NBER WORKING PAPER SERIES

TEACHERS, SCHOOLS, AND ACADEMIC ACHIEVEMENT

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

Working Paper 6691
http://www.nber.org/papers/w6691

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1998

We are grateful to Kraig Singleton, Jaison George and Dan O'Brien for excellent research assistance, and we thank seminar participants at UC Berkeley, UC Davis, UC San Diego, the Public Policy Research Institute, Stanford University, University of Texas, and Texas A&M University for their many helpful comments. Hanushek and Rivkin thank the Donner Foundation and the Smith Richardson Foundation for funding, and Kain thanks the Smith-Richardson Foundation and the Spencer Foundation. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Teachers, Schools, and Academic Achievement
Eric A. Hanushek, John F. Kain,
and Steven G. Rivkin
NBER Working Paper No. 6691
August 1998
JEL No. I2, H4

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 role of teacher differences and other aspects of schools. Unique matched panel data from the Harvard/UTD Texas Schools Project permit distinguishing between total effects and the impact of specific, measured components of teachers and schools. While schools are seen to have powerful effects on achievement differences, these effects appear to derive most importantly from variations in teacher quality. A lower bound suggests that variations in teacher quality account for at least 7½ percent of the total variation in student achievement, and there are reasons to believe that the true percentage is considerably larger. The subsequent analysis estimates educational production functions based on models of achievement growth with individual fixed effects. It identifies a few systematic factors — a negative impact of initial years of teaching and a positive effect of smaller class sizes for low income children in earlier grades — but these effects are very small relative to the effects of overall teacher quality differences.

Eric A. Hanushek
Wallis Institute of Political Economy
University of Rochester
Rochester, NY 14627
and NBER
hanu@troi.cc.rochester.edu

John F. Kain
Green Ctr. for the Study of Science and Society
University of Texas at Dallas
P.O. Box 830688
Richardson, TX 75083-0688
jkain@utdallas.edu

Steven G. Rivkin
Amherst College
Department of Economics
Amherst, MA 01002
rivkin@acesgr.econ.amherst.edu

Teachers, Schools and Academic Achievement

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

I. Introduction

Since the release of *Equality of Educational Opportunity* (the "Coleman Report") in 1966, the educational policy debate has often been reduced to a series of simplistic arguments and assertions about the role of schools in producing achievement.[1] The character of this debate has itself been heavily influenced by confusing and conflicting research. While this research has suffered from inadequate data, imprecise definitions of the underlying problems and issues have 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 the current policy debate and considers how observed student performance can be used to address each. It then employs a unique panel data set of students in Texas that permits clear identification of the sources of differences in student achievement and of the relevance of a broad class of policies related to school resources. The empirical work, which extends the prior analysis in Kain and Singleton (1996), remains a preliminary investigation of the very complex achievement process found in schools. Nonetheless, findings diverge sharply from many commonly held policy perspectives, raising serious concerns about the current institutional structure of schools.

Some very basic questions that have arisen from prior work command a central position in most policy discussions. First, partly resulting from common interpretations of the Coleman Report, a surprising amount of controversy continues over whether schools "make a difference" or not. This

[1]The original Coleman Report (Coleman et al., 1966) was subjected to considerable criticism both for methodology and interpretation; see, for example, Hanushek and Kain (1972). The ensuing controversy led to considerable new research, but this new work has not ended the controversy; see Hanushek (1996, 1997) and Greenwald et al. (1996). Those discussions represent the starting point for this research.

issue comes down to a simple one of whether or not there are significant differences among schools in their ability to raise achievement. Second, to the extent that there are differences in the impact of schools, is it the teachers, the principals, or the organization of the schools that is the major source of the differences? Third, if in fact schools differ in their impact, are these differences systematically related to school resources or to measurable aspects of schools and teachers? (This third issue is in fact the genesis of the first, because the Coleman Report reported relatively small effects on student achievement of differences in the measured attributes of schools – a finding that has frequently been interpreted as indicating that there are no systematic quality differences among schools). Fourth, is the impact of any systematic differences in resources sufficient to justify policy initiatives designed to provide more of those resources?

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 those questions. The data contain 3rd, 4th, 5th and 6th grade test scores for one cohort of students (4th graders in 1994) and 3rd, 4th and 5th grade test scores for 4th graders in 1993 and 1995. The multiple cohorts and grades permit identification of the importance of specific teachers versus the school as a whole. And, the very large samples, exceeding 3,000 schools and one half million students, remove much of the ambiguity about the effects of observable teacher characteristics such as class size, experience and education. While small effects of these factors could be difficult to detect in many samples previously used, such is not the case here.

The repeated performance observations also provide an excellent means of controlling for student heterogeneity and the nonrandom matching of students, teachers and schools through the use of fixed effects models. 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 to place students into programs and classes. Therefore it is difficult to interpret between school or even between classroom differences in student achievement gains, because such differences confound school or teacher effects with the influence of individual and family factors. Even value-added regression techniques that account for observable differences in student characteristics that might be related to the rate at which students learn are unlikely to account for all relevant factors related to both achievement and the selection of schools and teachers. However, the availability of multiple tests permits a comparison of academic performance in two different grades. Any performance differential between grades cannot be driven by unchanging student attributes such as ability or motivation. Comparisons for students who remain in the same campus in two successive years further isolate the influence of teachers, since peer group and school characteristics are similar for both grades.

The results show large differences among schools in their impact on student achievement. These differences are centered on the differential impact of teachers, rather than the overall school organization, leadership or even financial condition. The differences among teachers are not, however, readily measured by simple characteristics of the teachers and classrooms. Initial experience is important in improving a teacher's impact on learning, and smaller classes appear to improve educational achievement for lower income but not for higher income students. Nonetheless, these variables explain only a small amount of the total observed variation in teacher quality. All of this suggests that policy initiatives must reflect the substantial heterogeneity of teachers, be it differences in effort or skill, if they are to have a significant impact on students.

The next section provides a detailed description of the Texas data on students and teachers. Section III focuses on the persistence of teacher performance over time. It decomposes educational quality differences into between-school and within-school components, focusing on the role of teachers, and places lower bounds on the contribution of teachers to variations in school achievement. Section IV investigates the extent to which observable teacher and classroom characteristics explain achievement differences. These regressions provide information on the impacts of class size, teacher experience and teacher education that is not confounded by student or school characteristics. The final section considers the policy implications of the findings, particularly the importance of resources relative to the overall contribution of teachers.

II. The Texas Database

The data that are used in this paper come from the data development activity of the Harvard/UTD Texas Schools Project, conceived of and directed by John Kain. Working with the Texas Education Agency (TEA), this project has combined a number of different data sources to compile an extensive data set on schools and achievement. Demographic information on students and teachers is taken from the PEIMS (Public Education Information Management System), which is TEA's state-wide 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. For each cohort there are over 200,000 students in over 3,000 public schools (see Appendix Table A1). In comparison to studies that use only a small sample of students from each school, these data provide much more precise estimates of school average test scores and test score gains.

The PEIMS student 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 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 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.[2] The criterion referenced tests

[2]Many special education 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

evaluate student mastery of grade-specific subject matter. We use test results for reading and mathematics, subjects that are examined in all grades. Reading and math tests each contain approximately 50 questions. Because the number of questions and average percent right varies across time and grades, we transform all test results into standardized scores with a mean of zero and variance equal to one. While this transformation eliminates any year-to-year movement in overall performance, it corrects for any perturbations in the specific test form that is employed. Identical means and variances for all tests facilitates the interpretation of the variance decompositions across time and grades, and the empirical findings are robust to a number of transformations including the raw percentage correct. 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.[3]

Student data are merged with information on teachers using unique school-grade identifiers. The data provide information on each teacher for each year in which they teach. 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.

---

tests, either because of an exemption or because of repeated absences on testing days. The third grade tests in 1992 was actually taken in October 1991. This sampling difference from other years leads to little noticeable impact on the empirical results obtained from excluding this set of observations, but, given the sample sizes available, we report just the results that exclude this grade observation.

[3]For an explicit analysis of the achievement of special education students, see Hanushek, Kain, and Rivkin (1998).

Aggregation of teachers within grades has different implications for the subsequent analyses. The initial variance decompositions focus on the amount of variation in student achievement that can be attributed to systematic differences of schools and teachers. For this, the aggregation across teachers in a grade leads to understating the systematic contribution of schools and teachers. The subsequent estimation of production functions is, however, affected quite differently. Average teacher characteristics by subject, grade and year are assigned to each student. Though students cannot be linked directly to their teachers (except in schools with one teacher per grade), this aggregation of regular classroom characteristics should not introduce any bias into the regression estimates. It will increase the standard errors by ignoring any systematic relationship between achievement and teacher characteristics within grades. On the other hand, it overcomes what is possibly the largest form of selection within schools–that which occurs when some parents maneuver their children toward specific, previously identified teachers. This within-grade teacher selection is circumvented by looking at overall grade differences, which is equivalent to an instrumental variable estimator based on grade rather than classroom assignment.

III. The Importance of Schools and Teachers

What is the contribution of schools to student achievement? And, to what extent does any contribution represent factors common to a school building as opposed to specific teachers within the building? These questions are addressed by investigating the patterns of gains in student achievement within and across schools. We begin with a systematic decomposition of the variance of achievement gains and of correlations of school average gains across grades and time. The interpretation of these depends critically on the sources of underlying achievement variation including any nonrandom sorting of students and teachers across schools. While some past work has pursued portions of this, the limitations of previous data required the imposition of extremely strong assumptions to identify the various components of achievement gain. Separating the influences of schools, teachers, and families has been especially vexing – particularly when sorting by families and schools is acknowledged.

Overall Variance Decomposition

Past analyses of educational production functions have looked at both the level and the growth in achievement. 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 and other factors influence performance during, say, a given school year.[4] However, standard value added models do not account for unobservable factors that affect the rate of acquisition of new knowledge, and it is the availability of test score measures for more than two grades that enables us to control for such influences through the use of fixed effects models for achievement growth.

We use a fixed effects, value added framework to investigate mathematics and reading achievement in grades four, five and six. Equation (1) describes test score gain, Δ_{igst}, in a variance components framework: Achievement gain for student i in grade g in school s in year t is a function of a time invariant individual-specific growth component [γ_i], a school quality component that varies across grades (e.g. teacher quality, curriculum, or textbooks) [υ_{gst}], a school quality component that is constant across grades [δ_{st}] (e.g. resources,, school leadership, or institutional structure), and an idiosyncratic random error [ε_{igst}]. The idiosyncratic error includes all measurement error in the tests. The individual-specific component could include individual ability, parental background, or

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

neighborhoods to the extent each affects the growth in achievement; level effects are differenced out in calculating the gain in achievement.

$$\Delta A_{igst} = \gamma_i - \upsilon_{gst} + \delta_{st} + \varepsilon_{igst} \tag{1}$$

Suppressing the subscript t on the variance components, the total variance of test score gains at time t [$Var^T(\Delta A)$] can be written as:

$$Var^{T}(\Delta A) = \sigma^2_{\gamma} + \sigma^2_{\upsilon} + \sigma^2_{\delta} + 2\sigma_{\gamma\upsilon} + 2\sigma_{\gamma\delta} + 2\sigma_{\upsilon\delta} + \sigma^2_{\epsilon} \tag{2}$$

where σ_j^2 is the variance of component j and σ_{jk} is the covariance of components j and k. Similarly, the between school and grade (Var^{SG}) and between school (Var^S) components of variance can be written

$$Var^{SG}(\Delta A) = \sigma^2_{\gamma_{bgs}} + \sigma^2_{\upsilon} + \sigma^2_{\delta} + 2\sigma_{\gamma\upsilon} + 2\sigma_{\gamma\delta} + 2\sigma_{\upsilon\delta} + \sigma^2_{\epsilon_{bgs}} \tag{3}$$

$$Var^{S}(\Delta A) = \sigma^2_{\gamma_{bs}} + \sigma^2_{\upsilon_{bs}} + \sigma^2_{\delta} + 2\sigma_{\gamma\upsilon} + 2\sigma_{\gamma\delta} + 2\sigma_{\upsilon\delta} + \sigma^2_{\epsilon_{bs}} \tag{4}$$

where the subscript bgs refers to variation between grades and schools and the subscript bs refers to variation between schools. The nonrandom sorting of both students and teachers is made explicit in equations (2)-(4) by the three covariance terms and by allowing the between-school variance in γ and υ to differ from zero.

Many studies use the between school variation as a percentage of the total, Var^S / Var^T, to measure the contribution of school quality to achievement. This ratio is not, however, a clear indication of the possibilities for policy manipulation, because it represents a combination of technology, behavioral adjustments, and the sampling of achievement. More extensive sorting by teacher quality and greater variation in other factors that affect achievement (e.g. resources) will tend to increase the between school variance as a percentage of the total, while a policy of equalization of schools would drive this percentage toward zero. Yet as equations (3) and (4) show, this ratio may also either understate or overstate the true contribution of schools. By attributing all within school variation in achievement gains to non-school factors as is commonly done, this ratio may seriously understate the contribution of schooling. This problem is especially serious if the variation of teacher quality and other inputs within schools is large relative to the between school variation in school quality. This ratio also ignores the error in measuring test scores which is likely to be reduced markedly by aggregation to school average scores, again tending to reduce the observed contribution of schools when measured as a percentage of the total variance in achievement. On the other hand, the often noted problem of nonrandom selection of students into schools inflates the between school component of variance and introduces an upward bias into the calculation of the contribution of schooling.

Measurement error and Tiebout sorting make the identification of the total contribution of schools an extraordinarily difficult task, and in this paper we focus on two specific components of school quality: 1) the within school variation in teacher quality across grades; and 2) the variation in school quality attributed to differences in class size and other measurable school resources.[5] Equation (5) considers the variation in performance arising from grade difference that go beyond any overall differences across schools. The between-grade-within-school variance component in achievement

[5]In subsequent work we intend to use the ability to follow mobility across schools in order to learn about differences between schools in quality and about the motives for moving.

growth depends solely upon the magnitude of within school variations in teacher quality, in other school inputs (υ), in student quality, and in the random error within schools.[6]

$$(5) \qquad \mathrm{Var}^{SG}(\Delta A) - \mathrm{Var}^{S}(\Delta A) = (\sigma^2_{\upsilon_{bgs}} - \sigma^2_{\upsilon_{bs}}) + (\sigma^2_{\gamma_{bgs}} - \sigma^2_{\gamma_{bs}}) + (\sigma^2_{\epsilon_{bgs}} - \sigma^2_{\epsilon_{bs}})$$

Across grades within individual schools, the influence of nonrandom student sorting is likely to be minor and aggregation by grade will reduce the problem of measurement error. Equation (5) does not completely isolate the within school variation in teacher quality per se, because υ includes all systematic aspects of between grade differences. Nonetheless, the within-school-between-grade variations of other school inputs and average student characteristics are likely to be small in comparison to the between school variations. More importantly, comparisons across cohorts at different grades and the removal of individual fixed effects can effectively isolate the contribution of within school differences in teacher quality to the total variation in achievement.

Table 1 reports the variance decomposition of math and reading test score gains for all 4th, 5th and 6th graders in Texas in the 1995 school year.[7] The between school variance as a percentage of the total is 5.5 percent for math and 3.3 percent for reading, while the contribution of the between school

---

[6]We assume that there is no within school correlation between the grade average qualities of students and school inputs including teachers.

[7]A parallel decomposition of the levels of test scores, instead of the value-added, permits comparison with some existing estimates of educational production functions. These calculations (not shown) indicate that the proportion of variance between schools and grades (years) is very close to that for the value-added form, but that a higher proportion is found between as opposed to within schools. The nature of this analysis, however, makes interpretation difficult; the random component includes historical educational inputs, individual differences in the level of performance, and the covariance of these. The complexity, including sorting of students according to these factors, implies no easy conclusions about how this decomposition might be biased from simple school and teacher effects.

Table 1. Variance Decomposition of Math and Reading Test Score Gains: 4th, 5th and 6th Grade Students in 1995

Math			Reading		
Total variance	Between school as a % of total	Between school and grade as a % of total	Total variance	Between school as a % of total	Between school and grade as a % of total
0.39	5.5%	15.3%	0.44	3.3%	8.9%

and grade variance rises to 15.3 percent for math and 8.9 percent for reading. Given that the within school variation in cohort average characteristics is likely to be small, this pattern of results makes a prima facie case that there are substantial within school differences in schools, likely arising from differences in teacher quality as well as other school inputs. Consequently, studies that use the between school variance component as an upper bound for the potential contribution of schooling may seriously underestimate the importance of schools.

Even here, the potential contribution of school differences is likely understated. First, any measurement error in the tests enlarges the overall variance in test performance and is implicitly attributed to the within school variance. Second, the aggregation across teachers within a grade also implicitly attributes within grade variation in teachers to the within school variance. While neither of these factors has substantial effects on standard regression analyses, each contributes to understating the estimated importance of systematic school and teacher factors.

Nevertheless, as highlighted in equation 5, the between school and grade component confound school influences with any differences in student characteristics, impeding the identification of the magnitude of school quality effects. A natural first step in isolating the contribution of teacher quality is to compare for two adjacent cohorts the correlation average of achievement cohort gains in different years (t_1 and t_2) but the same grade (j) with the correlation of gains for these adjacent cohorts in the same year (t_1) but different grades (j and k). Assuming that school wide factors remain constant during the time period (i.e., no change in δ_s), these two correlations equal

$$(6) \qquad r_{jt_1,jt_2} = \frac{\sigma_{\upsilon_{jt_1},\upsilon_{jt_2}} + \sigma^2_{\gamma_{bs}} + \sigma^2_{\delta} + 2\sigma_{\gamma\upsilon} + 2\sigma_{\gamma\delta} + 2\sigma_{\upsilon\delta}}{\sigma^2_{\gamma_{bgs}} + \sigma^2_{\upsilon} + \sigma^2_{\delta} + 2\sigma_{\gamma\upsilon} + 2\sigma_{\gamma\delta} + 2\sigma_{\upsilon\delta} + \sigma^2_{\epsilon_{bgs}}}$$

and

$$(7) \qquad r_{jt_1 kt_1} = \frac{\sigma_{\upsilon_{jt_1},\upsilon_{kt_1}} + \sigma^2_{\gamma_{bs}} + \sigma^2_{\delta} + 2\sigma_{\gamma\upsilon} + 2\sigma_{\gamma\delta} + 2\sigma_{\upsilon\delta}}{\sigma^2_{\gamma_{bgs}} + \sigma^2_{\upsilon} + \sigma^2_{\delta} + 2\sigma_{\gamma\upsilon} + 2\sigma_{\gamma\delta} + 2\sigma_{\upsilon\delta} + \sigma^2_{\epsilon_{bgs}}}$$

As equations (6) and (7) show, the difference between these two correlations is proportional to

$$(8) \qquad \sigma_{\upsilon_{jt_1}\upsilon_{jt_2}} - \sigma_{\upsilon_{jt_1}\upsilon_{kt_1}}$$

since time invariant cohort differences affect each correlation similarly.[8] Because most 4th, 5th and 6th grade students are taught by different teachers while a substantial percentage of teachers remain in the same grade and school in successive years, one would expect this difference to be positive and large if teacher quality is an important determinant of achievement and if there is substantial variation in teacher quality within schools. Of course between grade differences in the quality of other school inputs would also tend to widen this difference. On the other hand, changes over time in other school inputs (e.g. new principal, change in resources, curriculum etc.) would tend to reduce the difference between these two correlations.

Tables 2 and 3 present direct comparisons of the achievement gains for adjacent cohorts and provide the next piece of evidence on the importance of teacher heterogeneity. The tables show that the correlations in math (reading) gains between grades at one point in time are much smaller than the correlations across time for the same grades. In Table 2, the numbers in bold show that the correlation between school average math gains in 4th and 5th grade equals .00 and in 5th and 6th grade equals .12.

[8]Though entry and exit from Texas public schools alter the composition of each cohort over time, the overwhelming majority of students remain in the same school in a two year period unless districts divide grades across schools as, for example, with the move of students from elementary to junior high school.

Table 2. Correlations in School Average Math and Reading Test Score Gains for Adjacent Cohorts: 4th, 5th and 6th Graders in 1995

		Math			Reading		
		4th	5th	6th	4th	5th	6th
Math	4th	1.00					
	5th	**0.00**	1.00				
	6th	-	**0.12**	1.00			
Reading	4th	0.57	-0.01	-	1.00		
	5th	0.03	0.58	0.10	**0.05**	1.00	
	6th	-	0.10	0.51	-	**0.14**	1.00

Table 3. Correlations in School Average Math and Reading Test Score Gains for Adjacent Cohorts in Same Grade: 4th and 5th Grade in 1994 and 1995; 6th Grade in 1995 and 1996

1. 4th Grade

		Math		Reading	
		1994	1995	1994	1995
Math	1994	1.00			
	1995	**0.29**	1.00		
Reading	1994	0.57	0.18	1.00	
	1995	0.24	0.57	**0.27**	1.00

2. 5th Grade

		Math		Reading	
		1994	1995	1994	1995
Math	1994	1.00			
	1995	**0.33**	1.00		
Reading	1994	0.62	0.17	1.00	
	1995	0.22	0.58	**0.21**	1.00

3. 6th Grade

		Math		Reading	
		1995	1996	1995	1996
Math	1995	1.00			
	1996	**0.53**	1.00		
Reading	1995	0.63	0.41	1.00	
	1996	0.45	0.64	**0.45**	1.00

The corresponding correlations for school average reading gains are .05 and .14, respectively. In comparison, Table 3 shows that the correlations between school average math gains for students in the same grade in successive years equal .29 for 4th grade, .33 for 5th grade and .53 for 6th grade. The corresponding correlations for reading gains equal .27, .21 and .45 for 4th, 5th and 6th grade respectively.[9]

The much higher correlations between years for the same grade are consistent with the belief that differences in teacher quality make a substantial contribution to the variation in test score gains, but they are also consistent with the importance of a variety of other factors that vary systematically by grade. In an effort to isolate the impact of teachers, we divide schools into three groups on the basis of teacher turnover in each grade and subject: 1) less than 33 percent of teachers are the same in both years; 2) 33-99 percent of teachers are the same in both years; and 3) 100 percent of teachers are the same in both years. If differences in teacher quality are responsible for the correlations, it should be the case that the correlation in school average gains in successive years should be higher for schools in which most or all of the teachers teach in both years, and lower in schools in which there is greater turnover. Notice that we are not saying anything about the direct impact of turnover on achievement, only that persistence of achievement gains across years should decrease as turnover rises.

The top panel of Table 4 presents the correlations across schools in math and reading gains for grades 4, 5 and 6 by teacher turnover. The increase in correlations of achievement gain over time with more common teachers is exactly what would be expected when there is an important role for teacher quality. For 5th grade students, the correlation in school average math gains between successive cohorts is .18 in schools with less than 33 percent of teaching positions staffed by the same teacher in

---

[9]Not surprisingly, the school average math and reading gains for the same year and grade are highly correlated (.57 in 4th grade, .58 in 5th grade, and .51 in 6th grade). These correlations in Table 2 show gains for the same students, so directly include all common individual effects of overall ability, family influences, and the like.

Table 4. Correlations in School Average Math and Reading Test Score Gains and Between Grade Differences in Average Test Score Gains with Individual Fixed Effects: By Grade[a] and Teacher Turnover

1. School Average Gains

	Math				Reading			
	% same teachers in both years				% same teachers in both years			
	<33%	33-99%	100%	All	<33%	33-99%	100%	All
4th	0.20	0.28	0.42	0.29	0.18	0.30	0.31	0.27
5th	0.18	0.34	0.44	0.33	0.09	0.21	0.30	0.21
6th	0.40	0.56	0.54	0.53	0.33	0.49	0.42	0.45

2. Grade Differences in Average Gains with Individual Fixed Effects

	Math				Reading			
	% same teachers in both years				% same teachers in both years			
	<33%	33-90%	>90%	All	<33%	33-90%	>90%	All
4th v. 5th	0.02	0.28	0.39	0.28	0.11	0.20	0.26	0.21
5th v. 6th	0.01	0.25	0.38	0.25	0.03	0.15	0.21	0.15

a. 4th Graders in 1994 and 1995, 5th Graders in 1994 and 1995, and 6th Graders in 1995 and 1996.

both years, .34 in schools with between 33 percent and 99 percent staffed by the same teacher and .44 in schools with all of the same teachers in both years. The corresponding correlations for 5th grade reading gains are .09, .21 and .30 for high, medium and zero turnover schools respectively, and a quite similar pattern appears for 4th grade math and reading test score gains. The pattern is not quite as pronounced in 6th grade, though the correlations in high turnover schools are much lower than the rest.

Fixed Individual Effects and Specification Checks

The higher correlations of achievement gains in schools and grades with less teacher turnover provides additional support for the importance of teacher quality in determining achievement. However, there may be confounding factors unrelated to teacher quality that could be partly or even fully responsible for generating the observed pattern of correlations. In particular, the variance in the academic preparation and commitment of students may be much higher in schools with higher teacher turnover. If so, it would be imperative to account for school average differences in student characteristics.

The most direct and comprehensive approach first extracts individual student fixed effects from equation 1 and then considers the pattern of grade specific gains across years. At this point, the full value of having matched panels of achievement for different cohorts becomes apparent. Because our objective is to focus on the importance of teachers, we restrict our sample just to students who do not change schools, minimizing the effects of differences in other factors over the two years.[10] Using the panel data for two successive years for a given cohort, individual fixed effects (or average achievement gains for each student) are removed. The remaining achievement gains within each school reflect the between grade differences in teachers and schools plus any time varying error. We repeat the same

[10]If we included students who had changed schools over the two years, we could estimate individual grade effects on achievement, but this estimate would be based entirely on the school movers. Not only is it likely that these students are not representative of the students who did not change schools, but also it is no longer possible to separate the effects of teachers and of schools cleanly.

procedure for an adjacent cohort of students (i.e., the cohort going through the same grades one year after the first cohort). Comparison of the average achievement gains of, say, students in both the 4th and 5th grade in two successive cohorts within each school provides us a direct estimate of the persistence of relative quality that is not contaminated by student heterogeneity.[11] Evidence that the persistence in relative quality is stronger in schools with little teacher turnover indicates that teacher quality is an important determinant of achievement.

The bottom panel of Table 4 reports the between cohort correlations in the school average math and reading gain differentials for grades 4 and 5 and grade 5 and 6, by teacher turnover for both grades. The low turnover group in these calculations includes schools in which at least 90 percent of teachers were the same in both years (as opposed to the previously used 100 percent), due to the small number of schools that had all of the same 4th and 5th grade teachers in both years.[12] The results present a striking pattern in support of the belief that teacher quality is an important determinant of achievement. Virtually identical, the correlations in school average math gain differentials for grades 4 and 5 and grades 5 and 6 rise from close to zero for high turnover schools, to between .25 and .30 for schools with between 33 percent and 90 percent of the same teachers, and finally to almost .40 for schools in which over 90 percent of positions are staffed by the same teachers. While less pronounced,

---

[11]Note that this fixed effects framework is equivalent to a first differenced model in which the difference in gains between grades g and g-1 is a function of the difference in teacher, quality, other school characteristics and a random error. In other words, when gains are averaged across students, this becomes a difference-in-differences formulation. That is in fact the form that the actual estimation takes.

[12]There are about 200 schools in which the 4th and 5th grades have just a single teacher and about 100 in which the 5th and 6th grade have the same teacher and are found on the same campus. These schools are nonrepresentative of the state. Nonetheless, since there is an exact student-teacher match in these and since turnover of specific teachers can be followed, we replicate both the basic gain and the fixed effect estimation and compare the pattern of average gains in these schools. For math, the patterns are very similar to those in Table 4, but for reading the pattern of correlations holds only for the 5th and 6th grade estimation.

the same qualitative pattern emerges for reading gain differentials (the subject area where schools were previously shown to have less effect).

These correlations leave little doubt that there are important within school differences in teacher quality that play an important and systematic role in determining student achievement. There is no other plausible explanation for the finding that the difference in school average gains between grades g and g-1 is much more persistent over time in schools with lower teacher turnover. It is not differences in students, for student fixed effects are removed. Nor is it driven by school characteristics that affect all grades, for such influences are also differenced out. Nor is it driven by other school inputs that vary by grade, because they would not be expected to vary systematically with teacher turnover unless they are associated with teachers.

While we observe substantial within school differences in the average quality of teaching between grades, there are almost certainly similar differences among teachers within grades that we cannot observe. We now attempt to estimate a lower bound on the contribution of differences in teacher quality to the total variation in achievement gains.

Lower Bound Estimate of Teacher Quality Effects

The correlations in Table 4 provide strong evidence for the existence of substantial variations in teacher quality within schools. It is almost certain that there also is substantial sorting of teachers among schools, that would tend to amplify the role of teachers in explaining variations in student achievement. However, because administrators, neighborhood characteristics, peer groups and other factors are also distributed nonrandomly among schools, it is virtually impossible to identify any between school variation in teacher quality without explicitly specifying and measuring the differences. As a consequence, such between school variation per se is ignored, and we concentrate on estimating a lower bound to teacher effects (that would, among other things, hold if there were no between school sorting of teachers by quality). The estimation focuses on the potential contribution of teachers to within school differences in student gains between grades 4 and 5. We use these two grades and omit grade 6 because a large percentage of students attend 6th grade in a different campus than they attend grade 5 and we want to obtain a precise estimate of teacher effects uncontaminated by overall school differences.

The within-school-between-grade variation in achievement gains includes not only teacher quality but also other factors such as student characteristics, curriculum, class size, measurement error, etc., that may contribute to between grade differences. Therefore the effects of these other factors must be removed in order to estimate a lower bound of the contribution of teachers. Our strategy is to use information from schools with no teacher turnover in successive years to place a bound on the proportion of the within-school-between-cohort variance due to factors other than teacher quality differences. We can then decompose the total within-school-between-grade variation into a pure teacher effect and a combination of other influences (which may include additional, unidentified teacher effects).

These ideas can be fixed by looking at the between school and cohort variance and between school variance of achievement gains. The variance component, υ, from equation 1 is broken into two components: 1) teacher quality, τ; and 2) other school factors that vary by grade, θ, as in:

$$\upsilon_{gst} = \tau_{gst} + \theta_{gst} \tag{9}$$

This permits focusing on the independent effects of teachers. Equation (10) then describes the ratio of the between school and cohort variance divided by the between school variance in achievement gains.[13] The subscript ws refers to variance within schools.

[13]We assume that the within school covariance of τ and θ is zero across grades.

$$\text{(10)} \qquad \frac{\mathrm{Var}^{CS}(\Delta A)}{\mathrm{Var}^{S}(\Delta A)} = 1 + \frac{(\sigma^2_{\theta_{ws}} + \sigma^2_{\gamma_{ws}} + \sigma^2_{\epsilon_{ws}})}{\mathrm{Var}^{S}(\Delta A)} + \frac{\sigma^2_{\tau_{ws}}}{\mathrm{Var}^{S}(\Delta A)}$$

Notice that the between school and cohort variance and the between school variance are identical except for the within school variances of teacher quality, other school factors, student characteristics and the random error. In general it is not possible to separate these components, but the assumption that the within school variance of teacher quality equals zero in schools with no teacher turnover implies that

$$\text{(11)} \qquad \frac{(\sigma^2_{\theta_{ws}} + \sigma^2_{\gamma_{ws}} + \sigma^2_{\epsilon_{ws}})}{\mathrm{Var}^{S}(\Delta A)}\bigg|_{\text{no turnover}} = \frac{\mathrm{Var}^{CS}(\Delta A)}{\mathrm{Var}^{S}(\Delta A)}\bigg|_{\text{no turnover}} - 1$$

for schools with no teacher turnover. Since the between school and cohort variance and the between school variance are observed in the data, we can bound the proportion of the between school variance arising from factors other than teacher quality. Assume that the ratio of the between school and cohort variance to the between school variance in schools with no teacher turnover provides a good estimate of the within school variation in factors other than teacher quality for all schools. Rewriting equation (10) and using information from schools with no turnover to estimate the within school variance of other factors, the contribution of teacher quality to the variation in test score gains equals:

$$\text{(12)} \qquad \sigma^2_{\tau_{ws}} = \mathrm{Var}^{CS}(\Delta A) - \mathrm{Var}^{S}(\Delta A) - \mathrm{Var}^{S}(\Delta A) * \left(\frac{\mathrm{Var}^{CS}(\Delta A)}{\mathrm{Var}^{S}(\Delta A)}\bigg|_{\text{no turnover}} - 1\right)$$

The validity of this methodology rests on the assumption that the ratio of the between school and cohort variance to the between school variance for two fifth grade cohorts in schools with no teacher turnover provides a good estimate of the within school variation in factors other than teacher quality. While it is not possible to test this identifying assumption, it is likely that this ratio actually overstates the importance of non-teacher quality factors by assuming that teacher quality is constant over time in schools with no teacher turnover. An additional year of experience may improve or erode instructional skills, teachers may have bad years, and outside events may affect performance in the classroom. At the same time, a lower between cohort variance in average student characteristics that is correlated with teacher turnover will introduce a bias in the opposite direction. While small, we explicitly remove such differences below.

Alternative lower bound estimates of the contribution of teacher quality to the total variation in mathematics and reading achievement gains are presented in Table 5. As expected, teacher quality has a much larger impact on mathematics gains, and we focus our attention on those results. Row 1 presents lower bound estimates of the contribution of teacher quality for 4th and 5th grade students in 1995 who attend schools that have both grades. Columns 1 and 2 report the between school and between school and cohort variance as a percentage of the total variance in test scores, and column 3 reports the ratio of between school and cohort and between school variances for the 1995 and 1996 cohorts of 5th grade classes in schools that had identical teachers in both years. These three factors are used to calculate a lower bound estimate of the within school variance in teacher quality according to equations (11) and (12) as a proportion of the total variance. The lower bound in column 4 indicates that at least 3.5 percent of the total variance in math achievement gains should be attributed to average differences in teacher quality between the 4th and the 5th grade.

As noted above, between cohort differences in student characteristics might contaminate this measure. Therefore rows 2 and 3 use student fixed effects to control for between cohort differences in

Table 5. Lower Bound Estimates of the Contribution of Teacher Quality to the Variation in 4th and 5th Grade Mathematics and Reading Test Score Gains

	Student fixed effects	Measured teacher and classroom factors	I Between school and cohort variation as % of total variation	II Between school variation as a % of the total	III Between school and cohort variation as % of between school variation for schools with little teacher turnover	IV Lower Bound Estimate of Contribution of Teacher Quality (I - II*III)
1. Math						
	no	no	13.0%	6.5%	145.7%	3.5%
	yes	no	13.4%	7.7%	147.8%	2.0%
	yes	yes	13.4%	7.7%	148.2%	2.0%
2. Reading						
	no	no	7.4%	3.8%	158.5%	1.4%
	yes	no	6.9%	3.6%	162.8%	1.0%
	yes	yes	6.8%	3.5%	164.3%	1.1%

student characteristics. The estimates in row 2 are produced by stacking 4th and 5th grade gains and first removing individual fixed effects. The sample is divided into two parts: schools with more than 90 percent of the same 4th and 5th grade teachers in both cohorts; and schools with less than 33 percent of the teachers the same in both cohorts. It would be preferable to use schools with complete teacher turnover and schools with no teacher turnover, but the samples would be extremely small. Schools with more than 90 percent of the same teachers are considered the no turnover schools, and the ratio of the between school and cohort variance divided by the between school variance for these schools is shown in column 3. The between school and between school and cohort variances for schools with less than 33 percent of positions occupied by the same teachers are reported in columns 1 and 2. These three figures are again used to compute a lower bound estimate of the contribution of teacher quality, which in this case equals 2.0 percent.

The basic estimates in row 1 compare 4th and 5th grade gains in the same year and thus had very few teachers who taught both cohorts. In the fixed effects estimation, the sample includes only schools in which less than 33 percent of teaching positions were the same in successive years. Therefore a substantial minority of teachers taught both cohorts. Moreover, the sample used to produce the estimate of the within school variance of other factors included schools with different teachers for the two cohorts. In combination, these two problems depress the estimate of the contribution of teacher quality.

The final row of Table 5 uses the same sample as Row 2 but first removes any covariation with a number of measured school characteristics including class size, teacher experience and teacher turnover. This estimation provides direct information on the relative importance of the measured attributes of schools that are commonly the focus of policy compared to total teacher effects. The regression results are discussed in the next section, but the combined explanatory power of these factors is quite small and thus the lower bound estimate using these residuals is virtually identical to the lower bound estimate previously presented in the fixed effects calculations of row 2.

The results in Table 5 suggest that between grade differences in average teacher quality within schools account for at least 2.0 percent of the total variation in student test score gains. While this number might seem small, there are many reasons to believe that it vastly understates the importance of teacher quality differences.

First and perhaps most important, these calculations ignore within grade variation in teacher quality. If within and between grade differences in teacher quality are similar, the estimates can be multiplied by the average number of teachers per grade, which equals roughly 3.75, to obtain an estimate of the contribution of all within school differences in teacher quality. So doing raises the lower bound estimate to approximately 7.5 percent of the total variation in achievement gains,

Second, as discussed above, virtually all of the many methodological problems associated with the lower bound estimation tend to depress estimates of the contribution of teacher quality. Third, all between school variation in teacher quality is ignored, not because it is clearly unimportant but instead because it cannot be adequately identified within this general estimation strategy. Fourth, measurement error in the test scores is ignored. The achievement score gains are calculated from the difference between two test scores, and the fixed effects estimates are essentially the difference in gains. Such measurement error increases the total variance without raising the portion attributed to teachers, again depressing the estimates of teacher quality effects.[14]

[14]The prior estimates of the correlation of math and reading scores for the same students were all less than 0.6. In the early grades, where it is common to have just one primary teacher, these suggest that measurement error could be substantial, although measurement error and the effects of multidimensional abilities are not separately identified.

In sum, the estimated lower bound on the true contribution of teacher quality is substantial and the full effect of teachers is likely to be far greater than the estimates presented. Within the current structure of schools, considerable leverage is being exerted by the variations in teacher quality.

IV. Education Production Function Estimates

Even though the pattern of results in Table 5 suggests that observable teacher characteristics explain only a small portion of the variation in teacher quality, small but systematic relationships may provide information for the development of improved school policies. At the very least, many current policies and regulations are built on a set of assumed relationships between teacher effectiveness and teacher training, experience and turnover. An equally important issue that is at the heart of considerable current school policy debate is the role of class size in determining achievement. The pattern of correlations previously presented suggests at most a minor role for class size and other school factors not systematically related to teachers. Nonetheless, because of the ability to disentangle the separate influences on student achievement within the Texas sample, we pursue the estimation of traditional educational production functions based on the observable characteristics of teachers and schools.

Equation 13 describes the value-added empirical model that forms the basis of our examination of school resource effects on achievement,

$$\text{(13)} \qquad \Delta A_{igst} = X_i\beta + C_s\eta + S_{gst}\lambda + \gamma_i + \upsilon_{gst} + \delta_{st} + \varepsilon_{igst}$$

where X is a vector of individual and family background characteristics, C is a vector of community type dummy variables, S is a vector of school and teacher characteristics and β, η, and λ are production parameters. The family characteristics include information on race, ethnicity, gender, and whether the student is eligible to receive free or reduced price lunches. Teacher and school characteristics, described in Appendix Table A1, are computed separately for each grade and subject, and they include the average class size in regular class rooms,[15] the percentage of teachers with a master's degree and the percentage of teachers who fall into three experience categories: zero years, one year and two to four years (the omitted category is five years and above).[16] The value added form corrects for unmeasured past history of educational inputs and allows us to focus on the marginal effects of current inputs.

The data appendix describes in detail the construction of the school characteristics and sample selection criteria. In particular, considerable effort was made to eliminate measurement error in the school variables. Like most educational studies, this estimation relies on self-reported school data, and these data are prone to significant reporting errors. Unlike most studies, however, we have access to longitudinal information on key data, and therefore we can adjust for inconsistencies that occur over time. (See Appendix A). Importantly, the error components must be reinterpreted in this framework as unobservable individual, teacher and school factors, because we are now relying on just measured

[15]As Boozer and Rouse (1995) and others have pointed out, it is important to separate regular and special education students because class size and possibly other characteristics differ dramatically by population served and because special education students are much less likely to take tests. If the proportion of students in special education classes or the gap between regular classroom and special education class size differs across schools, estimates of the effect of class size based on the entire school average will be biased. Our measure of class size is the average class size for regular classrooms in specific grades and subjects. Thus, it is not contaminated by special purpose classes such as special education or bilingual education programs, even though those do mask some of the variation that exists. Separate analysis of special education is found in Hanushek, Kain, and Rivkin (1998).

[16]Including the percentage of teachers with 5-9 years of experience as a separate category did not change any of the results, and the hypothesis that teachers with 5-9 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.

differences in teachers and schools and not the total differences employed in the variance decompositions and correlational analyses.

Basic production function results

Table 6 reports ordinary least squares estimates of the effects of the teacher characteristics on 4th, 5th and 6th grade math and reading gains; Huber-White t-statistics are in parentheses.[17] These estimates suggest that the first and to a lesser extent the second year of experience significantly improve teacher quality, but that additional years rarely have a significant impact. There is also no evidence that post-graduate education improves the quality of teaching; the point estimates for the effects of a master's degrees are generally negative and always statistically insignificant from zero. Finally, the evidence on class size is somewhat mixed: there is a statistically significant negative relationship between class size and math and reading achievement in 4th and 5th grade but not in 6th grade.

The validity of these estimates rests on the assumption that none of the unobserved error components is systematically related to the teacher characteristics, but there are strong reasons to believe that at least one if not all of the error components are correlated with the teacher characteristics. First, even in this value added form, it is highly unlikely that a single measure of family socio-economic background controls for all variation in family characteristics related to both achievement and teacher characteristics. Correlation between unobserved student characteristics and teacher characteristics may result from the endogeneity of school choice and from the process by which schools match teachers with children. Second, teacher experience, education and class size are undoubtably correlated with unobservable school and teacher characteristics such as peer group quality, administrative structure, other school resources, and unobservable teacher and administrator ability.

[17]The individual and family background characteristics, three community type dummy variables indicating that the school is located in a small city, a large city, or a suburb, and dummy variables to allow for cohort effects are included in the gains regressions.

Table 6. Estimated Effects of Teacher Characteristics Grade on Math and Reading Test Score Gains, Grades 4-6 (Huber-White adjusted t statistics in parentheses)

	Math			Reading		
	4th	5th	6th	4th	5th	6th
class size	-0.0057	-0.0048	0.0005	-0.0021	-0.0016	0.0015
	[-3.70]	[-5.44]	[0.42]	[-1.95]	[-2.44]	[1.58]
% with graduate degree	-0.04	-0.01	-0.02	0.00	0.00	-0.02
	[-3.07]	[-0.58]	[-1.14]	[0.37]	[-0.58]	[-1.45]
% 0 years experience	-0.15	-0.11	-0.07	-0.10	-0.05	-0.04
	[-5.11]	[-5.41]	[-3.04]	[-5.05]	[-3.66]	[-2.23]
% 1 years experience	-0.03	-0.04	-0.03	-0.04	-0.03	-0.05
	[-1.14]	[-2.24]	[-1.45]	[-2.03]	[-2.41]	[-2.85]
% 2-4 years experience	-0.02	-0.02	-0.01	-0.01	0.00	-0.01
	[-1.06]	[-1.80]	[-0.91]	[-0.87]	[-0.52]	[-0.59]
% teachers same	0.04	0.00	-0.01	0.00	0.00	-0.01
	[3.69]	[0.23]	[-0.74]	[0.69]	[-0.72]	[-0.65]
R squared	0.0056	0.0039	0.0079	0.0112	0.0029	0.0040
Observations	335,643	500,536	362,195	334,650	499,199	362,083
Schools	3,031	3,010	1,937	3,031	3,010	1,937

Any such correlations will confound the effects of observable and unobservable characteristics and bias the estimates of the effects of the observable teacher characteristics on achievement. Moreover, the direction of the bias is ambiguous: student choice of schools likely inflates estimated effects, while a school's allocation of students to classes may well introduce a downward bias, possibly by assigning better teachers to larger classes or more difficult to educate students.

Fortunately, the repeated test scores available in the Texas data a unique opportunity to minimize the potential problems of omitted variables bias. Specifically, all time invariant factors that affect the *gains* in student achievement can be eliminated by introducing individual student fixed effects. The inclusion of individual fixed effects eliminates all observed and unobserved time invariant individual and community factors. In this estimation, the effect of school factors is identified by the their change between grades for individual students.[18]

The fixed effect estimates of equation (13) are presented in Table 7.[19] For each test, the first column reports results for fourth and fifth grade gains, while the second provides results for fifth and sixth grade gains. The analysis is separated in this way because preliminary evidence suggested differences in class size effects between grades. Further, a majority of students attends different campuses in 6th grade than in earlier grades, complicating the interpretation of the results.

The pattern of results is similar to that produced by the simple gains specification of Table 6, suggesting that any omitted individual-specific factors do not appear to yield significant estimation bias in traditional value added models. There is no evidence that teacher education is systematically related to achievement, and strong evidence of a positive return to the first two years of experience with

[18]Typically systematic variations in class size occur primarily between schools rather than among regular classrooms in the same school, but this is not true in Texas because of explicit policies to reduce class size in early grades.

[19]These regressions control for cohort differences in gains, but no other covariates are included.

Table 7. Estimated Effects of Teacher Characteristics on Math and Reading Test Score Gains with Individual Fixed Effects (Huber-White adjusted t statistics in parentheses)

	Math		Reading	
	grades 4 & 5	grades 5 & 6	grades 4 & 5	grades 5 & 6
class size	-0.0058	-0.0044	-0.0045	0.0001
	[-3.67]	[-3.32]	[-3.91]	[0.07]
% with graduate degree	-0.02	-0.02	0.00	-0.02
	[-0.86]	[-0.96]	[0.03]	[-1.43]
% 0 yrs experience	-0.12	-0.12	-0.10	-0.10
	[-4.42]	[-4.42]	[-4.54]	[-4.59]
% 1 year experience	-0.05	-0.11	-0.05	-0.09
	[-1.97]	[-4.45]	[-2.40]	[-4.85]
% 2-4 yrs experience	-0.01	-0.04	0.00	-0.02
	[-0.36]	[-2.15]	[0.16]	[-1.68]
% teachers same	0.02	-0.01	-0.01	-0.01
	[2.00]	[-0.74]	[-0.73]	[-1.00]
R squared	0.0010	0.0018	0.0008	0.0010
observations	286,948	273,377	286,606	272,736
schools	2,897	1,922	2,898	1,923

achievement gains approximately 0.1 standard deviations higher after the initial years of teaching. Class size continues to be significantly related to math and reading achievement for 4th and 5th graders, but there is no significant relationship between class size and reading achievement in the 5th and 6th grade specification.

Student fixed effects eliminate the influence of all time invariant individual effects, including not only ability differences but also fixed family, peer, and school effects. However, because a substantial percentage of students switch schools each year, differences in unobserved school or teacher characteristics could still contaminate the estimates. A new school has a different peer group, neighborhood, administrative structure, and other factors that might be related to both achievement and teacher characteristics. Even if families do not relocate, the switch from elementary to middle or junior high school may introduce omitted variables bias. This is particularly likely for the 5th and 6th grade specifications, because a large percentage of students attended 6th grade in a different campus than they attended 5th grade.

In order to mitigate any problems of unobserved school, teacher and neighborhood characteristics, we restrict the sample in two ways. First, we exclude all students who attend a different school in grade g from the one attended by the majority of their schoolmates in grade g-1 (Table 8). This eliminates the possibility that changes in neighborhood or school district characteristics biases the results but retains students who progress normally from one campus to another. Second, we further restrict the sample to students who attend grades g-1 and g at the same campus (Table 9). In terms of equation 13, this incorporates all time invariant school effects in the individual fixed effects. Unobserved differences between teacher quality (and other school factors) in grades g-1 and g remain, but there is little reason to believe that the observed teacher characteristic in a grade are systematically linked with the unobservables.

Table 8. Estimated Effects of Teacher Characteristics on Math and Reading Test Score Gains with Individual Fixed Effects for Nonmovers (Huber-White adjusted t statistics in parentheses)

	Math		Reading	
	grades 4 & 5	grades 5 & 6	grades 4 & 5	grades 5 & 6
class size	-0.0057	-0.0039	-0.0048	0.0002
	[-3.18]	[-2.53]	[-3.68]	[0.18]
% with graduate degree	-0.02	-0.01	0.00	-0.02
	[-0.85]	[-0.78]	[0.19]	[-1.21]
% 0 yrs experience	-0.12	-0.11	-0.09	-0.10
	[-3.90]	[-3.66]	[-4.05]	[-4.06]
% 1 year experience	-0.05	-0.11	-0.04	-0.10
	[-1.52]	[-3.79]	[-1.88]	[-4.69]
% 2-4 yrs experience	0.00	-0.04	0.00	-0.02
	[0.10]	[-2.23]	[0.27]	[-1.11]
% teachers same	0.02	0.00	-0.01	-0.01
	[1.56]	[-0.22]	[-0.69]	[-0/69]
R squared	0.0009	0.0016	0.0009	0.0010
observations	254,048	235,977	253,720	235,357
schools	2,778	1,842	2,778	1,842

Table 9. Estimated Effects of Teacher Characteristics on Math and Reading Test Score Gains with Individual Fixed Effects for Students at Same Campus (Huber-White adjusted t statistics in parentheses)

	Math		Reading	
	grades 4 & 5	grades 5 & 6	grades 4 & 5	grades 5 & 6
class size	-0.0066	-0.0018	-0.0044	0.0002
	[-3.53]	[-0.60]	[-3.11]	[0.11]
% with graduate degree	-0.02	0.00	0.00	0.00
	[-1.06]	[-.17]	[-0.25]	[0.14]
% 0 yrs experience	-0.11	-0.11	-0.09	-0.09
	[-3.39]	[-2.26]	[-3.51]	[-2.41]
% 1 year experience	-0.03	-0.09	-0.03	-0.06
	[-0.89]	[-2.08]	[-1.26]	[-2.02]
% 2-4 yrs experience	0.01	-0.08	0.01	-0.01
	[0.39]	[-2.29]	[0.56]	[-0.56]
% teachers same	0.02	0.00	-0.03	0.00
	[1.23]	[0.08]	[-2.29]	[0.12]
R squared	0.0010	0.0015	0.0006	0.0005
observations	211,017	77,381	210,815	77,167
schools	2,566	1,055	2,566	1,056

These sample restrictions are clearly nonrandom. They exclude students who switch schools. They also, in the second case, exclude entire school districts that organization instruction to locate 4th and 5th grades or 5th and 6th grades at different campus. While movers likely differ from stayers, such differences are unlikely to contaminate the coefficients because individual fixed effects are removed. Therefore the estimates generated by the sample of students who remain in the same campus should not be contaminated by unobserved individual, school, neighborhood or teacher characteristics, and they should be generalizable for all students and schools.

The increasingly restrictive samples yield results that are virtually identical to those previously presented. Tables 8 and 9 show that early experience continues to increase teacher quality, particularly in the first year, and there remains no evidence that completion of a master's degree increases teacher quality. The statistically significant negative effect of class size on math achievement in the 4th and 5th grade estimation holds with the fixed effects estimation and with the sample restricted to students who remain in the same campus. Again, also, effects on reading performance seem nonexistent in later grades.

Though the sample restrictions certainly mitigate or even eliminate problems of omitted variables bias, the pattern of results suggests a different type of specification error. It appears that the effect of class size varies by grade, meaning that the restriction of identical class size effects for grades g-1 and g is invalid. Table 10 reports estimates that permit class size effects to vary by grade using the restricted sample of students who remain in the same campus for successive grades. The results show that the effect of class size on math achievement is roughly 30 percent larger in 4th grade than in 5th grade and very small and statistically insignificant in 6th grade. The effect on reading achievement is almost 50 percent larger in 4th grade than in 5th grade, but it remains below that for math. The 5th grade class size coefficients estimated separately from the two samples are similar in magnitude for math but different in magnitude for reading. The small and insignificant 5th grade coefficient in the

Table 10. Estimated Grade Specific Effects of Class Size on Math and Reading Test Score Gains with Individual Fixed Effects for Students at Same Campus (Huber-White adjusted t statistics in parentheses)

	Math		**Reading**	
	grades 4 & 5	grades 5 & 6	grades 4 & 5	grades 5 & 6
4^{th} grade class size	-0.0075		-0.0059	
	[-2.71]		[-2.74]	
5^{th} grade class size	-0.0057	-0.0040	-0.0032	-0.0002
	[-2.92]	[-1.30]	[-2.20]	[-0.10]
6^{th} grade class size		0.0000		0.0013
		[0.02]		[0.57]

reading specification that includes 5th and 6th grade students may reflect a small sample size that inhibits the estimation of more precise coefficients, or it may reflect true parameters that are indistinguishable from zero.

Heterogeneity of students

One issue of policy relevance is whether or not all students respond equally to school resources or treatments. Summers and Wolfe (1977), the experimental evidence from Project STAR (Word et al., 1990), and others suggest that small class sizes are particularly important for disadvantaged students. The previous estimates considered the entire sample without regard to the possibility of such differential effects.

Table 11 presents estimates of class size effects that are allowed to vary by income, as measured by eligibility for free or reduced price lunch. Similar to other research, there is some support for larger effects for lower income students. The class size coefficients for both 4th and 5th grade math achievement are highly significant for low income students but small and not statistically significant for others, and the class size coefficient for 4th grade reading achievement is almost twice as big for low income students. Similarly small and insignificant coefficients are observed for 5th grade reading and 6th grade math and reading specifications for all students. In contrast, neither the evidence on teacher experience nor the analysis of unobserved teacher quality effects (not reported) shows systematic differences by family income. Note that the larger effects for low income students compared to higher income students do not reflect simple diminishing returns to reducing class size, as average class sizes are actually smaller for lower income students.

The division by income must also be put into perspective in terms of achievement. The average test score for higher income students is less than 0.1 standard deviations higher than for students eligible for free or reduced lunch. This considerable overlap in the achievement distributions

Table 11. Estimated Effects of Class Size on Math and Reading Test Score Gains with Individual Fixed Effects for Students at Same Campus, by eligibility for free or reduced price lunch (Huber-White adjusted t statistics in parentheses)

Free or reduced price lunch	Math	Reading	Descriptive Statistics: Average Class Size	Observa-tions	Schools
4th Grade					
Eligible	-0.013 [-3.64]	-0.0082 [-2.62]	19.6	79,027	2,533
Not Eligible	-0.0034 [-1.16]	-0.0047 [-2.18]	19.8	131,990	2,519
5th Grade					
Eligible	-0.0080 [-3.11]	-0.0027 [-1.25]	22.2	79,027	2,533
Not Eligible	-0.0031 [-1.45]	-0.0018 [-1.25]	22.8	131,990	2,519
6th Grade					
Eligible	0.0007 [0.16]	-0.0038 [-1.18]	22.3	29,305	1,046
Not Eligible	0.0002 [0.07]	-0.0004 [-0.16]	23.1	48,076	1,036

suggests that the role for teachers is truly substantial, contrary to many past interpretations that only family background really matters.

Magnitude of class size effects

From just the prior coefficient estimates, it is difficult to assess whether these class size effects are large or small. Though it is difficult to place a monetary value on increases in math or reading scores in elementary school, we can provide some indication of magnitudes. We initially report the impact of a 10 student, or 50 percent, reduction in class size in terms of its effect on achievement gains relative to total and to between school[20] variation in achievement. We will also report the percentage of total variation in gains and the percentage of the between school variation in gains explained by differences in class size. In each case the estimates from the flexible specification that allows for different effects in each grade will be used.

Class size exerts the largest impact in 4th grade where the estimates imply that a 10 student reduction in class size would increase achievement gains by .11 and .08 standard deviations of the total distribution of gains and by .30 and .31 standard deviations of the between school distribution of gains for math and reading, respectively. The estimated effect sizes for 5th grade are smaller: the same reduction in class size would increase achievement gains by .09 and .05 standard deviations of the total distribution of gains and by .26 and .19 standard deviations of the between school distribution of gains for math and reading respectively.

The evidence also suggests larger effects from targeting low income students: A 10 student reduction in class size is estimated to increase achievement gains by .18 and .12 standard deviations of the total distribution of gains and by .51 and .43 standard deviations of the between school distribution of gains for math and reading respectively. The estimated effect sizes of a 10 student reduction in class

[20]We use the between school and year variation to account for the fact that schools change over time.

size for 5th grade low income students falls to .13 and .04 standard deviations of the total distribution of gains and by .36 and .16 standard deviations of the between school distribution of gains for math and reading, respectively.

Note that the validity of these projections rests on the assumption of a linear relationship between achievement and class size, which may not hold for a statewide class size reduction of 10 students which is 2½ standard deviations in the current distribution of class sizes. Therefore these estimates are probably more relevant for schools that currently have much larger than average class sizes.

A second way of determining the importance of class size relative to other factors is to measure its explanatory power. Table 12 reports the partial R square values for the 4th, 5th and 6th grade math and reading specifications, and also expresses explanatory power as a percentage of the between school variance in achievement gains. The results show that class size explains less than 0.1 percent of the total variation in achievement gains and less than 0.5 percent of the between school and year variance of test score gains in all specifications.[21] Thus the contribution of class size to the variation in achievement gains is less than one twentieth of the contribution of teacher quality differences despite the substantial class size differences within schools.[22]

V. Conclusions

Policy interest in education has been very high, reflecting both the public and scientific agreement about the importance of schooling. The difficulty in identifying clear, efficacious policies

---

[21]Calculations that use the OLS estimates produce similar findings.

[22]The standard deviation of the within school difference in class size between 4th and 5th grade equals 3.3, quite similar to the between school variation in class size for the 5th grade.

Table 12. Percent of Total and Between School and Year Unexplained Variance in Test Score Gains Explained by Average Class Size (based on separate effects for each grade in Table 10)

	Math		Reading	
	5th v. 4th grade gain	6th v. 5th grade gain	5th v.4th grade gain	6th v. 5th grade gain
% Total Variance Explained	0.05%	0.04%	0.02%	0.00%
% Between School and Year Variance Explained	0.4%	0.3%	0.3%	0.1%

has, however, led to confusion in discussions and, more ominously, to proposals to scale back educational efforts.

Past research attempts to clarify the impact of schools on student performance have tended to worsen the situation by providing conflicting and unreliable conclusions. Part of the remaining controversies from past research (see, for example, Burtless [1996], Card and Krueger [1996] and Hanushek [1996]) reflect differences in the questions being asked. A significant portion of the controversy, however, results from making inferences with limited and incomplete data. Here we employ the extensive data set currently being developed through the Harvard/UTD Texas Schools Project. This unique data set, painstakingly developed over several years, covers entire cohorts of students in the State of Texas and permits entirely new approaches to the analysis of the determinants of student performance.

1. Schools matter importantly for student achievement.

The issue of whether or not there is significant variation in the quality of schools has lingered, quite inappropriately, since the original Coleman Report. This analysis identifies large differences in the quality of schooling in a way that rules out the possibility that they are driven by nonschool factors. Even if none of the between school variation in achievement is attributed to schools, it is clear that school quality is an important determinant of academic performance and an important tool for raising the achievement of low income students. At the same time, the analysis suggests that resource differences explain at most a small part of the difference in school quality, raising serious doubts that additional expenditures would substantially raise achievement under the current institutional structure.

2. Variations among teachers dominate school quality differences.

Through a series of increasingly refined tests of the source of achievement variations, we conclude that the most significant component is heterogeneity among teachers. Even if teachers were randomly distributed among schools (highly unlikely) and all of the between school variation in achievement were to result from other school inputs (it is even more unlikely that students are randomly distributed among schools), differences in teacher quality would swamp all other school inputs. The importance of class size in particular pales in comparison to that of teacher quality. The lower bound estimates suggest that differences in teacher quality explain at least 7.5 percent of the total variation in measured achievement gains, and probably much more.[23]

The identification of specific teacher effects and of significant heterogeneity within school buildings has immediate implications for the consideration of educational policy. While one thrust of policy discussion has focused on the quality of school leadership as the place for leveraging change, the existing structure of schools and school operations does not support that as a particularly likely place for change. At least with the current constraints of school organization and hiring regulations, systematic differences in leadership do not appear to be the driving force behind school quality differences.

Much of the debate about schools in recent years has centered on the potential for choice to improve schools. One argument along these lines is that income currently determines the amount of choice that is available. By choosing particular residential locations, middle- and upper-income families can exercise choice and can get their children into good schools. Lower-income parents do not have those options readily available, and thus their children are relegated to inferior schools. But this analysis suggests that such differences are significantly less important than much of the current discussion indicates, because within-school heterogeneity in quality hits all income groups. The divergence between perception and actual operations undoubtedly reflects the differences between overall level of student performance and the grade-by-grade growth in achievement. In any case, these

[23]Importantly, these estimates of the potential variation explained by teacher variations implicitly attribute all measurement error in the tests to within school variation in gains, thus lowering the apparent impact of between teacher variations.

results suggest that the success of a voucher or public school choice program in raising achievement would hinge largely upon the impact on the quality of teaching. Altered incentive structures might increase teacher effort and improve hiring practices, but, while theoretically appealing, there is currently little hard evidence to support such beliefs. Much more needs to be learned it about ways to improve the quality of teaching, as the problems of the current system are far deeper than simply selecting into a "good" or "bad" school.

3. Class size, teacher education and teacher experience are a small component of variations in school quality.

Much of the confusion about the importance of schools has been a reflection of the inability to describe quality differences in terms of simple characteristics of teachers and schools. The previous analysis, building on a variance-components model, did not identify what leads to differences; it only identified the existence of strong systematic differences in the expected gains in achievement owing to having different teachers.

To relate this analysis to previous work, we have estimated standard forms of explicit educational production functions. Four major conclusions emerge from this.

- Measured factors capture just a small proportion of the differences among teachers and schools.[24]
- 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 ability over the first few years of a career, although there is little evidence that improvements continue after the first couple of years.

[24]Our analysis employs a very small number of measured attributes – those that directly affect school costs. A larger set of attributes and more disaggregation to individual teachers could increase the estimated importance of measured characteristics.

- Class size appears to have a significant effect on the academic achievement of children from low-income families in the 4th and 5th grades but any effect declines with grade level.

The estimated relationship between achievement and graduate degrees and experience opens questions about the prevalence of teacher pay scales that reward these characteristics. At a minimum, these results raise serious doubts about policies that require or strongly encourage graduate education for teachers. And, while there may be ways to speed the acquisition of the relevant skills that come with early experience, these findings do not lead to immediate general policies. It is, however, frequently asserted that suburban districts with greater ability to choose among a pool of applicants hire teachers who gain their initial experience in city schools. Such policies would yield distributional effects, even though the effects on average performance across the state could be minimal.[25]

The most significant question revolves around class size effects. Previous work has led to a wide range of interpretations. Hanushek (1997) argues that the weight of past econometric evidence does not support a conclusion that class size is important. On the other hand, there are those who believe that a myriad of statistical problems obfuscates the strong and significant relationship between achievement and resources.[26]

The analysis here finds that class size has little impact on the achievement of children not from low income families, but a positive effect on the math and reading achievement of low income children in 4th and 5th grade. The fact that a sample exceeding one half million observations and three

---

[25]We have not attempted to analyze the patterns and determinants of resources and inputs across districts here. See, however, Kain and Singleton (1996).

[26]For example, some would place great weight on findings from the randomized experiment in the 1980s in the State of Tennessee and other natural experiments (e.g. Word et al., 1990; Angrist and Lavy, 1996) showing that reductions in class size significantly raise achievement in some grades. These are nonetheless also subject to contradictory findings (e.g. Hoxby, 1996).

thousands schools generates very small and mostly insignificant estimates of class size effects for the non-low income population strongly suggests that policies to reduce 4th, 5th or 6th grade class size for these students will not be cost effective. Whether the magnitude of class size effects for low income children justifies additional reductions in class size is quite difficult to evaluate.[27] It is clear, nonetheless, that teacher quality differences are much more important than variations in class size.

4. Policy makers face a dilemma in attempting to improve school quality.

The limited effectiveness of resource policies within the current structure compared with the much larger potential of policies that affect the quality of the teaching force underscore the clear dilemma of educational policy. Accurate measurement of the value added of teachers and schools is extraordinarily difficult for both policy makers and consumers to obtain given the normal standards of current reporting of student achievement. As a result, incentives within schools are not closely related to value added, and even the efficacy of individual actions—say, through Tiebout moving—is open to serious question. But failure to deal directly with the information and incentives issues leaves a collection of expensive and often ineffective policy options that ignore the primary determinant of school quality.

[27]The analysis of class size policies must address a series of fundamental issues: are the estimates of effect sizes relevant over the entire range of class sizes? What is the monetary value on achievement gains? and, what are the costs of class size reductions, particularly if we consider the average salaries required to reduce class size without reducing average teacher quality? Finally, the analysis would have to consider the alternative uses of the resources. Current spending on preschool, after school and summer programs is quite low particularly for lower income children, and to be efficacious it must be the case that the gains from resources devoted to class size exceed the benefits of using such resources in other ways.

References

Angrist, Joshua D., and Victor Lavy. "Using Maimondides' rule to estimate the effect of class size on scholastic achievement". NBER Working Paper 5888, January 1997.

Ballou, Dale, and Michael Podgursky. *Teacher pay and teacher quality*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1997.

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

Burtless, Gary. "Introduction and summary." In *Does money matter? The effect of school resources on student achievement and adult success*, edited by Gary Burtless, 1-42. Washington, DC: Brookings, 1996.

Card, David, and Alan B. Krueger. "Labor market effects of school quality: Theory and evidence." In *Does money matter? The effect of school resources on student achievement and adult success*, edited by Gary Burtless, 97-140. Washington, DC: Brookings, 1996.

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

Finn, Jeremy D., and Charles M. Achilles. "Answers and Questions about class size: A statewide experiment." *American Educational Research Journal* 27, no. 3 (Fall 1990): 557-77.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. "The effect of school resources on student achievement." *Review of Educational Research* 66, no. 3 (Fall 1996): 361-96.

Hanushek, Eric A. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources* 14, no. 3 (Summer 1979): 351-88.

—. "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature* 24, no. 3 (September 1986): 1141-77.

—. "School resources and student performance." In *Does money matter? The effect of school resources on student achievement and adult success*, edited by Gary Burtless, 43-73. Washington, DC: Brookings, 1996.

—. "Assessing the effects of school resources on student performance: An update." *Educational Evaluation and Policy Analysis* 19, no. 2 (Summer 1997): 141-64.

Hanushek, Eric A., and John F. Kain. "On the value of `equality of educational opportunity' as a guide to public policy." In *On equality of educational opportunity*, edited by Frederick Mosteller and Daniel P. Moynihan, 116-45. New York: Random House, 1972.

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

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

Hoxby, Caroline Minter. "The effects of class size and composition on student achievement: New evidence from natural population variation". Department of Economics, Harvard University (mimeo), July 1996.

—. "How teachers' unions affect education production." *Quarterly Journal of Economics* 111, no. 3 (August 1996): 671-718.

Kain, John F., and Kraig Singleton. "Equality of Educational Opportunity revisited." *New England Economic Review* (May/June 1996): 87-111.

Krueger, Alan B. "Experimental estimates of education production functions". NBER Working Paper 6051, June 1997.

Tiebout, Charles M. "A pure theory of local expenditures." *Journal of Political Economy* 64 (October 1956): 416-24.

U.S. Department of Education. *Digest of Education Statistics, 1997.* Washington, DC: National Center for Education Statistics, 1997.

Word, Elizabeth, John Johnston, Helen Pate Bain, B. DeWayne Fulton, Jayne Boyd Zaharies, Martha Nannette Lintz, Charles M. Achilles, John Folger, and Carolyn Breda. *Student/teacher achievement ratio (STAR), Tennessee's K-3 class size study: Final summary report, 1985-1990.* Nashville, TN: Tennessee State Department of Education, 1990.

Appendix A. Texas School Data

The data set used in the education production function regressions combines test score data with information on teachers and schools. In an effort to reduce problems associated with measurement error, a number of observations are excluded from the data set. As noted in the text, all participants in special or bilingual education are excluded, as are the bottom one percent of test scores in each year and the top and bottom one percent of test score gains. Measurement error in the teacher characteristics is perhaps even more of a problem. 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 was 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, and 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 35 are set to missing prior to the computation of school averages for each grade. It is our understanding that very few elementary schools in Texas have actual class sizes 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. Finally, in the specifications that are restricted to students who remain in the same school for successive years, students whose class size changes by more than 15 students are excluded.

Estimates of class size effects increased in magnitude following these exclusions.

Appendix Table A1. Variable Means and Standard Deviations

grade	Test Scores: math gain	Test Scores: reading gain	class size	Teacher Characteristics: %graduate degree	% 0 years experience	% 1 year experience	% 2-4 years experience	Observations
4th	-0.04	-0.07	19.8	24.8	5.6	5.6	17.2	335,643
	0.69	0.67	2.3					
5th	0.01	0.01	22.8	25.8	5.7	5.8	16.5	500,536
	0.59	0.63	3.4					
6th	0.05	0.05	22.8	24.5	7.8	6.9	17.6	362,195
	0.56	0.62	3.7					