TEACHERS, SCHOOLS, AND ACADEMIC ACHIEVEMENT

Steven G. Rivkin, Eric A. Hanushek, and John F. Kain

September 2003

Abstract

Considerable controversy surrounds the impact of schools and teachers on the achievement of students. This paper disentangles the separate factors 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. Nonparametric lower bound estimates of the variance in teacher quality based entirely on within-school heterogeneity indicate that teachers have powerful effects on achievement, though little of the variation in teacher quality is explained by observable characteristics such as education or experience. First and second year teachers are systematically less effective than more experienced peers, and students do benefit from smaller classes, particularly in grades 4 and 5. The effects on mathematics and reading achievement of a very costly ten student reduction in class size similar to that undertaken in a variety of U.S. states 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.
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. 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 each. It then employs a unique panel data set of students in Texas to identify the sources of differences in student achievement and of 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.

---

1The 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.
By employing an extraordinarily rich data set for achievement of students in the State of Texas, we can provide quite precise answers to each of these questions. The data contain test scores spanning grades three through seven for three cohorts of students in the mid-1990s. The multiple cohorts and grades permit clear identification of teacher and school factors. And, the very large samples, coming from repeated observations on more than one-half million students in over 3,000 schools, permit the detection of quite small effects.

A primary objective of the initial empirical analysis is to obtain estimates of differences in teacher contributions to student learning that eliminate any 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 confounds school or teacher effects with the influences of unmeasured individual, family, school and neighborhood factors, making it difficult to interpret between-school or even between-classroom differences in achievement.

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 can substantially offset or even eliminate the disadvantage of low socioeconomic background. Moreover, they highlight key policy trade-offs. For example, the effects of a one standard deviation improvement in teacher quality exceed those of a very costly ten student reduction in class size. 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 bulk of research showing that observed teacher characteristics including experience and education explain little of the variation in student achievement.

The next section provides an overview of how variations in teacher quality affect the pattern of achievement growth within schools. Section II describes the estimation strategy within the context of prior analyses and shows how a lower bound on the impact of teacher quality can be identified by within-school differences in the growth of student achievement. Section III provides a detailed description of the Texas data on students and teachers. Section IV reports estimates of the variance in teacher quality based on the method developed in Section II, and Section V presents an extension of traditional analyses of the effects on outcomes of measured resources: class size, teacher education, and teacher experience. The final section considers the policy implications of the findings, particularly the importance of measured resources relative to the overall contribution of teachers.

I. 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 by teachers
explain little of the true variation in quality.

To motivate the concentration on teacher quality, we begin with aggregate statistics on the
performance of schools and teachers. Table 1 displays correlations of school average annual
mathematics and reading achievement gains in grades five, six, and seven between two cohorts of
students for all public elementary schools in Texas. 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
between-school 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
sixth and seventh grade reading performance, the correlation is -0.01. In contrast, the correlations
between school 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.

---

2These data, subsequently used in the detailed empirical analyses, are described in detail in Section III, below. All correlations relate just to students in schools that have both of the relevant grades.
Table 1. Correlations by grade and by cohort (school year) of average gains in student mathematics reading performance across common cohorts within schools

<table>
<thead>
<tr>
<th>Grade of Cohort I</th>
<th>Mathematics Grade of Cohort II</th>
<th>Reading Grade of Cohort II</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.32**</td>
<td>0.19**</td>
</tr>
<tr>
<td>6</td>
<td>0.12** 0.52**</td>
<td>0.13** 0.43**</td>
</tr>
<tr>
<td>7</td>
<td>0.05 0.46**</td>
<td>-0.01 0.44**</td>
</tr>
</tbody>
</table>

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 correlations weighted by average enrollment of the pairs.

*: Significant at 10% level; ** significant at 1% level
Table 2 presents some preliminary evidence in support of the view that commonly available teacher and classroom characteristics including experience, education, and class size explain little of the substantial variation in teacher quality that exists. The table reports the $R^2$ 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 below, 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); and the last substitutes teacher fixed effects for the observable teacher and classroom characteristics.

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, the inclusion of school rather than teacher fixed effects reduces the explanatory power by roughly two thirds (not shown), suggesting that much of the variation in teacher quality exists within rather than between schools.

Tables 1 and 2 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 fourth 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 non-random 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.
Table 2. Comparison of the Explanatory Power of Teacher Experience, Education, and Class Size with Teacher Fixed Effects in Explaining Achievement Gains

<table>
<thead>
<tr>
<th>Included explanatory variables</th>
<th>Mathematics</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Covariates</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Teacher Characteristics</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Teacher Fixed Effects</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>R Squared</td>
<td>0.0153</td>
<td>0.0183</td>
</tr>
<tr>
<td>Observations</td>
<td>93,734</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Dependent variables are mathematics and reading test score gains; sample includes only grades in a school with a single teacher.
II. 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 have potentially 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 nonparametric 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.

A. 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.\(^3\)

Equation (1) presents a conventional value-added model that describes the gain in student achievement \(\Delta A^c_{ijgs} \) for individual i in cohort c with teacher j in grade g of school s.

\[
\Delta A^c_{ijgs} = A^c_{ijgs} - A^c_{ijgs-1s} = X^c_{ijgs} \beta_X + T^c_{ijgs} \beta_T + S^c_{gs} \beta_S + f_i + \epsilon^c_{ijgs}
\]

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 \((\epsilon)\). Note that “inherent abilities” refer 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.\(^4\) Each of the inputs can be thought of as a vector of underlying components.

---

\(^3\)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 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.

\(^4\)The isolation of inherent student abilities does not rely on any presumption about their source (genetic, environmental, or an interaction of these). Any fixed learning 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 \textit{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.
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, for example, 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 included 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, for example, 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 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.5

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.

5Hanushek (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).
B. 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 nonparametric 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.

\[
\Delta A_{ijgs}^c = \gamma_i + \theta_j + \delta_s + \nu_{ijgs}^c
\]

Test score gain in grade g is written as an additive function of student (\(\gamma\)), teacher (\(\theta\)), and school (\(\delta\)) fixed effects along with a random error (\(\nu\)) that is a composite of time-varying components. The fixed student component captures the myriad family influences that persist over time including parental influences, family structure, and permanent income; similarly, the other two fixed factors incorporate the persistent effects of teacher and school quality including the persistent effects of resources, peers, curriculum, etc. Finally, the error term captures all time varying effects. Though the model highlights the fixed components of education production, the analyses below considers issues related to both fixed and time varying factors.

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. Many school policies – hiring, curriculum, school structure, etc. – emanate from school districts and will produce common elements in the teacher, school, and grade 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. 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.

C. Estimator of the Variance of Teacher Quality

In the nonparametric approach of equation (2), the variance of $\Theta$ 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 any between classroom variation would introduce serious impediments to the identification of the variance of teacher quality.

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 variance, and the aggregation to the grade level circumvents any problems resulting from...
classroom assignment.\textsuperscript{9} The cost of this aggregation is the loss of all within grade variation in teacher quality and the inability to trace out the quality of individual teachers.

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:

\[
\Delta A_{gs}^c = \gamma_{gs}^c + \theta_{gs}^c + \delta_s + \nu_{gs}^c
\]

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

\[
\Delta A_{gs}^c - \Delta A_{gs}^{c'} = (\gamma_{gs}^c - \gamma_{gs}^{c'}) + (\theta_{gs}^c - \theta_{gs}^{c'}) + (\nu_{gs}^c - \nu_{gs}^{c'})
\]

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 ($\theta$), in fixed student and family factors ($\gamma$), and an average error component that includes not only measurement errors but time varying individual, family, and school factors.

\textsuperscript{9}This IV estimator assumes that there are not strong complementarities between specific students and teachers, i.e., 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.
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 \( \gamma \) 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 \( g' \). For ease of presentation, we consider only students who remain in the same school for grades \( g-1 \) and \( g \).

\[
\begin{align*}
(5) \quad &\left( \Delta \theta^c_{gs} - \Delta \theta^c_{g's} \right) - \left( \Delta \theta^c_{gs} - \Delta \theta^c_{g's} \right) = \\
&\left[ \left( \bar{\theta}^c_{gs} - \bar{\theta}^c_{g's} \right) + \left( \bar{\theta}^c_{gs} - \bar{\theta}^c_{g's} \right) \right] \\
&+ \left[ \left( \bar{\nu}^c_{gs} - \bar{\nu}^c_{g's} \right) - \left( \bar{\nu}^c_{gs} - \bar{\nu}^c_{g's} \right) \right]
\end{align*}
\]

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 \( \nu \)) as determinants of the difference in the pattern of achievement gains.

Squaring both sides of equation (5) gives:

\[
(6) \quad \left[ \left( \Delta \theta^c_{gs} - \Delta \theta^c_{g's} \right) - \left( \Delta \theta^c_{gs} - \Delta \theta^c_{g's} \right) \right]^2 = \bar{\theta}^c_{gs}^2 + \bar{\theta}^c_{g's}^2 + \bar{\theta}^c_{gs}^2 + \bar{\theta}^c_{g's}^2 \\
-2 \left( \bar{\theta}^c_{gs} \bar{\theta}^c_{g's} + \bar{\theta}^c_{g's} \bar{\theta}^c_{gs} \right) \\
+2 \left[ \left( \bar{\theta}^c_{gs} \bar{\theta}^c_{g's} - \bar{\theta}^c_{gs} \bar{\theta}^c_{g's} \right) + \left( \bar{\theta}^c_{gs} \bar{\theta}^c_{g's} - \bar{\theta}^c_{gs} \bar{\theta}^c_{g's} \right) \right] \\
+ e
\]

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 \( \theta \) that includes all random error and cross product terms between teacher and other grade specific effects.
We now impose three straightforward assumptions that formally characterize the notion that teachers are drawn from common distributions over the restricted time period of our cohort and grade observations: 1) The variance of grade average teacher quality is the same for all cohorts and grades; 2) The covariance of grade average teacher quality for adjacent cohorts is the same for all grades; and, 3) 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:

\[
(7) \quad E \left[ \left( \Delta \theta^c_{g,s} - \Delta \theta^{c'}_{g,s'} \right) - \left( \Delta \theta^c_{g,s} - \Delta \theta^{c'}_{g,s'} \right) \right]^2 = 4(\sigma^2_{\theta_s} - \sigma_{\theta_s,\theta_s'}) + E(e_s)
\]

where \( \sigma^2_{\theta_s} \) is the variance of teacher quality in school s and \( \sigma_{\theta_s,\theta_s'} \) 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., proportion of teachers who are different equals 0). As long as each teacher performs equally well in both years, \( \sigma^2_{\theta_s} = \sigma_{\theta_s,\theta_s'} \), and teacher quality contributes nothing to student performance differences across cohorts.

On the other hand, consider schools in which cohorts c and c' have all different teachers (the proportion of teachers who are different equals 1). In this case the within school covariance of teacher

---

10 In some cases teachers may simply switch grades within the same school, but the majority of turnover results from teacher exits. 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 (forthcoming-b)). None, however, suggests that leavers are systematically more effective teachers than
quality equals zero. More generally, as the proportion different declines from one to zero, the difference between the variance and covariance terms decreases in proportion to the share of teachers who are different.

Equation (7) provides the basis for estimation of the within-school variance of teacher quality. The left-hand side in our subsequent 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.\textsuperscript{11}

One immediate 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 the estimated parameter 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., $\frac{4\sigma^2_{\theta_i}}{3}$) 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 100 percent turnover in schools with 6 teachers per grade. In order to

\textsuperscript{11}Note that we use teacher turnover as a method of identifying the variance in teacher quality. Implicitly, we are assuming that the level of teacher turnover per se does not directly affect student achievement gains. We test this assumption within the general production function estimation (below) and cannot reject it.
account for such differences in the number of teachers and place all schools on a common metric, the proportion different must be divided by the number of teachers per grade.

The coefficient on the variable “proportion different divided by number of teachers” provides an estimate of the within-school variance in teacher quality, but there is strong reason to believe that the coefficient on turnover will underestimate the actual within-school variance in teacher quality. This follows from the 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, etc., reducing the expected coefficient on the turnover variable below the within-school variance. For this reason, we interpret this estimate as a lower bound on the within-school variance in quality.

Of course teacher turnover may also be precipitated or accompanied by other changes such as a new principal or superintendent or district induced curriculum changes (Ingersoll (2001)). These concomitant changes could, depending on their correlation with teacher turnover, bias the estimates of the variance in teacher quality. For example, if administrator turnover also leads to teacher turnover, any direct effects of new administrators on achievement growth could 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 explicitly incorporating whether or not the principal or superintendent changes over the observation period.

Finally, this framework 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. In any case, any 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.

III. 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). For each cohort there are more than 200,000 students in over 4,000 public elementary and middle schools. (For details on the database, see Appendix A and Appendix Table A1). 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 underlying 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 three through eight. These criterion referenced tests

---

12 Hanushek, Kain and Rivkin (forthcoming-b) find that teachers who switch schools tend to move to schools with higher achieving, higher income, and lower proportion minority student bodies.

13 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 15 percent of students do not take the tests, either because of an exemption or because of repeated absences on testing days. This
evaluate student mastery of grade-specific subject matter that is prescribed for students in the state. We focus on test results for mathematics and reading, which each contains approximately 50 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 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 scores obtained 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.\footnote{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.}

Student data are merged with information on teachers using unique school-grade identifiers. Because student and teacher data come from different reporting systems that are not directly linked, matching individual students with their teachers is not possible, but matching by school and grade is 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 used in the second part of the analysis.

IV. Lower Bound Estimates of the Importance 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. While we refine the estimation below, the pattern can be seen directly by observing the higher correlations in

rate of missing tests appears comparable to those for other high quality testing programs such as the National Assessment of Educational Progress.
student achievement across cohorts for schools with lower teacher turnover. Figures 1 and 2 plot the differences in mathematics and reading average gains in grades g and grade g-1 for adjacent cohorts. Each figure reports two plots, one for high teacher turnover schools (fewer than 25 percent teachers are the same) and the other for low teacher turnover schools (fewer than 25 percent of teachers are different). There are 316 high turnover and 1050 low turnover schools based on this definition.

Both the mathematics and reading figures reveal a much more positive correlation in gains across cohorts for lower teacher turnover schools. In fact 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.

Table 3 reports the basic results from the regression of the squared between-cohort difference in gains on the proportion of teachers who are different. The sample includes only students who remain in the same school for two successive grades, either 5th and 6th or 6th and 7th, only grades that have at least five students with valid test scores and nonmissing data on teacher turnover, and only schools with information on all three cohorts. Only grades five and six are used for the small number of schools with all three grades. The final sample has 1,512 schools in the mathematics specifications and 1,490 in the reading specifications, each of which contributes two observations on between-cohort differences.

Table 3 reports three specifications for both 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

---

15 The majority of students move from elementary to middle school at sometime between grades 5 and 7. Roughly 15 percent of schools with at least two of the three grades in this range have all three. The hypothesis that the effect of teacher turnover is the same across included grades cannot be rejected at any conventional level.
Figure 1. Plot of Grade Pattern of Mathematics Achievement Gains for Two Adjacent Cohorts by Teachers Turnover (hiturn==0 is 25 percent or fewer new teachers across cohorts; hiturn==1 is 75 percent or more new teachers across cohorts)
Figure 2. Plot of Grade Pattern of Reading Achievement Gains for Two Adjacent Cohorts by Teachers Turnover (hiturn==0 is 25 percent or fewer new teachers across cohorts; hiturn==1 is 75 percent or more new teachers across cohorts)
Table 3. Effect of Teacher Turnover on the Divergence of Mathematics and Reading Test Score Gains Between Cohorts  (absolute value of t-statistics in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>no fixed effects&lt;sup&gt;a&lt;/sup&gt;</th>
<th>individual and school fixed effects&lt;sup&gt;b&lt;/sup&gt;</th>
<th>individual and school-by-grade fixed effects&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion different</td>
<td>0.041</td>
<td>0.061</td>
<td>0.060</td>
</tr>
<tr>
<td>math teachers/number of teachers</td>
<td>(3.82)</td>
<td>(5.96)</td>
<td>(3.83)</td>
</tr>
<tr>
<td><strong>2. Reading</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion different</td>
<td>0.053</td>
<td>0.054</td>
<td>0.019</td>
</tr>
<tr>
<td>English teachers/number of teachers</td>
<td>(6.50)</td>
<td>(5.17)</td>
<td>(1.36)</td>
</tr>
</tbody>
</table>

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. Sample size is 3024 for the mathematics and 2980 for the reading specifications. The basic estimation form:

Equations have the same form for mathematics and for reading and for the analysis of gain patterns between grade 6 and 7. For \( \Theta \)=proportion different math teachers/#teachers:

\[
a. \quad (A_5^c - \bar{A}_5^c)^2 = \beta_\theta \Theta_{6s}^{c,c'c'} + \beta_{X} X_{6s}^{c,c'} + e_{6s}^{c,c'}
\]

\[
b. \quad [(A_6^c - \bar{A}_6^c) - (A_5^c - \bar{A}_5^c)]^2 = \beta_\theta \Theta_{5and6s}^{c,c'c'} + \beta_{X} X_{5and6s}^{c,c'} + e_{5and6s}^{c,c'}
\]

\[
c. \quad [(A_6^c - \bar{A}_6^c) - (A_6^c - \bar{A}_5^c)]^2 = \beta_\theta \Theta_{5and6s}^{c,c'c'} + \sum \omega_x + \beta_{X} X_{5and6s}^{c,c'} + e_{5and6s}^{c,c'}
\]
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 latter estimation directly controls for school specific unobservables 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 seventh 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 achievement gains among cohorts are strongly related to teacher turnover, while the results for reading are qualitatively similar but not quite as strong. The coefficients on proportion of math teachers who are different are all positive and significant at the one percent level. Except for the school-by-grade fixed effect specifications, all t-statistics exceed 3.80 in absolute value with a sample of 3,024 observations. Our preferred estimates (column 3) consider both individual and school-by-grade fixed effects in the divergence of cohort gains and yield a teacher turnover coefficient of 0.060. This estimate implies a lower bound estimate of the variance of teacher quality (measured in units of student achievement) equal to 0.015 (0.060/4). A variance of 0.015 means that a one standard deviation increase in average teacher quality for a grade raises average student achievement in the grade by at least 0.12 standard deviations of the total test score distribution. We discuss the magnitude of this effect below following consideration of potential problems with the estimates.

The coefficients for reading are quite similar to those for mathematics in magnitude and significance in the first two specifications but somewhat smaller and less precisely estimated in the
Teacher turnover in the untested subject is used to identify any concomitant disruption in the school. Thus, the number of teachers will not directly affect the variance in student performance, leading to the different measurement of the turnover variable for the tested and untested subject.

Specifically, the t-statistic exceeds 5 in value for the first two specifications but equals 1.36 in the full fixed effects model. One possibility is that there is less variation in teaching effectiveness in reading, but an alternative explanation is that there is not enough variation in reading achievement gains once all fixed effects are accounted for. In fact, though the variance of the dependent variable is quite similar for mathematics and reading for the first two specifications, the residual variance in reading is 11 percent smaller than in mathematics following the removal of school fixed effects in Column 3.

The main 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. Turnover in the nontested subject should capture the influences of any confounding factors related to teacher turnover in general. In this, only the difference in turnover between the tested and nontested subjects is used to identify the variance in teacher quality.\(^\text{16}\)

Table 4 is generated by the smaller sub-sample of schools with subject specialists (defined as schools that have no teachers in either 5th and 6th grade who teach both math and English). This sample is slightly more than 400 schools or roughly 30 percent of the full sample. Not surprisingly, the smaller sample size in combination with the inclusion of the other turnover variable reduces the precision of the estimates, particularly in the specifications that remove school fixed effects in the between-cohort difference in the pattern of gains. Nevertheless, the results remain similar to those reported in Table 3.

---

\(^{16}\)Teacher turnover in the untested subject is used to identify any concomitant disruption in the school. Thus, the number of teachers will not directly affect the variance in student performance, leading to the different measurement of the turnover variable for the tested and untested subject.
Table 4. Effect of Teacher Turnover on the Divergence of Mathematics and Reading Test Score Gains Between Cohorts, controlling for Teacher Turnover in other subjects$^a$
(absolute value of t-statistics in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>no fixed effects$^b$</th>
<th>individual fixed effects$^b$</th>
<th>individual school fixed effects$^b$</th>
<th>individual school-by-grade fixed effects$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion different math teachers/number of teachers</td>
<td>0.020</td>
<td>0.052</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>% English teachers same</td>
<td>-0.009</td>
<td>-0.017</td>
<td>-0.002</td>
<td></td>
</tr>
<tr>
<td>2. Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion different English teachers/number of teachers</td>
<td>0.007</td>
<td>0.041</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>% Mathematics teachers same</td>
<td>-0.007</td>
<td>-0.008</td>
<td>-0.022</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

a. The sample includes all 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 891.

b. See the estimation forms in the notes to Table 3.
All coefficients on proportion different math teachers are positive, statistically significant at the five percent level and similar in magnitude to those in Table 3, while the results for reading again are somewhat weaker, particularly in the final specification. In contrast, the hypothesis that teacher turnover in the untested subject is not positively related to the variation in achievement gains cannot be rejected at any conventional level in any specification; all of the coefficients are negative and small in magnitude. These results provide further support for that the Table 3 estimates of the within-school variance in teacher quality are not contaminated by changes in other factors.

A final robustness check examines only schools with a single teacher per grade. This quite select set of roughly 150 schools generates large, positive, and statistically significant estimates in both mathematics and reading for the first two specifications (see Appendix Table A2). Not surprisingly given the small sample sizes, the estimates for the full fixed effect specification are quite imprecise. Nevertheless, the mathematics coefficient is quite similar in magnitude and the reading coefficient is much larger than those in Tables 3 and 4.

Importantly, the true magnitudes of the variances in mathematics and reading teacher quality are likely to be far greater than the estimates presented here. First, as noted above, the identifying assumptions are likely to be violated in a way that biases 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 which would attenuate the coefficients.17

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

17The number of teachers is the average number in both cohorts for the grades included in each specification. The administrative data contain some ambiguities as to whether a teacher actually taught in a given year, likely introducing some error. For example, some teachers with class size coded as zero may actually have taught. Similarly, ambiguity is introduced when class size is missing.
teacher quality is an important determinant of readings and mathematics achievement.

V. 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 impact of 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. 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 or distort true causal impacts. To motivate our estimation, we highlight recent instrumental variables and random assignment methods used to estimate the effects of class size and then propose an alternative based on the comprehensive model developed above.

The search for better identification methods for resource effects led the State of Tennessee in the mid1980s to undertake a unique random assignment class size experiment, Project STAR. This search has also produced a pair of innovative papers that pursue a similar approach for obtaining estimates of “true” class size effects. We discuss the results and potential problems of the Tennessee

---

STAR experiment below, but first we focus on the two nonexperimental studies that are much closer in focus to our own work. Hoxby (2000) uses demographic changes to isolate exogenous class size variation and concludes that class size reductions do not improve student outcomes. Angrist and Lavy (1999) utilize a class size rule in Israeli law to identify class size effects, but contrary to Hoxby conclude that smaller classes systematically produce higher achievement. Despite the theoretical appeal of the methodologies used in these papers, the studies fail to produce a consistent set of findings. More important for this work, each potentially suffers from problems that raise questions about the consistency of the parameter estimates.

Hoxby (2000) used detrended variations in the size of birth cohorts to identify exogenous changes in class size in small Connecticut towns. Changes in cohort sizes, coupled with the lumpiness of classes in small school districts, can provide variations in class size that are unrelated to other factors. This promising methodology, however, suffers in application from having only limited aggregate performance data available. The study regresses school average scores for tests administered in the fall on class sizes for the same school cohort for previous grades, but it cannot separate recent student movers (for whom these class sizes are not relevant) from students who attended previous grades in the given school district. The assignment of incorrect class sizes for students who moved during the period of the study attenuates the estimated class size effects, biasing them toward zero.

Angrist and Lavy (1999) note that Maimonides’ Rule requires that Israeli classes cannot exceed 40 students, so that the lumpiness of classrooms may lead to large changes in class size when the number of students in a school approaches multiples of 40 (and the preferred class size is greater than 40). They use a regression discontinuity approach to identify the effects of class size, but many of their estimates also use class size variation other than that generated by the discontinuities. Unfortunately, the absence of large numbers of schools with enrollments around the points of discontinuity impedes the analysis, and Hoxby (2000) demonstrates the sensitivity to the definition of the size of regions around the discontinuity points. The fact that the results are driven by the very small schools in the
sample and are not consistent across grades raises further questions about their generalizability.

Both of these studies attempt to identify factors leading to exogenous variation in class size. 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 II, our approach 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.

**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 equations (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).

\[
\Delta y_{ig} = c + \lambda g + \beta x_{ig} + \gamma y_{ig} + \delta_{xy} + \omega g_{ig} + \upsilon_{ig} + \epsilon_{ig}
\]

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. The requirements for these instrumental approaches actually interact with the more general issues because the instruments must be uncorrelated with all of the selection factors and any omitted inputs to achievement. Thus, for example, no unmeasured elements of teacher quality or school selection by families can be correlated with the distribution of schools near different discontinuity points.

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
the percentage of teachers with a master’s degree, and the percentage of teachers who 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). 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 (\( \delta_{xy} \)) and school-by-grade fixed effects (\( \omega_{sg} \)). Note that 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 A 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

---

21 Including the percentages of teachers with 5-9 and 20 or more years of experience as separate categories did not change any of the results, and the hypotheses that teachers with 5-9 or 20 or more years of experience had a different impact from those with 10 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.
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 $\Delta A$ and the estimation of the individual $\gamma_i$’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 ($\delta_s$) 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 ($\delta_{sy}$) rather than simple school fixed effects in some 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 school will be eliminated, and the estimates are identified solely by within school by year differences across grades.22

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),23 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...

---

22Less substantively, we also allow for changes in the tests over time through inclusion of a fixed effect for year for each subject-grade test ($\tau_{gy}$).

23Note that the estimation explicitly controls for the effects of moving on the moving students’ achievement growth; see Hanushek, Kain and Rivkin (forthcoming-a).
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.

Impact of teacher and school characteristics

Table 5 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. Based on preliminary findings, class size effects are further allowed to differ by grade. The absolute values of Huber-White t-statistics adjusted for the clustering of students into schools are reported for all coefficients. Appendix Table A1 presents descriptive statistics for the school characteristics and achievement gain.

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

---

24 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 fourth 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.

25 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.
Table 5. Effects of Teacher Characteristics on 4th-6th Grade Gains in Mathematics and Reading Test Scores  
(absolute value of Huber-White adjusted t-statistics in parentheses, n=1,337,426 for mathematics and 1,334,853 for reading)

<table>
<thead>
<tr>
<th></th>
<th>no fixed effects</th>
<th>student and school fixed effects</th>
<th>student and school-by-year fixed effects</th>
<th>student, school by grade and school-by-year fixed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th grade</td>
<td>-0.0047</td>
<td>-0.0105</td>
<td>-0.0106</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>(2.01)</td>
<td>(3.44)</td>
<td>(4.17)</td>
<td></td>
</tr>
<tr>
<td>5th grade</td>
<td>-0.0037</td>
<td>-0.0085</td>
<td>-0.0081</td>
<td>-0.0050</td>
</tr>
<tr>
<td></td>
<td>(4.49)</td>
<td>(6.91)</td>
<td>(5.33)</td>
<td>(3.53)</td>
</tr>
<tr>
<td>6th grade</td>
<td>-0.0001</td>
<td>-0.0030</td>
<td>-0.0037</td>
<td>-0.0022</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(2.51)</td>
<td>(3.03)</td>
<td>(2.21)</td>
</tr>
<tr>
<td>7th grade</td>
<td>0.0008</td>
<td>0.0028</td>
<td>0.0034</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(1.98)</td>
<td>(2.25)</td>
<td>(1.19)</td>
</tr>
<tr>
<td>Experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% 0 years</td>
<td>-0.08</td>
<td>-0.10</td>
<td>-0.13</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(7.24)</td>
<td>(6.44)</td>
<td>(7.34)</td>
<td>(4.36)</td>
</tr>
<tr>
<td>% 1 year</td>
<td>-0.04</td>
<td>-0.06</td>
<td>-0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(3.07)</td>
<td>(3.82)</td>
<td>(2.74)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>% 2 years</td>
<td>-0.02</td>
<td>-0.04</td>
<td>-0.06</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(1.42)</td>
<td>(2.80)</td>
<td>(3.02)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>% 3-5 years</td>
<td>-0.01</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(1.03)</td>
<td>(2.27)</td>
<td>(1.75)</td>
<td>(0.98)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% with graduate degree</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(2.69)</td>
<td>(1.58)</td>
<td>(1.80)</td>
<td>(1.54)</td>
</tr>
</tbody>
</table>
### Table 5. Continued

<table>
<thead>
<tr>
<th></th>
<th>no fixed effects</th>
<th>student and school fixed effects</th>
<th>student and school-by-year fixed effects</th>
<th>student, school by grade and school-by-year fixed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2. Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>class size</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th grade</td>
<td>-0.0031</td>
<td>-0.0091</td>
<td>-0.0089</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>(1.69)</td>
<td>(3.80)</td>
<td>(4.35)</td>
<td></td>
</tr>
<tr>
<td>5th grade</td>
<td>-0.0001</td>
<td>-0.0033</td>
<td>-0.0035</td>
<td>-0.0036</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(3.44)</td>
<td>(3.03)</td>
<td>(2.75)</td>
</tr>
<tr>
<td>6th grade</td>
<td>0.0022</td>
<td>0.0002</td>
<td>-0.0010</td>
<td>-0.0024</td>
</tr>
<tr>
<td></td>
<td>(2.97)</td>
<td>(0.16)</td>
<td>(0.90)</td>
<td>(2.44)</td>
</tr>
<tr>
<td>7th grade</td>
<td>-0.0046</td>
<td>-0.0016</td>
<td>-0.0031</td>
<td>-0.0022</td>
</tr>
<tr>
<td></td>
<td>(5.70)</td>
<td>(1.19)</td>
<td>(2.04)</td>
<td>(1.42)</td>
</tr>
<tr>
<td><strong>Experience</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% 0 years</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.06</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(3.83)</td>
<td>(3.01)</td>
<td>(4.14)</td>
<td>(1.28)</td>
</tr>
<tr>
<td>% 1 year</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.06</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(3.87)</td>
<td>(2.86)</td>
<td>(4.03)</td>
<td>(0.75)</td>
</tr>
<tr>
<td>% 2 years</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.03</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(1.12)</td>
<td>(1.31)</td>
<td>(1.60)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>% 3-5 years</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.54)</td>
<td>(0.77)</td>
<td>(2.00)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% with graduate degree</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(2.48)</td>
<td>(0.75)</td>
<td>(0.36)</td>
<td>(0.98)</td>
</tr>
</tbody>
</table>

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.
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.\textsuperscript{26} Both the fourth and fifth grade class size coefficients are highly significant in both subjects, though the magnitude of the fifth grade effect is roughly three fourths as large as that for fourth grade in mathematics and less than half as large in reading. The sixth 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 0.003 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 triples the 4th grade coefficient and more than doubles the coefficient for 5th grade. Finally, the smaller though still significant coefficients in the full fixed effects model are consistent with the possibility that the school-by-grade and school-by-year fixed effects together aggravate problems associated with measurement error.

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 free lunch eligibility). The results in Table 6 generally do not support the belief that class size effects are substantially larger for disadvantaged 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 5.

One potential perspective on these estimates comes from Project STAR, the random assignment

\textsuperscript{26}However, the addition of school-by-grade fixed effects substantially reduces the magnitudes and significance levels of most estimates. Nevertheless, class size continues to exert a significant effect on mathematics and reading achievement in grades 5 and 6. It is 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 6. Effects of Class Size on Test Score Gains, by Family Income  (absolute value of Huber-White adjusted t statistics in parentheses)

| class size | Mathematics | | Reading | |
|------------|------------|------------|------------|
|            | disadvantaged students | not disadvantaged students | disadvantaged students | not disadvantaged students |
| 4th grade  | -0.0118 | -0.0102 | -0.0106 | -0.0086 |
|            | (4.49)  | (4.01)    | (5.01)    | (4.18)    |
| 5th grade  | -0.0077 | -0.0080 | -0.0027 | -0.0037 |
|            | (4.89)  | (5.23)    | (2.12)    | (3.17)    |
| 6th grade  | -0.0040 | -0.0036 | -0.0026 | -0.0014 |
|            | (3.11)  | (2.99)    | (2.15)    | (1.26)    |
| 7th grade  | 0.0038  | 0.0033 | 0.0013 | -0.0040 |
|            | (2.36)  | (2.23)    | (0.83)    | (2.86)    |

Note: Estimates come from a model that includes student and school-by-year fixed effects. Specifications also include percentage of teachers with a graduate degree, full sets of grade-by-year and teacher experience dummies, and indicators for subsidized lunch eligibility and a change of school prior to or during year.
experiment in class size reduction conducted in Tennessee (Word, et al. (1990)). While these experimental results are not directly comparable because they consider just grades K to 3, they indicate that a reduction 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 first grade – and not subsequent small class treatments (Krueger 1999), while we find that smaller classes still have an effect in 4th and 5th grade.

It is also noteworthy that students in the small classes outperformed schoolmates in larger classes in almost half of the schools involved in the STAR experiment. This experimental finding is consistent with the conclusions here that differences in teacher quality are very important and sufficiently large to offset the benefits of even sizeable reductions in class size.

The school-by-year fixed effect estimates in Column 3 of Table 5 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 roughly 0.12 standard deviations higher annual growth in mathematics achievement (and perhaps somewhat less than that in reading). This magnitude of change is equivalent to a class size reduction of approximately eleven to twelve students in fourth grade, fourteen or more students in fifth grade, and an implausibly large number of students in the higher grades. Note that these comparisons assume both no accompanying changes in teacher quality and linearity in class size effects. Linearity appears reasonable from nonparametric estimates for class sizes.

---

27Project 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 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).
between 10 and 35 students (results not reported).

Teacher Characteristics

The results for teacher experience generally support the notion that rookie 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 than in the other fixed effects models, consistent with the view that multiple fixed effects can exacerbate problems with measurement error even with such large data sets. 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 (forthcoming-b)). The estimated parameters in Table 5 combine the effects of on-the-job learning and of selective exit and mobility.

Table 7 presents the basic estimates of first year teaching on achievement for samples that exclude those who immediately leave teaching or switch districts (with individual and school fixed effects). The close similarity of the estimates compared to those in Table 5 for both mathematics and reading indicates that on-the-job learning is the dominant element of the experience effect. Importantly, these results also indicate that the average quality of those who quit teaching after one year is similar to the average quality of those who remain.
Table 7. Effects of Teacher Experience on Mathematics and Reading Test Score Gains, by New Teacher Transitions  
(absolute value of Huber-White adjusted t-statistics in parentheses)

<table>
<thead>
<tr>
<th>% 0 years experience</th>
<th>excluding teachers who exit or switch schools</th>
<th>excluding teachers who exit teaching</th>
<th>all teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>math</td>
<td>-0.10</td>
<td>-0.11</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>(4.54)</td>
<td>(5.27)</td>
<td>(6.44)</td>
</tr>
<tr>
<td>Observations</td>
<td>[1,185,928]</td>
<td>[1,210,734]</td>
<td>[1,337,426]</td>
</tr>
<tr>
<td>reading</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(1.94)</td>
<td>(2.04)</td>
<td>(3.01)</td>
</tr>
<tr>
<td>Observations</td>
<td>[1,184,343]</td>
<td>[1,208,997]</td>
<td>[1,334,853]</td>
</tr>
</tbody>
</table>

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.

V. 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.

1. Teachers and therefore schools matter importantly for student achievement.

The issue of whether or not there is significant variation in the education 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. The cumulative gain from having five years of teachers who are one standard deviation above the mean quality (compared to having all average teachers) would virtually eliminate the average 7th grade achievement gap between students from low income and those from non-low income families.

2. 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 resource effects by family income are small.

Partially consistent with recent experimental and statistical efforts to identify class size effects, we find that class size has a positive effect on mathematics and reading achievement, though the magnitude of the effect is small, particularly following fifth 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, more than just teacher salaries will be relevant if class size reduction also 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 elasticity of 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)).

3. 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, 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 existing personnel practices in the Texas public schools leave much to be desired.

The appropriate policy changes remain 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 (forthcoming)). 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.
Appendix A. 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) [http://www.utdallas.edu/research/greenctr/]. 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 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.

Beginning in 1993, the Texas Assessment of Academic Skills (TAAS) was administered each spring to eligible students enrolled in grades three through eight. We focus on test results for mathematics. The bottom one percent of test scores and the top and bottom one percent of test score gains are trimmed from the sample in order to reduce measurement error. Participants in bilingual or special education programs are also excluded from the sample because of the difficulty in measuring school characteristics for these students.

Student data are merged with information on teachers using unique school-grade 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 corrected experience for those years. If experience data were inconsistent for all the years, if there were two consistent patterns, or if correction would have imputed negative years of experience, no corrections were made. In any case, no teachers were excluded from the final sample on the basis of inconsistent experience data, though the results were 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 fourth 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 remained the same.

We attempt to reduce problems introduced by measurement in a number of ways. First, all reported class sizes that fall below 10 or above 25 in 4th grade (35 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 22 students, though some 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 35 students during this period. Next, we exclude a school’s observations for the grades in years in which: 1) the school is in the top one percent in terms of within grade differential in reported class size; and 2) the year to year change in average class size exceeded fifteen students. Estimates of class size effects increased in magnitude following these exclusions, suggesting that class size was measured with substantial error for these schools.
References


Appendix Table A1. Variable Means and Standard Deviations

<table>
<thead>
<tr>
<th>grade</th>
<th>math test score gain</th>
<th>reading test score gain</th>
<th>class size</th>
<th>% with graduate degree</th>
<th>% 0 years experience</th>
<th>% 1 year experience</th>
<th>observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th grade</td>
<td>-0.01</td>
<td>0.02</td>
<td>19.5</td>
<td>23.8</td>
<td>6.0</td>
<td>5.9</td>
<td>142,833</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.73)</td>
<td>(2.4)</td>
<td>(24.6)</td>
<td>(12.4)</td>
<td>(12.6)</td>
<td></td>
</tr>
<tr>
<td>5th grade</td>
<td>0.01</td>
<td>0.01</td>
<td>22.5</td>
<td>25.3</td>
<td>5.8</td>
<td>6.0</td>
<td>438,536</td>
</tr>
<tr>
<td></td>
<td>(0.64)</td>
<td>(0.68)</td>
<td>(3.6)</td>
<td>(26.7)</td>
<td>(13.7)</td>
<td>(13.6)</td>
<td></td>
</tr>
<tr>
<td>6th grade</td>
<td>0.02</td>
<td>0.02</td>
<td>22.2</td>
<td>24.6</td>
<td>7.4</td>
<td>6.9</td>
<td>456,400</td>
</tr>
<tr>
<td></td>
<td>(0.61)</td>
<td>(0.68)</td>
<td>(3.9)</td>
<td>(27.8)</td>
<td>(16.7)</td>
<td>(15.9)</td>
<td></td>
</tr>
<tr>
<td>7th grade</td>
<td>-0.02</td>
<td>-0.01</td>
<td>21.5</td>
<td>21.9</td>
<td>9.2</td>
<td>9.0</td>
<td>299,657</td>
</tr>
<tr>
<td></td>
<td>(0.55)</td>
<td>(0.66)</td>
<td>(4.1)</td>
<td>(26.8)</td>
<td>(18.4)</td>
<td>(18.0)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix Table A2. Effect of Teacher Turnover on the Divergence of Gains in Mathematics and Reading Test Scores Between Cohorts for Schools with One Teacher per Grade (absolute value of t-statistics in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>no fixed effects</th>
<th>individual and school fixed effects</th>
<th>individual and school-by-grade fixed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mathematics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion different</td>
<td>0.11</td>
<td>0.12</td>
<td>0.044</td>
</tr>
<tr>
<td>math teachers/number</td>
<td>(3.01)</td>
<td>(3.11)</td>
<td>(0.96)</td>
</tr>
<tr>
<td>of teachers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion different</td>
<td>0.17</td>
<td>0.21</td>
<td>0.071</td>
</tr>
<tr>
<td>English teachers/</td>
<td>(4.32)</td>
<td>(3.90)</td>
<td>(1.56)</td>
</tr>
<tr>
<td>number of teachers</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 302 for the mathematics and 266 for the reading specifications. See notes to Table 3 for different estimation forms.