quest for better teachers: Grading the states*. Washington, DC: Thomas B. Fordham Foundation.

Flyer, Fredrick, and Sherwin Rosen. 1997. "The new economics of teachers and education." *Journal of Labor Economics* 15,no.1 (part 2) (January):104-139.

Hanushek, Eric A. 1971. "Teacher characteristics and gains in student achievement: Estimation using micro data." *American Economic Review* 60,no.2 (May):280-288.

———. 1979. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources* 14,no.3 (Summer):351-388.

———. 1986. "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature* 24,no.3 (September):1141-1177.

———. 1992. "The trade-off between child quantity and quality." *Journal of Political Economy* 100,no.1 (February):84-117.

———. 1997. "Assessing the effects of school resources on student performance: An update." *Educational Evaluation and Policy Analysis* 19,no.2 (Summer):141-164.

———. 1997. "The productivity collapse in schools." In *Developments in School Finance, 1996*, edited by W. J. Fowler, Jr. Washington, DC: National Center for Education Statistics.

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. forthcoming. "Inferring program effects for specialized populations: Does special education raise achievement for students with disabilities?" *Review of Economics and Statistics*.

Hanushek, Eric A., and with others. 1994. *Making schools work: Improving performance and controlling costs*. Washington, DC: Brookings Institution.

Hanushek, Eric A., and Richard R. Pace. 1995. "Who chooses to teach (and why)?" *Economics of Education Review* 14,no.2 (June):101-117.

Hanushek, Eric A., and Steven G. Rivkin. 1997. "Understanding the twentieth-century growth in U.S. school spending." *Journal of Human Resources* 32,no.1 (Winter):35-68.

Krueger, Alan B. 1999. "Experimental estimates of education production functions." *Quarterly Journal of Economics* 114,no.2 (May):497-532.

Lakdawalla, Darius. 2001. "The declining quality of teachers." Working Paper 8263, National Bureau of Economic Research, Cambridge, MA

Murnane, Richard J. 1975. *Impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger.

Murnane, Richard J., and Barbara Phillips. 1981. "What do effective teachers of inner-city children have in common?" *Social Science Research* 10,no.1 (March):83-100.

Murnane, Richard J., Judith D. Singer, John B. Willett, James J. Kemple, and Randall J. Olsen. 1991. *Who will teach?* Cambridge: Harvard University Press.

National Commission on Teaching and America's Future. 1996. "What matters most: Teaching for America's future. New York: NCTAF.

Raymond, Margaret E., Stephen Fletcher, and Javier A. Luque. 2001. *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford University: CREDO.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2001. "Teachers, schools, and academic achievement." Working Paper No. 6691, National Bureau of Economic Research (revised)

Rothstein, Richard, and Karen Hawley Miles. 1995. *Where's the money gone? Changes in the level and composition of education spending*. Washington, DC: Economic Policy Institute.

Summers, Anita, and Barbara Wolfe. 1977. "Do schools make a difference?" *American Economic Review* 67,no.4 (September):639-652.