TEACHERS, SCHOOLS, AND ACADEMIC ACHIEVEMENT

By Steven G. Rivkin, Eric A. Hanushek, and John F. Kain[1]

This paper disentangles the impact of schools and teachers in influencing achievement with special attention given to the potential problems of omitted or mismeasured variables and of student and school selection. Unique matched panel data from the UTD Texas Schools Project permit the identification of teacher quality based on student performance along with the impact of specific, measured components of teachers and schools. Semiparametric lower bound estimates of the variance in teacher quality based entirely on within-school heterogeneity indicate that teachers have powerful effects on reading and mathematics achievement, though little of the variation in teacher quality is explained by observable characteristics such as education or experience. The results suggest that the effects of a costly ten student reduction in class size are smaller than the benefit of moving one standard deviation up the teacher quality distribution, highlighting the importance of teacher effectiveness in the determination of school quality.

KEYWORDS: Student achievement, teacher quality, school selection, class size, teacher experience.

1. INTRODUCTION

SINCE THE RELEASE of *Equality of Educational Opportunity* (the "Coleman Report") in 1966, the educational policy debate in the United States and elsewhere has often been reduced to a series of simplistic arguments and assertions about the role of schools in producing achievement.[2] The character of this debate has itself been heavily influenced by confusing and conflicting research. While this research has frequently suffered from inadequate data, imprecise formulation of the underlying problems and issues has been as important in obscuring the fundamental policy choices. This paper defines a series of basic issues about the performance of schools that are relevant for current policy debates and considers how observed student performance can be used to address

[2]The original Coleman Report (Coleman et al. (1966)) was subjected to considerable criticism both for methodology and interpretation; see, for example, Hanushek and Kain (1972). The ensuing controversy led to considerable new research, but this new work has not ended the controversy; see Hanushek (1996, 2003) and Greenwald, Hedges, and Laine (1996). Those discussions represent the starting point for this research.

each. It then employs a unique panel data set of students in Texas to identify the sources of differences in student achievement and the relevance of a broad class of policies related to school resources.

Some very basic questions that have arisen from prior work command a central position in most policy discussions. First, partly resulting from common misinterpretations of the Coleman Report, do schools "make a difference" or not? While a surprising amount of controversy continues over this issue, it comes down to a simple question of whether or not there are significant and systematic differences between schools and teachers in their abilities to raise achievement. Second, how important are any differences in teacher quality in the determination of student outcomes? Finally, are any quality differences captured by observable characteristics of teachers and schools including class size, teacher education, and teacher experience? If so, how large are the effects? This third issue is in fact the genesis of the first, because the Coleman Report reported relatively small effects of differences in the measured attributes of schools on student achievement—a finding that has frequently been interpreted as indicating that there are no systematic quality differences among schools.

An extraordinarily rich data set providing longitudinal information on individual achievement of students in the State of Texas permits analyses that yield quite precise answers to each of these questions. The data contain test scores spanning grades 3 through 7 for three cohorts of students in the mid-1990s. The multiple cohorts and grades coming from repeated observations on more than one-half million students in over three thousand schools permit the clear identification and detection of even very small teacher and school effects.

A primary objective of the initial empirical analysis is to obtain estimates of differences in teacher contributions to student learning that eliminate the major sources of possible contamination from student selection or teacher assignment practices. Because family choice of neighborhood and school depends on preferences and resources, students are nonrandomly distributed across schools (Tiebout (1956)). Schools also use student characteristics including assessments of ability and achievement to place students into specific programs and classes. Such nonrandom selection may easily contaminate estimates of school or teacher effects with the influences of unmeasured individual, family, school, and neighborhood factors.

Repeated performance observations for individual students and multiple cohorts provide a means of controlling explicitly for student heterogeneity and the nonrandom matching of students, teachers, and schools through the use of fixed effects models. The models control for fixed student, school-by-grade, and in some cases school-by-year effects and then relate remaining differences in achievement gains between grades and cohorts to differences in school characteristics or teachers. This variation in academic performance cannot be driven by unchanging student attributes such as ability or motivation or by unchanging school characteristics and policies that are either common across all

grades at a point in time or unique to specific grades. Moreover, the empirical models also account for potentially important time varying influences not captured by the student or school fixed effects. Therefore we are able to identify the impacts of schools and teachers uncontaminated by the many unobserved family and other influences that have plagued past research.

The results reveal large differences among teachers in their impacts on achievement and show that high quality instruction throughout primary school could substantially offset disadvantages associated with low socioeconomic background. These differences among teachers are not, however, readily measured by simple characteristics of the teachers and classrooms. Consistent with prior findings, there is no evidence that a master's degree raises teacher effectiveness. In addition, experience is not significantly related to achievement following the initial years in the profession. These findings explain much of the contradiction between the perceived role of teachers as the key determinant of school quality and the body of research showing that observed teacher characteristics including experience and education explain little of the variation in student achievement.

Students also appear to benefit from smaller classes, particularly in grades 4 and 5. In comparison to the gains from higher teacher quality, however, the estimates indicate that even a very costly ten student reduction in class size such as that undertaken in some U.S. states produces smaller benefits than a one standard deviation improvement in teacher quality.

The next section provides an overview of patterns of achievement gains that suggests the presence of substantial within school variation in teacher quality. Section 3 describes the empirical approach used to generate a lower bound estimate of the within school variation in teacher quality. Section 4 provides a detailed description of the Texas data on students and teachers. Section 5 reports estimates of the variance in teacher quality based on the method developed in Section 3, and Section 6 presents an extension of traditional analyses of the effects of measured resources: class size, teacher education, and teacher experience on achievement. The final section considers the policy implications of the findings, particularly the importance of measured resources relative to the overall contribution of teachers.

2. SCHOOLS AND TEACHERS

Students and parents refer often to differences in teacher quality and act to ensure placement in classes with specific teachers. Such emphasis on teachers is largely at odds with empirical research into teacher quality. 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 academic achievement and future academic and labor market success. It may be that parents and students overstate the importance of teachers, but an alternative explanation is that measurable

characteristics such as teacher experience, education, and even test scores of teachers explain little of the true variation in quality.

To motivate the concentration on teacher quality, we begin with aggregate statistics on the variation in student achievement. Table I displays correlations of school average annual mathematics and reading achievement gains in grades 5, 6, and 7 between two cohorts of students for all public elementary schools in Texas.[3] The diagonal elements report correlations for the same grade in adjacent years, while the off-diagonal elements report correlations for adjacent grades in the same year.

The striking difference in magnitudes of the diagonal and off-diagonal elements suggests the existence of substantial within-school heterogeneity in school quality. Remarkably, the correlation of between-cohort average gains in different grades in the same year (the off-diagonal terms) is quite small despite the homogeneity of family backgrounds and peers within most schools and despite the common school organization, leadership, and resources for the two cohorts. Indeed for comparisons of 6th and 7th grade reading performance, the correlation is -0.01. In contrast, the correlations of between-cohort average gains in the same grade in adjacent years (the diagonal terms) are much larger. A number of factors may explain this pattern, but perhaps the most obvious explanation is that there will be many common teachers for two cohorts when observed in the same grade, while virtually all of the teachers will be different when comparing cohort performance across grades at a single point in time.

Table II reports the R^2 values from a series of achievement gain regressions for reading and mathematics performance run over the sample of schools and grades in which there is only a single teacher per subject. (As we discuss be-

TABLE I

PEARSON CORRELATION COEFFICIENTS OF SCHOOL AVERAGE TEST SCORE GAINS
IN MATHEMATICS AND READING ACROSS GRADES AND YEARS

Grade of Cohort I	Mathematics Grade of Cohort II			Reading Grade of Cohort II		
	5	6	7	5	6	7
5	0.32**			0.19**		
6	0.13**	0.52**		0.13**	0.43**	
7		0.05	0.46**		−0.01	0.44**

Notes: Cohort I attended 4th grade in 1994; Cohort II attended 4th grade in 1995. Thus, for example, Cohort I is attending the 6th grade during the same academic year that Cohort II is attending the 5th grade. All calculations are weighted by the average enrollment of the pairs.

*significant at 10% level; **significant at 1% level.

[3]These data, subsequently used in the detailed empirical analyses, are described in detail in Section 3. All correlations relate just to students in schools that have both of the relevant grades.

TABLE II

COMPARISON OF THE EXPLANATORY POWER OF TEACHER EXPERIENCE, EDUCATION, AND
CLASS SIZE WITH TEACHER FIXED EFFECTS IN EXPLAINING ACHIEVEMENT GAINS

	Mathematics				Reading			
Included explanatory variables								
Student covariates	yes	yes	yes	yes	yes	yes	yes	yes
Teacher characteristics	no	yes	no	no	no	yes	no	no
Teacher fixed effects	no	no	yes	no	no	no	yes	no
School fixed effects	no	no	no	yes	no	no	no	yes
R squared	0.0151	0.0182	0.1640	0.0949	0.0085	0.0093	0.0903	0.0507
Observations		89,414				81,897		

Notes: Dependent variables are mathematics and reading test score gains; sample includes only grades in a school with a single teacher for that subject.

low, these are the only schools in which students can be matched to their actual teachers.) The first column for each subject is based on a specification with only student characteristics and year dummies; the second column adds measured teacher and classroom characteristics (teacher experience, teacher education, and class size); the third column substitutes teacher fixed effects for the observable teacher and classroom characteristics; and the final column employs school rather than teacher fixed effects. The results demonstrate quite clearly that the observable school and teacher characteristics explain little of the between-classroom variation in achievement growth despite the fact that a substantial share of the overall achievement gain variation occurs between teachers. Importantly, even though the sample includes just schools with a single teacher per grade, the inclusion of school rather than teacher fixed effects reduces the explanatory power by over forty percent, suggesting that much of the variation in teacher quality exists within rather than between schools.

Tables I and II are consistent with the existence of substantial variation in teacher quality not explained by observable teacher characteristics. However, other factors could clearly enter into these two simple comparisons, making it necessary to utilize more comprehensive methods to identify the variance of teacher quality and importance of observable factors. For example, a high performing 4th grade teacher could leave less room for subsequent gains; the curriculum could affect specific grade levels in differing ways across school districts; test measurement errors could obscure the relationships; there may be nonrandom sorting across schools; or some schools may have more or less effective leadership. The next section develops a comprehensive model of student learning that provides the analytical framework for the estimation of the variance of teacher quality.

3. THE IDENTIFICATION OF TEACHER EFFECTS

In this section we develop an estimator of the variance of teacher quality that avoids problems of student selection and administrator discretion that potentially have biased prior attempts. This estimator is based upon patterns of within-school differences in achievement gains and ignores variations in teacher quality across schools, because such variation cannot readily be disentangled from student differences and the contributions of other school factors. This strategy yields a lower bound estimator for the importance of teacher quality that relies upon minimal maintained assumptions about the underlying achievement process. Importantly, we do not focus solely on measurable characteristics of teachers or schools as is typically done in this literature but instead rely on student outcomes to assess the magnitude of total teacher effects, regardless of our ability to identify and measure any specific components. This semiparametric approach provides both an estimate of the role of teacher quality in the determination of academic achievement and information on the degree to which specific factors often used in determining compensation and hiring explain differences in teacher effectiveness.

3.1. *Basic Model of Student Achievement*

Academic achievement at any point is a cumulative function of current and prior family, community, and school experiences. A study of the entire process would require complete family, community, and school histories, and such data are rarely if ever available. Indeed, the precise specification of what to measure is poorly understood. In the absence of such information, analyses that study the contemporaneous relationship between the level of achievement and school inputs for a single grade are obviously susceptible to omitted variables biases from a number of sources.

An alternative approach focuses on the determinants of the *rate* of learning over specific time periods. The advantage of the growth formulation is that it eliminates a variety of 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. While such a value-added framework by no means eliminates the potential for specification bias, the inclusion of initial achievement as a means to account for past inputs reduces dramatically the likelihood that omitted historical factors introduce significant bias.[4]

Equation (1) presents a conventional value-added model that describes the gain in student achievement (ΔA_{ijgs}^c) for individual i in cohort c with teacher j

---

[4]One restriction of this formulation 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

in grade g of school s:

$$
\begin{aligned}
(1) \qquad \Delta A_{ijgs}^{c} &= A_{ijgs}^{c} - A_{ij'g-1s'}^{c} \\
&= X_{ig}^{c}\beta_X + T_{jgs}^{c}\beta_T + S_{gs}^{c}\beta_S + f_i + \varepsilon_{ijgs}^{c}.
\end{aligned}
$$

This gain, measured as the difference between a student's test scores in grades g and $g-1$, depends on family background (X); teacher characteristics (T); school characteristics (S); inherent student abilities (f); and a random error (ε). Note that the term "inherent abilities" refers to the set of cognitive skills, motivation, and personality traits that affect the rate of achievement growth but that do not change during the school years being considered.[5] Each of the inputs can be thought of as a vector of underlying components.

Formulations similar to equation (1) have been estimated in a variety of circumstances in order to identify the causal link between a student outcome such as achievement or years of schooling on the one hand and a school characteristic such as class size on the other (see, e.g., Murnane (1975) or Summers and Wolfe (1977)). Much research has focused on the development of methods to eliminate any remaining biases, and we address this concern as well. However, a potentially much more important issue is the possibility that the measured teacher and school factors do not adequately capture important differences in the quality of education.

An alternative approach attempts to circumvent the problem of inadequate measures of quality through the estimation of classroom fixed effects on achievement gains (see, e.g., Hanushek (1971), Armor et al. (1976), Murnane and Phillips (1981)). These analyses of covariance capture all between-classroom differences in achievement gains controlling for any included regressors. The resulting classroom differences in average achievement gain have been interpreted as reflecting teacher quality, since the teacher is the most

---

cumulative achievement production process, it is virtually impossible to isolate any continuing effects of specific school factors.

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. The alternative formulations do place different restrictions on the form of the achievement process. See Hanushek (1979) for a discussion of value-added models. Subsequent analysis, relying on expected expansions of our database, will explore alternative specifications.

[5]The isolation of inherent student abilities does not rely on any presumption about their source (genetic, environmental, or an interaction of these). Any fixed differences that affect the rate of learning will be incorporated in this term. This formulation goes beyond typical discussions that concentrate just on how fixed ability, family, and motivational terms affect the level of achievement at a point in time. Here we explicitly allow for the possibility that *ceteris paribus* some children will acquire knowledge at different rates even after allowing for variations in initially observed achievement. Further, these differences do not have to be unidimensional.

obvious factor differing across classrooms. However, problems from test measurement errors and potential school and classroom selection effects may be even more serious for these types of models than in those that use observable measures, making the interpretation of these as direct estimates of the teacher component problematic.[6]

The central estimation problem results from the processes that match students with teachers, and schools. Not only do families choose neighborhoods and schools, but principals and other administrators assign students to classrooms. Because these decision makers utilize information on students, teachers and schools, information that is often not available to researchers or measured with error, the estimators are quite susceptible to biases from a number of sources. The following section develops an empirical model designed to avoid these problems and to identify the variations in the quality of instruction.

3.2. *An Extended Specification of Education Production*

Rather than attempting to define each variable in the education process, we begin by thinking in terms of the total systematic effect of students, families, teachers, and schools. In this, we depart from the parametric approach of equation (1) that involved measuring a small set of inputs in their natural units and move to a semiparametric approach with inputs measured in achievement, or output, units. Equation (2) describes a decomposition of education production during grade g into a set of fixed and time varying factors:

$$(2) \qquad \Delta A_{ijgs}^c = \gamma_i + \theta_j + \delta_s + \upsilon_{ijgs}^c.$$

Test score gain in grade g is written as an additive function of student (γ), teacher (θ), and school (δ) fixed effects along with a random error (υ) that is a composite of time-varying components. The fixed student component captures the myriad family influences including parental education and permanent income that affect the rate of learning; the fixed school factor incorporates the effects of stable school characteristics including resources, peers, curriculum, etc. Finally, the teacher component captures the average quality of teacher j over time. Of course families, schools, and teachers all change from year to year, and such changes receive considerable attention in the analysis below.

Equation (2) is not intended to be a comprehensive model of the achievement determination process, and moreover we do not attempt to identify each of the separate components. Rather, it provides a framework for the specific models used to study the effects of teacher quality and school resource differences. We have not, for example, distinguished any role for school districts.

---

[6]Hanushek (1992) does provide suggestive evidence that teachers are the primary component by showing that classroom gains for individual teachers tend to be highly correlated across time (for different groups of students).

Many school policies—hiring, curriculum, school structure, etc.—emanate from school districts and will produce common elements in the teacher and school effects specified in equation (2). While the study of district effects is clearly important, particularly in a policy context, our focus on within-school achievement differences to avoid the difficulties associated with the endogeneity of school and district choice precludes identification of separate district effects.[7] Moreover, school fixed effects also capture any systematic differences across districts and communities, so there is no econometric reason to specify separate district or community components in this estimation. We do, however, address district related issues as they are relevant to the identification of teacher quality and school resource effects.[8]

3.3. *Estimator of the Variance of Teacher Quality*

In the semiparametric approach of equation (2), the variance of θ measures the variation in teacher quality in terms of student achievement gains. One could estimate this variance directly using between-classroom differences in average achievement gains. We do not adopt this approach for a number of reasons, not the least of which is the inability to match students to specific teachers. Yet even if students could be matched with teachers and the analysis considered only within-school variation in outcomes, both the intentional placement of students into classrooms on the basis of unobservables and the need to account for the contribution of measurement error to the between-classroom variation would introduce serious impediments to the identification of the variance of teacher quality.[9]

Consequently, we adopt a very different method that makes use of information on teacher turnover and grade average achievement gains to generate a lower bound estimate of the within-school variance in teacher quality. This approach avoids the need to identify and to estimate separately the test error

[7]The role of district environment and policies is a topic that we intend to pursue in the future. That analysis however, requires a different estimation strategy that, importantly, does not permit the precise identification of teacher influences that we pursue here.

[8]The model also imposes the assumption of additive separability in order to simplify the presentation. We explore the possibility that the magnitudes of school resource effects vary by student characteristics, allowing for the most commonly cited type of potential complementarity. In addition, we recognize that the matching of students and teachers likely affects the average rate of learning in a school, and the subsequent inclusion of school and school-by-grade fixed effects captures any differences that are maintained across our observation period.

[9]This discussion can be directly linked to prior estimation of classroom fixed effects, which develop classroom gains after conditioning on measurable characteristics of students or schools. See, for example, Hanushek (1971), Armor et al. (1976), and Murnane and Phillips (1981). In such cases, the interpretation of the individual and school components of equation (3) would relate directly to dimensions not captured by the included characteristics, and the test measurement errors would remain.

variance, and the aggregation to the grade level circumvents any problems resulting from classroom assignment.[10] The cost of this aggregation is the loss of all within grade variation in teacher quality and the inability to trace out the teacher quality distribution.

Equation (3) represents average achievement gain in grade g in school s for cohort c as an additive function of grade average student and teacher fixed effects, a school fixed effect, and the grade average error:

$$(3) \qquad \overline{\Delta A_{gs}^c} = \overline{\gamma_{gs}^c} + \overline{\theta_{gs}^c} + \delta_s + \overline{v_{gs}^c}.$$

With two different cohorts of students (c and c'), we can compare average gains in the same grade:

$$(4) \qquad \overline{\Delta A_{gs}^c} - \overline{\Delta A_{gs}^{c'}} = \left(\overline{\gamma_{gs}^c} - \overline{\gamma_{gs}^{c'}} \right) + \left(\overline{\theta_{gs}^c} - \overline{\theta_{gs}^{c'}} \right) + \left(\overline{v_{gs}^c} - \overline{v_{gs}^{c'}} \right).$$

Notice in equation (4) that all fixed school components from equation (3) drop out because they exert the same effect for both cohorts. These eliminated factors include fixed aspects of peers, school administration, technology, and infrastructure as they affect the *growth* in achievement, even if they are grade specific. They also include systematic (time invariant) sorting of teachers by school or district that comes from a district's salary or general attractiveness along with its standard teacher assignment practices. The difference in cohort average achievement gains is thus a function of the between-cohort differences in teacher quality (θ), in fixed student and family factors (γ), and an average error component that includes not only measurement errors but time varying individual, family, and school factors.

Though we do report estimates of the variance in teacher quality based on simple between-cohort achievement differences for a single grade, cohort average differences in (γ) contaminate estimates of the variance in teacher quality. Consequently, we concentrate on the difference between adjacent cohorts in the *pattern* of average gains in grades g and g'. In order to control fully for student fixed effects, we limit the sample to students who remain in the same school for grades $g-1$ and g:

$$(5) \qquad \left(\overline{\Delta A_{gs}^c} - \overline{\Delta A_{g's}^c} \right) - \left(\overline{\Delta A_{gs}^{c'}} - \overline{\Delta A_{g's}^{c'}} \right)$$
$$= \left[\left(\overline{\theta_{gs}^c} - \overline{\theta_{g's}^c} \right) - \left(\overline{\theta_{gs}^{c'}} - \overline{\theta_{g's}^{c'}} \right) \right] + \left[\left(\overline{v_{gs}^c} - \overline{v_{g's}^c} \right) - \left(\overline{v_{gs}^{c'}} - \overline{v_{g's}^{c'}} \right) \right].$$

---

[10]This estimator assumes that there are not strong complementarities between specific students and teachers, that is, that the effects of teachers is linear and separable as in equation (2). Yet as long as schools maintain similar assignment practices from year to year, as discussed below, even such complementarities will not contaminate the estimates. Additionally, changes in assignment practices will tend to bias estimates of the variance in teacher quality downward, reinforcing our interpretation of the estimator as a lower bound on teacher quality variance.

As equation (5) shows, taking the difference between average gains in grades g and g' eliminates all fixed student and family differences, leaving only cohort-to-cohort differences in the grade average difference in teacher quality and time varying student and school factors (contained in υ) as determinants of the difference in the pattern of achievement gains.

Squaring both sides of equation (5) gives

$$(6) \quad \left[\left(\overline{\Delta A_{gs}^c} - \overline{\Delta A_{g's}^c}\right) - \left(\overline{\Delta A_{gs}^{c'}} - \overline{\Delta A_{g's}^{c'}}\right)\right]^2$$

$$= \overline{\theta_{gs}^c}^2 + \overline{\theta_{g's}^c}^2 + \overline{\theta_{gs}^{c'}}^2 + \overline{\theta_{g's}^{c'}}^2 - 2\left(\overline{\theta_{gs}^c}\,\overline{\theta_{gs}^{c'}} + \overline{\theta_{g's}^c}\,\overline{\theta_{g's}^{c'}}\right)$$

$$+ 2\left[\left(\overline{\theta_{gs}^c}\,\overline{\theta_{g's}^{c'}} - \overline{\theta_{gs}^c}\,\overline{\theta_{g's}^c}\right) + \left(\overline{\theta_{g's}^c}\,\overline{\theta_{gs}^{c'}} - \overline{\theta_{gs}^{c'}}\,\overline{\theta_{g's}^{c'}}\right)\right] + e.$$

The squared difference leads to a natural characterization of the observed achievement differences between cohorts as a series of terms that reflect variances and covariances of the separate teacher effects plus a component e that includes all random error and cross product terms between teacher and other grade specific effects.

We now impose three assumptions that formally characterize the notion that teachers are drawn from common distributions over the restricted time period of our cohort and grade observations: (i) The variance of grade average teacher quality is the same for all cohorts and grades; (ii) the covariance of grade average teacher quality for adjacent cohorts is the same for all grades; and (iii) the covariance of grade average teacher quality for grades g and g' for adjacent cohorts equals the covariance of grade average teacher quality for grades g and g' for each cohort. For ease of exposition, we also make the simplifying assumption that each school has one teacher per grade, but this is relaxed later.

Applying these assumptions and taking the expectation of equation (6) yields

$$(7) \quad E\left[\left(\overline{\Delta A_{gs}^c} - \overline{\Delta A_{g's}^c}\right) - \left(\overline{\Delta A_{gs}^{c'}} - \overline{\Delta A_{g's}^{c'}}\right)\right]^2 = 4\left(\sigma_{\theta_s}^2 - \sigma_{\theta_s^c \theta_s^{c'}}\right) + E(e_s),$$

where $\sigma_{\theta_s}^2$ is the variance of teacher quality in school s and $\sigma_{\theta_s^c \theta_s^{c'}}$ is the covariance of teacher quality across cohorts in a school.

The key to the identification of the magnitude of the within-school variance of teacher quality comes from the first element on the right-hand side—the within-school variance of grade average teacher quality minus the within-school covariance of quality across cohorts. Consider first schools in which the two cohorts have the same teacher in each grade (i.e., the proportion of teachers who are different equals zero). As long as teachers perform equally well in both years, $\sigma_{\theta_s}^2 = \sigma_{\theta_s^c \theta_s^{c'}}$, and teacher quality contributes nothing to student performance *differences* across cohorts.

On the other hand, consider schools in which cohorts c and c' have different teachers in each grade (the proportion of teachers who are different

equals one).[11] In this case the within-school covariance of teacher quality equals zero. Importantly, this is not to say that schools hire randomly, for as we discuss below there can be little doubt that hiring practices and characteristics related to teacher job preferences differ substantially across schools. Rather, it says that the covariance across teachers in the deviation from the mean teacher quality in a school is zero.

Equation (7) provides the basis for estimation of the within-school variance of teacher quality. The left-hand side in most regressions is the squared divergence of the grade pattern in gains across cohorts, which we regress on the proportion of teachers who are different. Ignoring the possible confounding influences of other factors and maintaining the assumption that teacher quality remains unchanged in the absence of turnover, the coefficient on this proportion divided by four will provide a consistent estimate of the within-school variance in teacher quality.[12]

One empirical complication arises because most schools do not have a single teacher for each grade. Rather the number of teachers varies by school, and consequently the coefficient on the turnover variable would not have a straightforward interpretation. Because the achievement gains and the effects of teachers are averaged across the teachers in a grade, we actually have the variation of the mean in each school, and the relationship of turnover to the within-school variance will depend on the number of teachers. For example, in a sample of schools with three teachers per grade, the coefficient on proportion different would provide an estimate of four times one third (i.e., $4\sigma^2_{\theta_s}/3$) of the within-school variation in teacher quality. This also means that fifty percent turnover in schools with three teachers per grade would lead to the same expected squared cohort difference in grade average difference in gains as one hundred percent turnover in schools with six teachers per grade. In order to account for such differences in the number of teachers and place all schools on a common metric, the proportion differ-

---

[11]Note that such differences result from both teacher departures and grade changes. There is an extensive related literature on the determinants of teacher turnover, indicating that salary, working conditions, and alternative wage opportunities do affect the probability of exiting a school (cf. Dolton and van der Klaauw (1995, 1999), Murnane and Olsen (1989), Stinebrickner (2002), Hanushek, Kain, and Rivkin (2004b)). None, however, suggests that leavers are systematically more effective teachers than stayers, an issue to which we return below. Moreover, our analysis of within-school patterns of student performance implicitly controls for the overall determinants of turnover and focuses solely on the implications of turnover for performance. Regardless of any differences between leavers and stayers, the within-school covariance of grade average quality equals zero in 100 percent turnover schools as long as any changes in hiring procedures are not systematically related with the quality of leavers.

[12]Note that we use teacher turnover as a method of identifying the variance in teacher quality. Implicitly, we assume teacher turnover does not directly affect student achievement gains except for the possibility of systematic quality differences by teacher experience. We test this assumption within the general production function estimation (below) and cannot reject it.

ent must be divided by the number of teachers per grade, and the coefficient on this variable provides an estimate of the within-school variance in teacher quality.

Our empirical strategy focuses on the estimation of a lower bound on the variation of teacher quality, and in that regard a variety of factors that suggest downward bias in our turnover estimator are not problematic. First is the almost certain violation of the assumption that the variance and covariance terms are equal in schools without turnover. Even in the absence of teacher turnover, there is almost certainly some difference in teacher quality from year-to-year due to changes in pedagogy, personal problems, learning (particularly for beginning teachers), etc., reducing the expected coefficient on the turnover variable below the true within-school variance.

Measurement error in the teacher turnover variable would tend to exacerbate any such downward bias. The administrative data have missing information on key variables, and it is not always clear who teaches which subjects. Consequently, there is some error introduced into the calculations of both the percentage of teachers who differ from cohort to cohort and in the number of teachers per grade, and the ratio of the two may thus contain a nontrivial amount of noise.

More worrisome for our approach, however, is that there are also two potentially important sources of upward bias. First is the standard problem of omitted variables. Teacher turnover may be precipitated or accompanied by other changes such as a new principal or superintendent or district induced curriculum changes (Ingersoll (2001)). If, for example, administrator turnover also leads to teacher turnover, any direct effects of new administrators on achievement growth would introduce an upward bias if they were not accounted for. In the empirical work below, we take a number of steps to control for potentially confounding time-varying factors including controls for the numbers of principal and superintendent changes over the observation period. We also perform various sensitivity analyses directed at these issues.

Second is the possibility that teachers who exit are not drawn randomly from the teacher quality distribution. If attrition and quality are systematically related, the average teacher quality in high turnover years will tend to differ systematically from the average quality of new hires. Consider the possibility that high quality teachers are more likely to exit. In this case, schools that obtain a particularly good draw of teachers in one year will tend to experience both a greater turnover following the year and a larger average difference in achievement gains than would be experienced with random attrition. This situation would lead to an upward bias in our estimator, as would the opposite case where low quality teachers are more likely to exit. Even if attrition and quality are uncorrelated, if teachers in the tails of the distribution are more likely

to exit, higher turnover schools will tend to have higher cohort differences in achievement gains, again biasing our estimator upward.[13]

Appendix A demonstrates that a major departure from random exiting in the form of higher probabilities in either or both tails of the distribution can introduce substantial upward bias. In the absence of student/teacher matches, we have little information on the actual distribution of departures. Moreover, the literature on teacher turnover is not very informative on the quality distribution of any school attrition.[14] A general presumption, particularly in more policy-related analyses, is that union restrictions, the single salary schedule for teachers, and the lack of performance incentives related to student achievement mute any relationship between teacher quality and attrition, but this is clearly speculative.[15] Fortunately, we do have student/teacher matches for a single district, and we use that information to provide empirical evidence on the likely magnitude and direction of any nonrandom turnover induced bias.

Finally, this framework relies on just the variation in teacher quality that is found within schools and ignores all variation in teacher quality across schools. If all schools were to hire randomly from a common pool, the between-school variance would equal zero, but this is almost certainly not the case. Rather schools able to offer higher salaries or better working conditions choose among a larger pool of applicants and likely enjoy higher average teacher quality, though the difficulty predicting productivity on the basis of education credentials and interviews almost certainly allows for substantial within-school heterogeneity.[16] In the extreme, if schools were perfectly arrayed in their hiring, all variations in quality would be between schools. In any event, the between-school differences would have to be added to the estimates reported below to obtain an estimate of the total variation in the quality of instruction.

---

[13]Note that heavy attrition in just one tail also implies drift in the average quality of teachers, which would inappropriately add to our estimate of the within-school variance (and which we explicitly assume is not the case).

[14]Much of the turnover literature (footnote 11) relates to opportunity costs by specialties (e.g., math and science), but these studies are more relevant for secondary schools and do not directly address issues of quality. Another approach investigates attrition by the teacher's own test score (see Murnane et al. (1991)) and finds some relationship suggesting that higher scoring teachers are more likely to leave, but neither this relationship nor the relationship between teacher test scores and student achievement is very strong. The one direct study relating attrition to classroom performance finds that principal evaluations early in the teaching career are positively correlated with continued teaching. At the same time, while teacher value-added based on student achievement is also positively related to retention of teachers, the estimates are statistically insignificant (Murnane (1984)), perhaps because of the small samples.

[15]For example, The Teaching Commission (2004, p. 46) notes: "once teachers have passed a probationary period, it is notoriously difficult to dismiss those whose performance is inadequate. In 2002, for instance, only 132 of 78,000 teachers in New York City's massive school system were removed for poor performance." However, no analyses of decisions before tenure or of more informal actions are available.

[16]Hanushek, Kain, and Rivkin (2004b) find that teachers who switch schools tend to move to schools with higher achieving, higher income, and lower proportion minority student bodies.

4. THE TEXAS DATABASE

The data used in this paper come from the UTD Texas Schools Project, conceived of and directed by John Kain. Data are compiled for all public school students from administrative records in Texas, allowing us to use the universe of students in the analyses. We use data for three cohorts: 3rd through 7th grade test scores for one cohort (4th graders in 1995) and 4th through 7th grade test scores for the other two (4th graders in 1993 and 1994).[17] For each cohort there are more than 200,000 students in over 3,000 public elementary and middle schools. (For details on the database, see Appendix B and Table B1; currently available data along with variable definitions and estimation programs are found in Rivkin, Hanushek, and Kain (2005).) In comparison to studies that use only a small sample of students from each school, these data permit much more precise estimates of school average test scores and test score gains.

The administrative data contain a limited number of student and family characteristics including race, ethnicity, gender, and eligibility for a free or reduced price lunch. Students who switch public schools anywhere within the state of Texas can be followed just as those who remain in the same school or district. Although explicit background measures are relatively limited, the panel feature can be exploited as described previously to account implicitly for time invariant individual and school effects on achievement.

Beginning in 1993, the Texas Assessment of Academic Skills (TAAS) was administered each spring to eligible students enrolled in grades 3 through 8.[18] These tests are designed to evaluate student mastery of the grade-specific subject matter that is prescribed for students in the state.[19] We focus on test results for mathematics and reading, derived from tests of approximately fifty questions. Because the number of questions and average percent correct varies across time and grades, we transform all test results into standardized scores with a mean of zero and variance equal to one, though the empirical findings

---

[17]Note that, while we have 3rd grade test information, our analysis begins at 4th grade because of the focus on achievement gains.

[18]Many special education and Limited English Proficiency (LEP) students are exempted from the tests, as are other students for whom the test would not be educationally appropriate. In each year roughly fifteen percent of students do not take the tests, either because of an exemption or because of repeated absences on testing days. This rate of missing tests appears comparable to those for other high quality testing programs such as the National Assessment of Educational Progress.

[19]The TAAS tests are generally referred to as criterion referenced tests, because they refer directly to pre-established curriculum or learning standards. The common alternative is norm referenced tests that cover general subject matter appropriate for the subject and grade but that are not as closely linked to the specific state teaching standards. In principle, all students could achieve the maximum score on a criterion referenced test with no variation, while norm referenced tests focus on obtaining information about the distribution of different skills across the tested population. In practice, scores on commonly available criterion referenced and norm referenced tests are highly correlated across students.

are robust to a number of transformations including the raw percentage correct. The bottom one percent of test scores (all less than or equal to expected scores from random guesses) are trimmed from the sample in order to reduce measurement error. Participants in bilingual or special education programs are also excluded from the samples used in estimating teacher quality and resource effects because of the difficulty in measuring teacher and school characteristics for these students.[20]

Student data are merged with grade average information on teachers by subject. Because student and teacher data come from different reporting systems that are not directly linked, matching students with their specific teachers is not possible. Teacher personnel data provide information on experience, highest degree earned, and the class size, subject, grade, and population served for each class taught. This information is used to construct subject and grade average characteristics for teachers in regular classrooms. In the early grades teachers tend to teach all subjects, while in junior high most specialize. We consider those who self identify as general teachers as teachers of both mathematics and reading.

5. LOWER BOUND ESTIMATES OF THE VARIANCE OF TEACHER QUALITY

The estimation of the within-school variance in teacher quality relies on the notion that teacher turnover increases the variance in student outcomes across grades and cohorts in a school. Although we refine the estimation below, the pattern can be seen directly by observing the higher correlations in student achievement across cohorts for schools with lower teacher turnover (fewer than twenty five percent of teachers are different) than schools with high turnover (fewer than twenty five percent of teachers are the same). The correlations are 0.40 in math and 0.26 in reading respectively for the low turnover schools and 0.22 in math and 0.14 in reading for the high turnover schools. Of course other factors correlated with teacher turnover could also produce this pattern, and it is necessary to turn to our more structured model in order to identify the importance of teacher quality in the determination of achievement gains. Note that on average roughly one third of teachers are new to a grade and subject in any year. This is roughly double the rate of school leaving, meaning that incumbent teachers tend to change grades or subjects every five years or so.

---

[20]For an explicit analysis of the achievement of special education students, see Hanushek, Kain, and Rivkin (2002). Kain and O'Brien (1998) provide additional analysis of special education students along with information on the performance of limited English proficiency (LEP) students. These students are included in the calculations of class sizes for the analysis below when they receive instruction in regular classrooms.

5.1. *Basic Estimates*

Table III reports basic estimates from the regression of the squared between-cohort difference in gains on the proportion of teachers who are different and other covariates. The sample includes only students who remain in the same school for two successive grades, either 5th and 6th or 6th and 7th, and only grades that have at least five students with valid test scores and nonmissing data on teacher turnover.[21] Just grades 5 and 6 are used for the small number of schools with all three grades.[22] The final sample has 3,076 schools in the mathematics specifications and 3,086 in the reading specifications.

The three left hand columns in Table III report results from the three specifications for mathematics and reading in order to isolate the sensitivity of the estimates to the different fixed components of achievement growth. The first regresses the squared difference in 5th (or 7th) grade gains between cohorts on 5th (or 7th) grade teacher turnover; the second and third regress the squared difference in the difference of 5th (or 7th) and 6th grade gains between cohorts on the turnover of 5th (or 7th) and 6th grade teachers combined. As described previously, using the difference in gains between the two grades controls for both student and school fixed effects in gains. Finally, the third specification adds an additional school fixed effect directly into the regression, identifying the variance in teacher quality on the basis of the difference in turnover rates between the first and second cohorts and the second and third cohorts. This last estimation, which captures school specific variations in the grade pattern of performance, directly controls for systematic school and grade specific unobservables that may be correlated with turnover. All three specifications also include a dummy variable identifying the precise cohort comparison, the inverse of enrollment (because the variance of measurement error in student performance is inversely proportional to enrollment), the use of 7th grade information, and the numbers of new principals and superintendents. The measures of new school and district leadership capture time varying policy factors that could simultaneously affect teacher turnover and student achievement.

The results show that differences in mathematics and reading achievement gains among cohorts are strongly related to teacher turnover. All coefficients

---

[21]An additional observation in the reading sample was also excluded, because the grade average gain was more than six standard deviations from the mean (higher than any other school). It turned out to be a single teacher whose students' average gain in the previous year was quite close to the mean and who did not teach in the subsequent year. In addition, the average gain in the subsequent grade was roughly four standard deviations below the mean, far different than the positive gain reported for the prior cohort taught by the same teacher. We believe there is overwhelming evidence of either cheating or miscoding. The exclusion of this observation did not have a large impact on the estimates except in the full fixed effect model.

[22]The majority of students move from elementary to middle school sometime between grades 5 and 7. Roughly fifteen percent of schools with at least two of the three grades in this range have all three.

TABLE III

EFFECT OF TEACHER TURNOVER ON THE DIVERGENCE OF MATHEMATICS AND READING TEST
SCORE GAINS BETWEEN COHORTS (STANDARD ERRORS IN PARENTHESES)

	No Fixed Effects[a]	Individual and School Fixed Effects[b]	Individual and School-by-Grade Fixed Effects[c]	Individual and School Fixed Effects[b]	Individual and School-by-Grade Fixed Effects[c]
1. Mathematics					
Proportion of teachers who are different/number of teachers	0.080 (0.017)	0.090 (0.015)	0.050 (0.021)	0.080 (0.016)	0.045 (0.021)
Absolute change in proportion of teachers with no experience				0.033 (0.016)	0.027 (0.023)
2. Reading					
Proportion of teachers who are different/number of teachers	0.067 (0.013)	0.082 (0.014)	0.036 (0.018)	0.078 (0.015)	0.029 (0.018)
Absolute change in proportion of teachers with no experience				0.015 (0.015)	0.041 (0.020)

Notes: All equations include the inverse of the number of students, numbers of new principals and superintendents in the school during adjacent years, a grade 7 dummy variable, and a cohort dummy variable. The sample includes all students who remain in the same school for grades 5 and 6 (or 6 and 7). Sample size is 3,076 for the mathematics and 3,086 for the reading specifications.

Equations have the same structure for mathematics and for reading. (The analyses of gain patterns between grades 6 and 7 take the same form as those for grades 5 and 6 that are shown.) For Φ = proportion different math (or reading) teachers/#teachers and adjacent cohorts (c and c'), the specifications take the following forms:

(a) $(\bar{A}_5^c - \bar{A}_5^{c'})^2 = \beta_\theta \Phi_{6s}^{c,c'} + \beta_X X_{6s}^{c,c'} + e_{6s}^{c,c'}$,

(b) $[(\bar{A}_6^c - \bar{A}_5^c) - (\bar{A}_6^{c'} - \bar{A}_5^{c'})]^2 = \beta_\theta \Phi_{5 \text{ and } 6,s}^{c,c'} + \beta_X X_{5 \text{ and } 6,s}^{c,c'} + e_{5 \text{ and } 6,s}^{c,c'}$,

(c) $[(\bar{A}_6^c - \bar{A}_5^c) - (\bar{A}_6^{c'} - \bar{A}_5^{c'})]^2 = \beta_\theta \Phi_{5 \text{ and } 6,s}^{c,c'} + \delta_s + \beta_X X_{5 \text{ and } 6,s}^{c,c'} + e_{5 \text{ and } 6,s}^{c,c'}$,

where δ_s is a fixed effect for school s.

are positive and significant at the five percent level, and except for the school-by-grade fixed effect specifications, all *t*-statistics exceed 4.5 in absolute value. The declines in coefficient magnitudes for the full fixed effect specifications are consistent with measurement error induced attenuation bias, but they may also reflect the presence of omitted variables bias in the other specifications. In order to avoid as much as possible the introduction of any upward biases, we concentrate here on the full fixed effect coefficients of 0.050 and 0.036. These imply lower bound estimates of the within school variance of teacher quality (measured in units of student achievement) equal to 0.0125 (0.050/4) and 0.009 (0.036/4) for mathematics and reading respectively. This means that a one standard deviation increase in average teacher quality for a grade raises average student achievement in the grade by at least 0.11 standard deviations of the total test score distribution in mathematics and 0.095 standard deviations in reading.

These estimates suggest the existence of substantial within school variation in teacher quality, but they combine average differences across the experience distribution with skill differences not related to experience. As we demonstrate in the direct estimation of educational production functions below, the learning curve appears to be quite steep in the first year or two of teaching before flattening out. Because many of the teachers new to a grade are in their first year, the share of the variance due to differences between beginning and experienced teachers might be quite sizeable. Fortunately, we can identify the effects of beginning teachers by including the absolute change in the share of teachers in their first year as an additional variable.[23]

The final two columns of Table III present estimates from the two fixed effect specifications that include the absolute change in the share of beginning teachers. These estimates suggest that quality differences between new and experienced teachers account for only ten percent of the teacher quality variance in mathematics and somewhere between five and twenty percent of the variance in reading. The addition of the change in the share of teachers with one year of experience (not shown) has virtually no effect on the estimates.

5.2. *Specification Checks*

The consistency of the estimator relies on the assumption that the turnover variable is unrelated to the error. One important threat to the estimation strategy is the possibility that unobserved changes over time in schools may be correlated with teacher turnover. A comprehensive control for other time varying factors in the schools comes from looking at turnover of teachers not involved in the specific subject. Specifically, by looking at schools that use separate teachers for mathematics and English, we can include English teacher turnover as a control variable in the modeling of math performance and mathematics teacher turnover in the modeling of reading achievement.[24]

Table IV reports the results for fixed effect specifications that include turnover in the untested subject. These estimates are generated from the smaller subsample of schools with subject specialists (defined as schools that have no teachers in either of the two sampled grades who teach both math and English), which is roughly thirty percent of the full sample. The results for mathematics remain highly significant though somewhat smaller in the first two specifications and are significant only at the ten percent level in the full fixed effects model, which is not that surprising given the substantial reduction in

---

[23]Because we are looking at variance in outcomes across cohorts, any significant change either up or down in the proportion of teachers in their initial year of experience has a similar impact, thus making the absolute value appropriate.

[24]Because teacher turnover in the untested subject is used to identify any concomitant disruption in the school, the number of teachers in that subject will not directly affect the variance in student performance. Therefore this turnover variable is not divided by the number of teachers in the untested subject.

TABLE IV

EFFECT OF TEACHER TURNOVER ON THE DIVERGENCE OF MATHEMATICS AND READING TEST
SCORE GAINS BETWEEN COHORTS, CONTROLLING FOR TEACHER TURNOVER IN OTHER
SUBJECTS[a] (STANDARD ERRORS IN PARENTHESES)

	Individual and School Fixed Effects[b]			Individual and School-by-Grade Fixed Effects[b]		
1. Mathematics						
Proportion different math teachers/number of teachers	0.059 (0.015)	0.058 (0.015)	0.069 (0.016)	0.034 (0.021)	0.034 (0.021)	0.035 (0.021)
Absolute change in proportion math teachers with no experience			−0.029 (0.013)			−0.005 (0.020)
Proportion of same English teachers		−0.006 (0.010)	−0.008 (0.010)		0.002 (0.014)	0.002 (0.014)
2. Reading						
Proportion different English teachers/number of teachers	0.027 (0.016)	0.024 (0.016)	0.010 (0.016)	0.001 (0.021)	−0.001 (0.021)	−0.005 (0.022)
Absolute change in proportion English teachers with no experience			0.042 (0.015)			0.013 (0.021)
Proportion of same mathematics teachers		−0.017 (0.011)	−0.016 (0.011)		−0.020 (0.013)	−0.020 (0.013)

[a]The sample includes all students who remain in the same school for grades 5 and 6 (or 6 and 7) in schools with no teacher offering both English and math instruction. All equations include the inverse of the number of students, numbers of new principals and superintendents in the school during adjacent years, a grade 7 dummy variable, and a cohort dummy variable. The sample size is 855.

[b]Table III notes describe the estimation specifications.

sample size. In contrast, the English teacher turnover coefficients in the reading test score regressions become quite small and insignificant in all specifications, raising concern that confounding factors in this estimation method could be driving the results. In this sample, the impact of inexperienced teachers is very imprecisely estimated. Importantly, comparisons across specifications for a common sample reveal that the inclusion of turnover information for the untested subject has virtually no effect on the other turnover estimate in either fixed effect specification.

The question remains as to why the estimates in Table IV are uniformly smaller than those reported in Table III. An important difference between the samples for the respective tables is the balance between 5th and 7th grade classrooms. It is almost always the case that junior high schools use subject specific teachers, while elementary schools use a single teacher for most subjects. Consequently the vast majority of schools with subject specific teachers include grades 6 and 7, while the majority of all schools in the sample include grades 5 and 6. Systematic differences by grade in the effects of teachers on test scores could therefore account for the observed pattern of results.

Table V reports estimates that allow the effect of turnover to vary by grade combination based on the full sample used in Table III. The coefficients suggest that the variance in teacher quality declines in mathematics as students progress through school, though the interaction term becomes insignificant in the full fixed effect model. On the other hand, it appears that within school differences in teacher quality are quite substantial in reading in elementary school but explain little or none of the variation in outcomes in junior high. In both subjects the pattern of estimates in Table V explain the differences between Tables III and IV. Interestingly, this pattern of diminishing effects will repeat itself in the production function estimates below, suggesting either that school

TABLE V

GRADE DIFFERENCES IN THE EFFECTS OF TEACHER TURNOVER ON THE DIVERGENCE OF MATHEMATICS AND READING TEST SCORE GAINS BETWEEN COHORTS (STANDARD ERRORS IN PARENTHESES)[a]

	Individual and School Fixed Effects	Individual and School-by-Grade Fixed Effects	Individual and School Fixed Effects	Individual and School-by-Grade Fixed Effects
1. Mathematics				
Proportion of teachers who are different/number of teachers	0.113 (0.018)	0.063 (0.026)	0.096 (0.019)	0.036 (0.027)
6th & 7th grades interaction with proportion different[b]	−0.075 (0.031)	−0.036 (0.044)	−0.068 (0.032)	−0.024 (0.047)
Absolute change in proportion of teachers with no experience			0.026 (0.022)	0.052 (0.032)
6th & 7th grades interaction with absolute change[b]			0.018 (0.033)	−0.035 (0.048)
2. Reading				
Proportion of teachers who are different/number of teachers	0.115 (0.017)	0.066 (0.022)	0.114 (0.018)	0.059 (0.023)
6th & 7th grades interaction with proportion different[b]	−0.092 (0.028)	−0.081 (0.037)	−0.101 (0.029)	−0.083 (0.038)
Absolute change in proportion of teachers with no experience			0.004 (0.020)	0.048 (0.027)
6th & 7th grades interaction with absolute change[b]			0.030 (0.031)	−0.011 (0.039)

[a] Table III notes describe the sample and estimation specifications.

[b] Interaction between an indicator for the grade 6 and 7 observations and specified variable.

and teacher quality differences have much smaller effects on achievement in junior high or that the test results do a poor job of capturing differences in school quality in those grades.

There remains one other potential source of bias that must be addressed. Although controls for any concomitant changes to teacher turnover address the problem of omitted variables, they do not resolve the potential problem of nonrandom teacher attrition described above. As noted previously, the estimation relies upon the assumption that turnover is uncorrelated with quality and is not drawn heavily from either of the tails of the quality distribution. Since our estimator is identified by the assumption of random departures, we cannot readily test this assumption within our model and data.

Fortunately, for one large Texas school district we have developed some additional data that link student test score gains with individual teachers.[25] Although we cannot account for unobservable selection into classes, sampling error, and the other factors that we explicitly worry about in this paper, we can use these data to compute a within-school measure of quality: average student achievement gains for each teacher minus the average for all teachers in the same school that year. We can then calculate attrition probabilities based on this quality measure and use these probabilities to estimate the impact of any nonrandom attrition on our estimator of the variance of teacher quality.

Table VI describes the distribution of teachers placed into twenty quality categories along with the probabilities of exit for each group. We create these categories by dividing the range of teacher average gains relative to the school average into twenty intervals of equal length. (Because of concerns about outliers, we drop the top and bottom one percent of gains, but the results are invariant to this sampling procedure as we show below.) Within each category we use the mean gain as the index of quality. Since the division into twenty categories is arbitrary, we examine the sensitivity of the results to changes in the number of intervals.

With random departures there would be no systematic differences in the probability of exiting. This does not appear to be the case in Table VI, as attrition clearly declines with quality, probably in part due to the fact that first year teachers have the highest attrition. On the other hand, attrition does not appear to be concentrated in the tails of the distribution, the key element described in Appendix A. (Note that there are very few teachers in the lowest quality category that is an outlier in the exit rate at 42.9 percent.)

We now use the method developed in the simulations in Appendix A to estimate the bias introduced by deviations from random departure of the type observed in Table VI. Table C1 shows that the nonrandom attrition leads to a very slight increase (less than one percent) in the estimated standard deviation of teacher quality. This result also holds if the number of quality intervals is doubled or tripled or if observations in the tails of the distribution are retained

---

[25]These data are described in Hanushek et al. (2005).

TABLE VI

TEACHER EXIT RATES BY QUALITY OF INSTRUCTION RELATIVE TO OTHERS
IN THE SCHOOL FOR TEACHERS IN A LARGE TEXAS DISTRICT

Quality Index	Frequency (Percent)	Exit Rate (Percent)
−1.56	0.17	42.9
−1.41	0.20	11.8
−1.27	0.45	23.7
−1.11	0.56	23.4
−0.94	1.17	30.6
−0.79	1.73	26.2
−0.63	2.86	22.2
−0.48	5.08	22.6
−0.32	9.58	21.3
−0.16	15.29	20.6
−0.01	21.35	20.2
0.14	16.65	17.65
0.29	10.58	18.51
0.45	6.51	18.35
0.60	3.55	12.79
0.76	2.07	17.34
0.92	0.96	25.00
1.07	0.62	13.46
1.22	0.43	13.89
1.38	0.19	0.00

Notes: The sample includes all teachers in grades 4–8 in one large Texas district. The measure of quality is the difference between average student gain in mathematics for a teacher and the average gain for all other teachers in the school. These relative gains are divided into twenty equal intervals, and the index for each interval is the interval mean. Frequency is the percentage of all teachers in the city in the category, and exit rate is the percentage of teachers who leave the school at the end of the year.

in the sample. Therefore, even if attrition is not random for the sample as a whole, as long as it is not far more concentrated in the tails than is observed for this single large district, it is extremely unlikely that it would introduce much if any upward bias.[26]

A final robustness check examines only schools with a single teacher per grade. This quite select sample generates large, positive, and statistically significant estimates in both mathematics and reading for the first two specifications (see Table C2). Not surprisingly given the extremely small sample sizes, the estimates for the full fixed effect specification remain positive but are quite imprecise.

---

[26]Note that the estimates of within school variation in quality based on individual teachers are three times as large as our lower bound estimates in Table III. Of course, these estimates do not deal with the selection effects that are the heart of the estimation here. They also include potentially important measurement error.

Importantly, the true magnitudes of the variances in mathematics and reading teacher quality are likely to be larger than the estimates presented here. First, the identifying assumptions are likely to be violated in ways that bias downward the extent of actual teacher quality differences within schools. Second, the measures of teacher turnover and number of teachers likely contain some error, and the ratio of the two may in fact have substantial measurement error that would likely attenuate the coefficients. For example, the exclusion of schools with large changes in the number of teachers in a grade from year to year, an indicator of problematic data, tends to increase coefficient magnitudes and the precision of virtually all estimates. Finally, we focus on just one component of the variance in teacher quality, the within-school variance. All between-school variation in teacher quality is ignored—not because of a belief it is small, but rather because it cannot be readily separated from other factors. Thus, there can be little doubt that teacher quality is an important determinant of reading and mathematics achievement in elementary school and mathematics achievement in junior high school.

6. EDUCATION PRODUCTION FUNCTION ESTIMATES

The frequently employed implicit assumption that schools are homogenous institutions is clearly contradicted by the finding of substantial within-school heterogeneity in teacher quality. These results also contrast sharply with the much smaller estimated differences in teacher and school quality that comes from studies investigating the impacts of specific school or teacher characteristics. Nevertheless, because teacher salaries are closely linked with experience and formal education and because class size reductions have been a widely discussed and often used policy tool, a better understanding of the effects of these specific factors remains important. From a policy viewpoint, a comparison of the costs and benefits of smaller classes or more educated and experienced teachers with those of improved general teacher quality would be particularly informative.

The results from the existing large body of literature on the effects of school resources on a variety of outcomes remain highly variable, in large part, we believe, because of difficulty of controlling for other relevant achievement inputs due to both conceptual and data limitations.[27] The main concern is that either explicit resource allocation rules—such as the provision of compensatory funds for poor achievers—or simple omitted variables problems could mask

---

[27]For summaries of the education production function literature, see Hanushek (1986, 2003), Levačić and Vignoles (2002), and Woessmann (2004). This work has been quite varied and controversial (Burtless (1996)). While concentrated on analyses of test score performance, continuing attention has also turned to longer run impacts on labor market outcomes (see, e.g., Card and Krueger (1992), Betts (1995), Heckman, Layne-Farrar, and Todd (1996), Dearden, Ferri, and Meghir (2002), and Dustmann, Rajah, and van Soest (2003)).

or distort true causal impacts. A set of more recent studies focuses specifically on identifying factors leading to exogenous variation in class size in order to uncover causal impacts.[28] Unfortunately, identification of truly exogenous determinants of class size, or resource allocations more generally, is sufficiently rare that other compromises in the data and modeling are frequently required. These jeopardize the ability to obtain consistent estimates of resource effects and may limit the generalizability of any findings.

As described in Section 3, our framework eliminates directly the most troubling potential endogeneity problems that are the focus of the alternative instrumental variables approaches. The large samples also permit detection of small effects that may differ by grade or student demographic characteristics, allowing us to distinguish between low power of tests and the true lack of a relationship.

6.1. Empirical Specification of Resource Models

Equation (8) describes the value-added empirical model that forms the basis of our examination of school resource effects on achievement. This is a modified version of equation (2) that adds a vector of school resource characteristics (SCH) measured at the grade level and a set of observable, time varying family characteristics (X):

$$(8) \qquad \Delta A_{ijgs}^c = SCH_{gs}^c \lambda + X_{ig}^c \beta + \underbrace{\gamma_i + \delta_{sy} + \omega_{gs} + \upsilon_{ijgs}^c}_{\text{composite error}}.$$

The family characteristics include indicator variables for students who switch schools and students who are eligible to receive a free or reduced price lunch. Teacher and school characteristics are computed separately for each grade and subject, and they include the average class size in regular classrooms,[29] the proportion of teachers with a master's degree, and the proportion of teachers who

---

[28]A variety of different approaches have been applied to sort out the causal influence of school resources including instrumental variables approaches relying upon various circumstances of the schooling institutions (e.g., Angrist and Lavy (1999), Feinstein and Symons (1999), Hoxby (2000), Woessmann and West (forthcoming), Dobbelsteen, Levin, and Oosterbeek (2002), Robertson and Symons (2003), and Bonesrønning (2004)) and direct consideration of potential pre-treatment selection factors (e.g., Dearden, Ferri, and Meghir (2002)).

[29]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. Both special purpose classes and student achievement for special education and Limited English Proficiency (LEP) students are eliminated from this estimation. At the same time, special education students in regular classroom instruction are included in the calculation of class size because they will affect the resources

fall into four experience categories: zero years, one year, two years, and three or four years (with the omitted category being five years and above).[30] The composite error terms should be reinterpreted as the unobserved components of students and schools. Note that we have added two additional error terms: school-by-year fixed effects (δ_{sy}) and school-by-grade fixed effects (ω_{gs}). These absorb the school fixed effects previously considered.

Unlike most educational studies, we concentrate specifically on the actual class sizes reported by regular classroom teachers rather than the more common pupil-teacher ratios for a school. Further, considerable attention was given to the elimination of measurement error in the school variables. We have access to longitudinal information on key data and can therefore adjust reports for inconsistencies that occur over time. Data Appendix B describes in detail the construction of the school characteristics and sample selection criteria.

Virtually all prior analyses of school resource effects have estimated specifications similar to equation (8) in either level or growth form, but none has been able to account for all of the fixed components of the composite error term. The elimination of these factors in the estimation of equation (8) addresses virtually all of the concerns typically raised about estimation of educational production functions. For example, arguments about simultaneity arising from compensatory resource allocations based on student performance are directly eliminated, since the level and expected rate of gain of achievement for each student are explicitly dealt with through the investigation of ΔA and the estimation of the individual γ_i's. The removal of school fixed effects would also control for time invariant school characteristics that might be related to the included teacher and school characteristics.

Though the removal of simple school fixed effects (δ_s) would eliminate the confounding influences of fixed school factors including stable curriculum, neighborhood factors, peer characteristics, school and district leadership, and school organization, changes over time in other school factors may be correlated to changes in the included teacher and school characteristics. Consider the possibility that other events in a school—leadership changes, curricular developments, student perceptions and flows, or the like—influence achievement directly and are correlated with changes in school and teacher characteristics. Importantly, the availability of a number of cohorts permits the inclusion of school-by-year fixed effects (δ_{sy}) rather than simple school fixed effects in some

---

allocated to regular instruction students in those classrooms. Separate analysis of special education is found in Hanushek, Kain, and Rivkin (2002).

[30]Including the percentages of teachers with five to nine and twenty or more years of experience as separate categories did not change any of the results, and the hypotheses that teachers with five to nine or twenty or more years of experience had a different impact from those with ten or more years of experience was rarely rejected at any conventional significance level. The class size and teacher education estimates also remained unchanged if average experience was used in place of the experience categories.

specifications in order to account for any such systematic year-to-year changes in school factors. Any pattern of events or policies common to the neighborhood and school will be eliminated, and the estimates are identified solely by within-school-by-year differences across grades.[31]

We believe an extremely strong case can be made that the remaining differences in class size and other teacher characteristics emanate from two uncontaminated sources: random differences between cohorts in the number of students who transfer in or out of the school as students age (i.e., changes in enrollment);[32] and school or district induced changes in class size policies that are unlikely to be systematically related to the time varying error components of individual students, controlling for student and school-by-year fixed effects in achievement gains.[33]

This approach to estimation goes well beyond what has been possible even with the specialized effects of institutional structure that have entered into past instrumental variables estimation. A concern, however, is that the signal to noise ratio falls with the removal of the multiple fixed effects, thus making it difficult to estimate the remaining elements of the specification. We consider this possibility below.

6.2. *Impact of Teacher and School Characteristics*

Table VII reports the full range of estimates obtained from value-added models that progressively contain no fixed effects; student and school fixed effects; student and school-by-year fixed effects; and, finally, student, school-by-year, and school-by-grade fixed effects.[34] Based on preliminary findings, class size effects are further allowed to differ by grade. Robust standard errors that account for the correlation of unobservables within a school are reported for all coefficients.[35] Table B1 presents descriptive statistics for the school characteristics and achievement gain.

[31]Less substantively, we also allow for changes in the tests over time through inclusion of a fixed effect for year for each subject-grade test (τ_{gy}).

[32]Note that the estimation explicitly controls for the effects of moving on the moving students' achievement growth; see Hanushek, Kain, and Rivkin (2004a).

[33]The availability of multiple cohorts also permits the inclusion of school-by-grade fixed effects, though at a cost of losing the ability to identify variable effects in the single 4th grade cohort. This may be important if, as suggested to us by Caroline Hoxby, school average achievement and class size change in a systematic way as students progress through school. However, the lack of systematic differences in class size by student demographic composition in any grade suggests that such problems are very minor if they exist at all. In the most complete model, coefficients are identified by school-by-grade-by-year differences in characteristics and achievement gains.

[34]Related to the work in the prior section, we also included (not shown) the level of teacher turnover in each year but found that it never had a systematic influence on student achievement. Stable differences in teacher turnover for each school are removed with the school fixed effects.

[35]Robust standard errors in Tables VII–IX are clustered at the school level to correct for general autocorrelations among the errors across cohorts of students attending the same school; for a discussion of the issue in a related context, see Bertrand, Duflo, and Mullainathan (2004).

TABLE VII

Effects of Teacher and School Characteristics on 4th–7th Grade Gains in
Mathematics and Reading Test Scores (Robust Standard Errors in Parentheses;
$n = 1{,}336{,}903$ for Mathematics and $1{,}330{,}791$ for Reading)

	No Fixed Effects	Student and School Fixed Effects	Student and School-by-Year Fixed Effects	Student, School-by-Grade and School-by-Year Fixed Effects
1. Mathematics				
Class size				
4th grade	−0.0049	−0.0106	−0.0107	n.a.
	(0.0023)	(0.0040)	(0.0037)	
5th grade	−0.0043	−0.0085	−0.0081	−0.0055
	(0.0010)	(0.0017)	(0.0024)	(0.0018)
6th grade	−0.0014	−0.0037	−0.0041	−0.0027
	(0.0010)	(0.0017)	(0.0020)	(0.0013)
7th grade	0.0002	0.0025	0.0032	0.0011
	(0.0009)	(0.0020)	(0.0024)	(0.0023)
Experience				
Proportion	−0.085	−0.103	−0.128	−0.073
0 years	(0.012)	(0.021)	(0.028)	(0.023)
Proportion	−0.043	−0.066	−0.055	−0.002
1 year	(0.013)	(0.022)	(0.028)	(0.023)
Proportion	−0.018	−0.045	−0.055	−.002
2 years	(0.013)	(0.021)	(0.030)	(0.022)
Proportion	−0.012	−0.031	−0.030	−0.017
3–5 years	(0.010)	(0.018)	(0.022)	(0.018)
Education				
Proportion with	−0.025	−0.018	−0.023	−0.021
graduate degree	(0.009)	(0.017)	(0.021)	(0.020)

6.2.1. *Class size*

The results reveal statistically significant effects of class size on both mathematics and reading achievement gains, but the impact declines markedly as students progress through school and tends to be smaller and less significant in reading than in mathematics. The discussion concentrates on the model that removes school-by-year fixed effects, because 4th grade estimates cannot be produced for models that contain school-by-grade fixed effects with only the single available 4th grade cohort.

The estimated effects of class size are quite similar quantitatively and qualitatively across specifications that include student and either school or school-by-year fixed effects.[36] Both the 4th and 5th grade class size coefficients are

---

[36]However, the addition of school-by-grade fixed effects substantially reduces the magnitudes and significance levels of estimates in mathematics though not in reading. Nevertheless, class size continues to exert a significant effect on mathematics achievement in grades 5 and 6. It is

TABLE VII—CONTINUED

	No Fixed Effects	Student and School Fixed Effects	Student and School-by-Year Fixed Effects	Student, School-by-Grade and School-by-Year Fixed Effects
2. Reading				
Class size				
4th grade	−0.0031	−0.0090	−0.0092	n.a.
	(0.0017)	(0.0031)	(0.0029)	
5th grade	0.0000	−0.0033	−0.0032	−0.0043
	(0.0007)	(0.0012)	(0.0018)	(0.0016)
6th grade	0.0021	0.0000	−0.0003	−0.0021
	(0.0009)	(0.0013)	(0.0019)	(0.0013)
7th grade	−0.0046	−0.0022	−0.0028	−0.0013
	(0.0008)	(0.0017)	(0.0024)	(0.0020)
Experience				
Proportion	−0.041	−0.045	−0.064	−0.026
0 years	(0.010)	(0.019)	(0.023)	(0.021)
Proportion	−0.037	−0.042	−0.070	−0.002
1 year	(0.010)	(0.018)	(0.023)	(0.020)
Proportion	−0.004	−0.006	−0.018	0.002
2 years	(0.010)	(0.019)	(0.025)	(0.020)
Proportion	0.001	0.014	0.002	0.018
3–5 years	(0.009)	(0.015)	(0.020)	(0.017)
Education				
Proportion with	−0.014	−0.004	0.001	0.010
graduate degree	(0.007)	(0.014)	(0.018)	(0.017)

Note: All specifications include a full set of grade-by-year dummies and indicators for subsidized lunch eligibility and a change of school prior to or during year. Robust standard errors in Tables VII–IX are clustered at the school level to correct for general autocorrelations among the errors across cohorts of students attending the same school; for a discussion of the issue in a related context, see Bertrand, Duflo, and Mullainathan (2004).

highly significant in both subjects, though the magnitude of the 5th grade effect is roughly three-fourths as large as that for 4th grade in mathematics and less than half as large in reading. The 6th grade effects are quite small, and by 7th grade class size appears to have little systematic effect on achievement. We discuss the magnitude of these estimates below. Note that the very large samples permit the precise estimation of quite small effects of less than 0.004 standard deviations.

The pattern of estimated class size effects also reveals the importance of controlling for student fixed effects. The inclusion of student fixed effects

not possible to know for certain the extent to which change with the addition of school-by-grade fixed effects results from the elimination of further biases as opposed to the exacerbation of any problems with measurement error.

TABLE VIII

EFFECTS OF CLASS SIZE ON TEST SCORE GAINS, BY FAMILY INCOME
(ROBUST STANDARD ERRORS IN PARENTHESES)

	Mathematics		Reading	
	Disadvantaged Students	Not Disadvantaged Students	Disadvantaged Students	Not Disadvantaged Students
Class size				
4th grade	−0.0118	−0.0103	−0.0111	−0.0087
	(0.0038)	(0.0037)	(0.0030)	(0.0029)
5th grade	−0.0077	−0.0079	−0.0027	−0.0033
	(0.0025)	(0.0024)	(0.0019)	(0.0018)
6th grade	−0.0044	−0.0040	−0.0022	−0.0007
	(0.0021)	(0.0020)	(0.0019)	(0.0017)
7th grade	0.0036	0.0031	0.0012	−0.0037
	(0.0026)	(0.0024)	(0.0023)	(0.0022)

Note: Estimates come from a single mathematics regression and a single reading regression. The models include student and school-by-year fixed effects, separate class size, and teacher experience variables for students eligible for a subsidized lunch (disadvantaged) and those not eligible during a given school year, proportion of teachers with a graduate degree, full sets of grade-by-year dummies, and indicators for subsidized lunch eligibility and a change of school prior to or during year.

triples the 4th grade coefficient and more than doubles the coefficient for 5th grade.[37]

An important and often studied question is whether lower income students receive larger benefits from class size reduction. In order to examine this claim we relaxed the restriction that class size effects were the same by income (measured by subsidized lunch eligibility). The results in Table VIII generally do not support the belief that class size effects are substantially larger for disadvantaged (subsidized lunch eligible) students. Class size effects are roughly 20 percent larger for disadvantaged students in 4th grade but actually smaller in 5th grade. Both the grade pattern and the comparable mathematics and reading results are very similar to the results in Table VII.

One potential perspective on these estimates comes from Project STAR, the random assignment experiment in class size reduction conducted in Tennessee (Word et al. (1990)).[38] While these experimental results are not directly comparable because they consider just grades *K* to 3, they indicate that a reduction

[37]The progressively more stringent estimates found across the columns does introduce some instability in the estimates, particularly in the final column. The smaller though still significant coefficients in the full fixed effects model for mathematics are consistent with the possibility that the school-by-grade and school-by-year fixed effects together aggravate problems associated with measurement error, but the results for reading go in the opposite direction.

[38]Project STAR randomly assigned a large group of kindergarten students to regular sized classes (22–25 students), regular sized classes with an aide, or small classes (13–17 students). It was designed to follow these students through grade 3, but there were significant attrition problems and subsequent additions of students to the experiment. Achievement tests were given

of eight students per class yields kindergarten achievement gains in math and reading of 0.17 standard deviations, which is roughly 60 percent larger than our 4th grade result for mathematics and reading. However, the deeper inconsistency that cannot be resolved here is that the experimental results indicate that virtually all of the achievement gain in STAR is associated with the first year in a small class—generally kindergarten or 1st grade—and not subsequent small class treatments (Krueger (1999)), while we find that smaller classes still have an effect in 4th and 5th grade.

The STAR experiment also reveals very large variation in student performance across individual classrooms. Specifically, all randomization occurred within each experimental school, and students in the large classes outperformed schoolmates in smaller classes in almost half of the schools (Hanushek (1999b)). This experimental finding is consistent with the conclusions here that differences in teacher quality within schools are quite large.

The school-by-year fixed effect estimates in column 3 of Table VII provide the basis for a simple comparison of policy alternatives. While it is difficult to estimate the cost of improving teacher quality, our lower bound estimates of the variation in quality found just within schools indicate that one standard deviation in quality is worth at least 0.11 standard deviations higher annual growth in mathematics achievement and 0.095 standard deviations higher annual growth in reading in elementary school. This magnitude of change is equivalent to a class size reduction of approximately ten students in 4th grade and thirteen or more students in 5th grade, and an implausibly large number in 6th grade. In 7th grade there appears to be no significant benefit from smaller classes in mathematics, while in reading neither class size nor teacher quality appears to exert a substantial effect on achievement. Note that these comparisons assume both no accompanying changes in teacher quality and linearity in class size effects, the latter of which appears reasonable based on semiparametric estimates for class sizes between 10 and 35 students (results not reported).

6.2.2. *Teacher characteristics*

The results for teacher experience generally support the notion that beginning teachers and to a lesser extent second and third year teachers in mathematics perform significantly worse than more experienced teachers. There may be some additional gains to experience in the subsequent year or two, but the estimated benefits are small and not statistically significant in both mathematics and reading in any of the fixed effect specifications. Similar to the case for class size, the results in the full fixed effect model in column 4 are much weaker

---

at the end of each grade, and a comparison showed that students in small classes outperformed those in regular classes in their first experimental year (K or 1) but that no additional gains were made. See Hanushek (1999b) and Krueger (1999).

than in the other fixed effects models, consistent with the view that multiple fixed effects can exacerbate problems with measurement error. The addition of school-by-grade fixed effects reduces the magnitude of all coefficients, and only the estimated effect of proportion of new teachers on math achievement gain is significant.

Importantly, the teacher experience effect conceptually combines two very distinct phenomena. First, new teachers may need to go through an adjustment period where they learn the craft of teaching along with adjusting to the other aspects of an initial job. Second, a number of the early 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 (Hanushek, Kain, and Rivkin (2004b)). The estimated parameters in Table VII combine the effects of on-the-job learning and of selective exit and mobility.

Table IX presents the basic estimates of first year teaching on achievement (with individual and school fixed effects) for samples that exclude those who immediately leave teaching or switch schools. The close similarity of the estimates across the samples compared to those in Table VII for both mathematics and reading indicates that on-the-job learning is the dominant element of the experience effect. Importantly, these results also suggest that the average quality of those who quit teaching after one year is similar to the average quality of those who remain, providing additional support for the validity of the estimates of the variance in teacher quality.

TABLE IX

EFFECTS OF PROPORTION OF TEACHERS WITH ZERO YEARS OF EXPERIENCE ON
MATHEMATICS AND READING TEST SCORE GAINS, BY NEW TEACHER TRANSITIONS
(ROBUST STANDARD ERRORS IN PARENTHESES)

Outcome Measure	Excluding Teachers Who Exit Teaching or Switch Schools	Excluding Teachers Who Exit Teaching	All Teachers
1. Mathematics			
Proportion of teachers with 0 years experience	−0.105 (0.030)	−0.114 (0.028)	−0.103 (0.021)
Observations	[1,185,329]	[1,210,155]	[1,336,903]
2. Reading			
Proportion of teachers with 0 years experience	−0.040 (0.024)	−0.040 (0.023)	−0.045 (0.019)
Observations	[1,181,611]	[1,206,139]	[1,330,791]

Note: Estimates come from a model that includes student and school fixed effects. Specifications also include the percentage of teachers with a graduate degree, full sets of class size variables, and grade-by-year dummies and indicators for subsidized lunch eligibility and a change of school prior to or during year.

Finally, 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 statistically insignificant.

7. CONCLUSIONS

Prior investigations of school and teacher effects have raised as many questions as they have answered, in large part because of the difficulties introduced by the endogeneity of school and classroom selection and in part because of the failure of observable teacher characteristics to explain much of the variation in student performance. The models and data used in this paper permit us to draw a number of sharp conclusions about public elementary education and to provide clear answers for the questions raised in the Introduction.

(i) *Teachers and therefore schools matter importantly for student achievement.* The issue of whether or not there is significant variation in school quality has lingered, quite inappropriately, since the original Coleman Report. This analysis identifies large differences in the quality of instruction in a way that rules out the possibility that the observed differences are driven by family factors.

The Coleman Report also popularized the issue of whether family influences are "more important" than school influences. This is not the relevant question for policy, which should focus on whether the benefits produced by any intervention justify the costs. Though our analysis does not consider the costs of raising teacher quality, the estimated variation in the quality of instruction clearly reveals an important role for schools and teachers in promoting economic and social equality. Even if none of the between-school variation in achievement is attributed to schools or teachers, it is clear that school policy can be an important tool for raising the achievement of low income students and that a succession of good teachers could, by our estimates, go a long way toward closing existing achievement gaps across income groups. At the very least, more must be known about the feasible means of providing such consistently high quality teachers.

(ii) *Achievement gains are systematically related to observable teacher and school characteristics, but the effects are generally small and concentrated among younger students.* This analysis used a fixed effects approach to identify the causal relationship between achievement and key school resources. Four major conclusions emerge from this work.

- Similar to most past research, we find absolutely no evidence that having a master's degree improves teacher skills.
- There appear to be important gains in teaching quality in the first year of experience and smaller gains over the next few career years. However, there is little evidence that improvements continue after the first three years.
- Class size appears to have modest but statistically significant effects on mathematics and reading achievement growth that decline as students progress through school.

• Any differences in school resource effects by family income are small.

Partially consistent with recent experimental and statistical efforts to identify class size effects, we find that lowering class size has a positive effect on mathematics and reading achievement, though the magnitude of the effect is small, particularly following 5th grade. The costs of class size reduction have not been well estimated, but they are likely to exceed the proportional increase in the number of teachers needed to staff the smaller classes. First, class size reduction almost certainly leads to more support expenditure, increased building requirements, and the like. Second, and more directly relevant to this discussion, it is highly unlikely that the supply of teacher quality is perfectly elastic, so that expansion of the teacher work force, at least in the short run, is likely to lead either to increased salary demands or a reduction in teacher quality. Moreover, the potential tradeoff between teacher quality and class size is probably most acute in difficult to staff schools serving largely disadvantaged student populations (Hanushek (1999a), Jepsen and Rivkin (2002)).

(iii) *The disjuncture between estimates of the variation of teacher quality and the explanatory power of measured teacher characteristics creates a clear dilemma for policy makers.* Though it is tempting to tighten standards for teachers in an effort to raise quality, the results in this paper and elsewhere raise serious doubts that more restrictive certification standards, education levels, etc. will succeed in raising the quality of instruction. Rather the substantial differences in quality among those with similar observable backgrounds highlight the importance of effective hiring, firing, mentoring, and promotion practices. Research shows that principals can, when asked, separate teachers on the basis of quality (Murnane (1975), Armor et al. (1976)), but the substantial variation documented in this paper strongly suggests that personnel practices in the Texas public schools are very imperfect.

One dimension of policy does, nonetheless, deserve special attention. Economically disadvantaged students systematically achieve less than more advantaged students, on average falling some 0.6 standard deviations behind.[39] While we find little reason to believe that school resources have a larger impact on disadvantaged students, we do know that low income and minority students face higher teacher turnover and tend to be taught more frequently by beginning teachers (Hanushek, Kain, and Rivkin (2004b)). Because beginning teachers, regardless of their ultimate abilities, tend to perform more poorly, policies should be developed to both keep more senior teachers in the classrooms of disadvantaged students and to mitigate the impact of inexperience. These may include improved mentoring of new teachers and policies designed specifically to cut down teacher turnover. Of course, it goes without saying that

---

[39]The measure of family income is eligibility for a free or reduced price school lunch. This measure, while quite commonly used because of its availability in administrative records, is an imprecise categorization of economic circumstances.

effective policies will pay particular attention to the substantial variation in teacher quality.

The desirability of specific policy changes remains quite speculative because of the limited experience with alternative organizational forms, incentives, and accountability policies. A very appealing though untested approach to raising teacher quality would move the focus away from the state legislatures and schools of education and toward principals and other administrators (Hanushek and Rivkin (2004)). In the presence of incentives such as expanded choice, school report cards, or other types of accountability systems, administrators would likely alter their behavior and personnel policies in ways that benefit students. In particular, there would likely be much more focus on student outcomes of interest. Not only would improved personnel policies likely raise the performance level of existing teachers, there is strong reason to believe that a closer link between rewards and performance would improve the stock of teachers. Of course inappropriate incentives likely lead to adverse outcomes, and it is imperative that schools learn from their mistakes and evolve toward more effective systems of school governance.

Dept. of Economics, Amherst College, Amherst, MA 01002, U.S.A.; sgrivkin@amherst.edu,

Hoover Institution, Stanford University, Stanford, CA 94305, U.S.A.; hanushek@stanford.edu; http://www.hanushek.net,

and

University of Texas at Dallas (deceased).

APPENDIX A: THE EFFECT OF NONRANDOM TEACHER ATTRITION ON TURNOVER-BASED ESTIMATOR OF TEACHER QUALITY VARIATION

The estimator of teacher quality derived from equation (7) assumes that the error term (e) is uncorrelated with teacher turnover. If, however, there is systematic teacher attrition that varies by quality, the estimator may no longer be a lower bound but may in fact overestimate the variance in quality. This specifically would be the case if attrition is concentrated in the tails of the quality distribution. It is most natural to think of this as a problem of sample selection where teachers who depart have a different distribution in terms of quality than those who remain. Thus, schools with turnover would tend to have a different quality distribution for teachers.

The nature of the problem with selective attrition using our estimator is easiest to see in the simpler comparison of the squared difference in grade g gains for successive cohorts, although it would easily generalize to the full estimator. The subtraction of 5th grade average gain from 6th grade average gain for a cohort removes any student and school fixed effects (including overall hiring practices) but does not address problems related to nonrandom teacher departures.

TABLE A1

UNDERLYING DISTRIBUTION OF TEACHER
QUALITY FOR NEW HIRES

Relative Teacher Quality (q)	Frequency: $f(q)$
−1	0.25
0	0.50
1	0.25

The potential impact of selective attrition is directly seen from a simple simulation using a trinomial quality distribution. Table A1 describes a distribution of new hires that has a variance of quality equal to 0.5. With this distribution of new hires, it is possible to simulate the estimator of school quality both with random departures and with systematic departures that differ across the distribution.

First, consider the turnover-based estimator of the variance in teacher quality when there are random departures. Table A2 begins with the distribution of teacher quality in Table A1 and then assumes that teachers leave randomly (and are replaced by a random selection of teachers according to the distribution in Table A1). Consequently there are nine possible transitions, three for each of the period 0 quality categories.

In this simple one grade example, the expected period 0/period 1 difference in quality is two times the variance in teacher quality (instead of four times the variance as derived in the full estimator that considers deviations across grades and cohorts). Table A2 shows that the estimator yields the true variation in quality when there is random hiring and departures.

Consider, however, the identical estimator with strongly nonrandom departures characterized by probabilities of departure of 0.5, 0.0, and 0.5 for the

TABLE A2

TRANSITION MATRIX AND VARIANCE ESTIMATE WITH RANDOM ATTRITION

Relative Teacher Quality (q_0) Period 1	Relative Teacher Quality (q_1) Period 2	Transition Frequency: $f(q_1, q_0)$	Squared Quality Difference $(q_1 - q_0)^2$
	−1	0.0625	0
−1	0	0.125	1
	+1	0.0625	4
	−1	0.125	1
0	0	0.250	0
	+1	0.125	1
	−1	0.0625	4
+1	0	0.125	1
	+1	0.0625	0

Notes: Weighted sum of squared differences = 1.0; estimated variance = 1/2 squared differences = 0.5.

TABLE A3

TRANSITION MATRIX AND VARIANCE ESTIMATE WITH NONRANDOM ATTRITION
CONCENTRATED IN THE TAILS OF THE QUALITY DISTRIBUTION

Relative Teacher Quality (q_0) Period 1	Relative Teacher Quality (q_1) Period 2	Transition Frequency: $f(q_1, q_0)$	Squared Quality Difference $(q_1 - q_0)^2$
	−1	0.125	0
−1	0	0.250	1
	+1	0.125	4
	−1	0.0	1
0	0	0.0	0
	+1	0.0	1
	−1	0.125	4
+1	0	0.250	1
	+1	0.125	0

Notes: Weighted sum of squared differences $= 1.5$; estimated variance $= 1/2$ squared differences $= 0.75$.

three quality groups in Table A1. Table A3 describes the transition probabilities, sum of squared quality differences, and the simulated variance estimates. If departures were as concentrated in the tails of the distribution as they are in this example, our method would overstate the variance in teacher quality by 50 percent: 0.75 instead of 0.5. Note that this upward bias would also arise if all departures were concentrated in only one of the tails of the distribution.

In general, if attrition is weighted toward the tails of the quality distribution the turnover-based estimator will tend to overestimate the variance of quality, and the opposite will hold if attrition is concentrated in the center of the quality distribution.

APPENDIX B: TEXAS SCHOOL DATA

The data that are used in this paper come from the data development activity of the UTD Texas Schools Project of the University of Texas at Dallas; see Kain (2001). 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 data base. Test score results and a limited amount of student demographic information are stored in a separate data base maintained by TEA and must be merged with the student data on the basis of unique student IDs. Data are compiled for all public school students in Texas, allowing us to use the universe of students in the analyses. In this paper all of the information on students comes from the test score data base, and we combine student information from the Texas Assessment of Academic Skills (TAAS) data base with teacher and school information contained in the PEIMs data base for three student cohorts: 3rd through 7th grade test

TABLE B1

VARIABLE MEANS AND STANDARD DEVIATIONS

	Math Test Score Gain	Reading Test Score Gain	Class Size	Teacher Characteristics			Observations
				% with Graduate Degree	% 0 Years Experience	% 1 Year Experience	
4th grade	−0.01 (0.70)	−0.02 (0.73)	19.5 (2.3)	23.7 (24.3)	6.1 (12.4)	5.9 (12.5)	143,314
5th grade	0.01 (0.64)	0.01 (0.68)	22.6 (3.6)	25.1 (26.2)	5.9 (13.7)	6.0 (13.6)	438,561
6th grade	0.02 (0.61)	0.02 (0.68)	22.1 (3.9)	24.5 (27.4)	7.4 (16.6)	6.9 (15.7)	455,438
7th grade	−0.02 (0.55)	−0.01 (0.66)	21.5 (4.2)	22.0 (26.7)	9.2 (18.4)	8.9 (18.0)	299,590

scores for one cohort (4th graders in 1995) and 4th through 7th grade test scores for the other two (4th graders in 1993 and 1994).[40]

Beginning in 1993, the Texas Assessment of Academic Skills (TAAS) was administered each spring to eligible students enrolled in grades 3 through 8. We focus on test results for mathematics and reading. The bottom one percent of test scores 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 students who split time between regular classrooms and special programs.

Student data are merged with information on teachers using unique school identifiers. The personnel data provide information on all Texas public school teachers for each year. Experience and highest degree earned are reported, as are the class size, subject, grade, and population served for each class taught. Although the currently available data do not permit linking individual students with specific teachers, the available information is used to construct subject and grade average characteristics for teachers in regular classrooms.

In an effort to reduce problems associated with measurement error, a number of observations are excluded from the data set. The following paragraphs describe in detail the construction of the variables and the sample selection procedures.

Measurement error in the teacher characteristics is an important issue. In many cases reported teacher experience in one year does not correspond with reported teacher experience for other years. If the experience sequence is valid except for one or two years that do not follow from the others, we correct ex-

---

[40]Note that, while we have 3rd grade test information, our analysis begins at 4th grade because of the focus on achievement gains.

perience for those years. If experience data are inconsistent for all the years, if there are two consistent patterns, or if correction would impute negative years of experience, no corrections are made. In any case, no teachers are excluded from the final sample on the basis of inconsistent experience data, though the results are not sensitive to their inclusion, possibly because we used discreet experience categories.

The case of average class size is somewhat more complicated. Teachers were asked to report the average class size for each class they taught that was of a different size. Unfortunately, many teachers appear to have reported the total number of students taught per day. This becomes particularly problematic for schools that move from general to subject specific teachers. Consider a school with two 4th grade classes of twenty students in which the two teachers each teach all subjects. If the school switches to math and reading specialists for 5th grade and each teaches one subject for each class, they will report class sizes of forty if they report total number of students served. It will appear that class sizes doubled as students aged, when in fact they remain the same.

In order to reduce problems introduced by measurement, all reported class sizes that fall below ten or above twenty five in 4th grade (thirty five in higher grades) are set to missing prior to the computation of school averages for each grade. By statute, 4th grade classes are not supposed to exceed twenty two students, though some schools receive waivers to provide slightly larger classes. It is our understanding that very few elementary schools in Texas have actual class sizes in later grades that exceed thirty five students during this period. Estimates of class size effects increased in magnitude following these exclusions, suggesting that class size was measured with error for these schools.

Access to the administrative data on student performance is currently restricted by U.S. federal law. Further information on data access along with the specific variable definitions, data construction, and data that may currently be released are found in Rivkin, Hanushek, and Kain (2005).

APPENDIX C: ALTERNATIVE TEACHER QUALITY ESTIMATES

TABLE C1
TEACHER QUALITY STANDARD DEVIATION ESTIMATES CALCULATED FROM SQUARED DIFFERENCE IN QUALITY FOR PERIODS 0 AND 1, BASED ON OBSERVED DISTRIBUTIONS OF TEACHER QUALITY AND DEPARTURE RATES

Number of Teacher Quality Intervals	σ Assuming Random Departures	σ Assuming Empirical Distribution of Departures
20 (Table VI)	0.395	0.399
40	0.397	0.401
60	0.397	0.402
30 with tails	0.422	0.427

TABLE C2

EFFECT OF TEACHER TURNOVER ON THE DIVERGENCE OF GAINS IN MATHEMATICS AND
READING TEST SCORES BETWEEN COHORTS FOR SCHOOLS WITH ONE TEACHER
PER GRADE (STANDARD ERRORS IN PARENTHESES)

	No Fixed Effects	Individual and School Fixed Effects	Individual and School-by-Grade Fixed Effects
1. Mathematics			
Proportion different math teachers/number of teachers	0.124 (0.039)	0.117 (0.039)	0.042 (0.047)
2. Reading			
Proportion different English teachers/number of teachers	0.181 (0.037)	0.180 (0.049)	0.061 (0.042)

Notes: All equations include the inverse of the number of students, numbers of new principals and superintendents in the school during adjacent years, and a cohort dummy variable. Sample size is 294 for the mathematics and 300 for the reading specifications. Table III notes describe the estimation specifications.

REFERENCES

ANGRIST, J. D., AND V. LAVY (1999): "Using Maimondides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533–575.

ARMOR, D. J., P. CONRY-OSEGUERA, M. COX, N. KING, L. MCDONNELL, A. PASCAL, E. PAULY, AND G. ZELLMAN (1976): *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools*. Santa Monica, CA: Rand Corp.

BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 114, 249–275.

BETTS, J. R. (1995): "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *Review of Economics and Statistics*, 77, 231–247.

BONESRØNNING, H. (2004): "The Determinants of Parental Effort in Education Production: Do Parents Respond to Changes in Class Size?" *Economics of Education Review*, 23, 1–9.

BOOZER, M. A., AND C. ROUSE (1995): "Intraschool Variation in Class Size: Patterns and Implications," Working Paper 5144, National Bureau of Economic Research, Cambridge, MA.

BURTLESS, G. (ED.) (1996): *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington, DC: Brookings.

CARD, D., AND A. B. KRUEGER (1992): "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, 100, 1–40.

COLEMAN, J. S., E. Q. CAMPBELL, C. J. HOBSON, J. MCPARTLAND, A. M. MOOD, F. D. WEINFELD, AND R. L. YORK (1966): *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

DEARDEN, L., J. FERRI, AND C. MEGHIR (2002): "The Effect of School Quality on Educational Attainment and Wages," *Review of Economics and Statistics*, 84, 1–20.

DOBBELSTEEN, S., J. LEVIN, AND H. OOSTERBEEK (2002): "The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition," *Oxford Bulletin of Economics and Statistics*, 64, 17–38.

DOLTON, P. J., AND W. VAN DER KLAAUW (1995): "Leaving Teaching in the UK: A Duration Analysis," *The Economic Journal*, 105, 431–444.

——— (1999): "The Turnover of Teachers: A Competing Risks Explanation," *Review of Economics and Statistics*, 81, 543–552.

DUSTMANN, C., N. RAJAH, AND A. VAN SOEST (2003): "Class Size, Education, and Wages," *Economic Journal*, 113, F99–120.

FEINSTEIN, L., AND J. SYMONS (1999): "Attainment in Secondary Schools," *Oxford Economic Papers*, 52, 300–321.

GREENWALD, R., L. V. HEDGES, AND R. D. LAINE (1996): "The Effect of School Resources on Student Achievement," *Review of Educational Research*, 66, 361–396.

HANUSHEK, E. A. (1971): "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," *American Economic Review*, 60, 280–288.

——— (1979): "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources*, 14, 351–388.

——— (1986): "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24, 1141–1177.

——— (1992): "The Trade-Off Between Child Quantity and Quality," *Journal of Political Economy*, 100, 84–117.

——— (1996): "A More Complete Picture of School Resource Policies," *Review of Educational Research*, 66, 397–409.

——— (1999a): "The Evidence on Class Size," in *Earning and Learning: How Schools Matter*, ed. by S. E. Mayer and P. E. Peterson. Washington, DC: Brookings Institution, 131–168.

——— (1999b): "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects," *Educational Evaluation and Policy Analysis*, 21, 143–163.

——— (2003): "The Failure of Input-Based Schooling Policies," *Economic Journal*, 113, F64–F98.

HANUSHEK, E. A., AND J. F. KAIN (1972): "On the Value of 'Equality of Educational Opportunity' as a Guide to Public Policy," in *On Equality of Educational Opportunity*, ed. by F. Mosteller and D. P. Moynihan. New York: Random House, 116–145.

HANUSHEK, E. A., J. F. KAIN, D. M. O'BRIEN, AND S. G. RIVKIN (2005): "The Market for Teacher Quality," Working Paper, National Bureau of Economic Research, Cambridge, MA.

HANUSHEK, E. A., J. F. KAIN, AND S. G. RIVKIN (2002): "Inferring Program Effects for Specialized Populations: Does Special Education Raise Achievement for Students with Disabilities?" *Review of Economics and Statistics*, 84, 584–599.

——— (2004a): "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools," *Journal of Public Economics*, 88/9–10, 1721–1746.

——— (2004b): "Why Public Schools Lose Teachers," *Journal of Human Resources*, 39, 326–354.

HANUSHEK, E. A., AND S. G. RIVKIN (2004): "How to Improve the Supply of High Quality Teachers," in *Brookings Papers on Education Policy 2004*, ed. by D. Ravitch. Washington, DC: Brookings Institution Press, 7–25.

HECKMAN, J. J., A. LAYNE-FARRAR, AND P. TODD (1996): "Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings," *Review of Economics and Statistics*, 78, 562–610.

HOXBY, C. M. (2000): "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 115, 1239–1285.

INGERSOLL, R. M. (2001): "Teacher Turnover and Teacher Shortages: An Organizational Analysis," *American Educational Research Journal*, 38, 499–534.

JEPSEN, C., AND S. G. RIVKIN (2002): "What Is the Trade-Off Between Smaller Classes and Teacher Quality?" National Bureau of Economic Research, Cambridge, MA.

KAIN, J. F. (2001): "The UTD Texas Schools Microdata Panel (TSMP): Its History, Use and Ways to Improve State Collection of Public School Data," Paper Prepared for The Secretary's Forum on Research and Value-Added Assessment Data, U.S. Department of Education; *http://utdallas.edu/research/tsp/index.htm.*

KAIN, J. F., AND D. M. O'BRIEN (1998): "A Longitudinal Assessment of Reading Achievement: Evidence from the Harvard/UTD Texas Schools Project," University of Texas at Dallas, Dallas, TX.

KRUEGER, A. B. (1999): "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 114, 497–532.

LEVAČIĆ, R., AND A. VIGNOLES (2002): "Researching the Links Between School Resources and Student Outcomes in the UK: A Review of Issues and Evidence," *Education Economics*, 10, 313–331.

MURNANE, R. J. (1975): *Impact of School Resources on the Learning of Inner City Children*. Cambridge, MA: Ballinger.

——— (1984): "Selection and Survival in the Teacher Labor Market," *Review of Economics and Statistics*, 66, 513–518.

MURNANE, R. J., AND R. OLSEN (1989): "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from Michigan," *Review of Economics and Statistics*, 71, 347–352.

MURNANE, R. J., AND B. PHILLIPS (1981): "What Do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research*, 10, 83–100.

MURNANE, R. J., J. D. SINGER, J. B. WILLETT, J. J. KEMPLE, AND R. J. OLSEN (1991): *Who Will Teach? Policies That Matter*. Cambridge, MA: Harvard University Press.

RIVKIN, S. G., E. A. HANUSHEK, AND J. F. KAIN (2005): "Variable Definitions, Data, and Programs for 'Teachers, Students, and Academic Achievement'," *Econometrica Supplementary Material*, 73, 2, *www.econometricsociety.org/ecta/supmat/4139data.pdf*.

ROBERTSON, D., AND J. SYMONS (2003): "Do Peer Groups Matter? Peer Group versus Schooling Effects on Academic Attainment," *Economica*, 70, 31–53.

STINEBRICKNER, T. R. (2002): "An Analysis of Occupational Change and Departure from the Labor Force," *Journal of Human Resources*, 37, 192–216.

SUMMERS, A. A., AND B. L. WOLFE (1977): "Do Schools Make a Difference?" *American Economic Review*, 67, 639–652.

THE TEACHING COMMISSION (2004): *Teaching at Risk: A Call to Action*. New York, NY: The Teaching Commission.

TIEBOUT, C. M. (1956): "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 64, 416–424.

WOESSMANN, L. (2004): "Educational Production in Europe," Paper Presented at 40th Meeting of *Economic Policy* in Amsterdam Ifo Institute for Economic Research at the University of Munich.

WOESSMANN, L., AND M. R. WEST (forthcoming): "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS," *European Economic Review*, forthcoming.

WORD, E., J. JOHNSTON, H. P. BAIN, B. D. FULTON, J. B. ZAHARIES, M. N. LINTZ, C. M. ACHILLES, J. FOLGER, AND C. BREDA (1990): *Student/Teacher Achievement Ratio (STAR), Tennessee's K-3 Class Size Study: Final Summary Report, 1985–1990*. Nashville, TN: Tennessee State Department of Education.