Schools and the Equal Opportunity Problem

edited by Ludger Woessmann
and Paul E. Peterson

CESifo Seminar Series

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Palatino on 3B2 by Asco Typesetters, Hong Kong. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Schools and the equal opportunity problem / Ludger Woessmann and Paul E. Peterson, editors.
p. cm. — (CESifo seminar series)
Includes bibliographical references and index.
ISBN 978-0-262-23257-9 (alk. paper)
1. Educational equalization—Europe—Congresses. 2. Educational equalization—United States—Congresses. 3. Educational change—Europe—Congresses. 4. Educational change—United States—Congresses. 5. Equality—Europe—Congresses. 6. Equality—United States—Congresses. I. Woessmann, Ludger. II. Peterson, Paul E.

LC213.3.E85S36 2007
379.2'6—dc22

2006046908

10 9 8 7 6 5 4 3 2 1

Contents

Series Foreword

This book is part of the CESifo Seminar Series. The series aims to cover topical policy issues in economics from a largely European perspective. The books in this series are the products of the papers and intensive debates that took place during the seminars hosted by CESifo, an international research network of renowned economists organized jointly by the Center for Economic Studies at Ludwig-Maximilians-Universität, Munich, and the Ifo Institute for Economic Research. All publications in this series have been carefully selected and refereed by members of the CESifo research network.

7

Some U.S. Evidence on How the Distribution of Educational Outcomes Can Be Changed

Eric A. Hanushek

Distributional issues are seldom far from the minds of U.S. educational policymakers. At a minimum, information is readily available on the proportion of students who fail to achieve some level of proficiency on standardized tests. Attention to such issues has even been written into U.S. federal law with the No Child Left Behind Act of 2001. But observing differences in performance and knowing what to do about them are not the same thing. Indeed, a variety of researchers and policymakers have argued that the schools cannot be expected to have much impact on the existing distribution of educational outcomes. The theme developed here is that many discussions have confused the potential for impact with current results based on the existing organization of schools.

This chapter assesses recent evidence on schools' potential impact on both the level and pattern of student achievement. The central quantitative estimates rely on a consistent set of analyses of the Texas Schools Project that my colleagues—principally Steven Rivkin and John Kain—and I have conducted. These estimated impacts, which range across a variety of separate areas of policy concern, provide powerful evidence of the influence of schools on achievement, but the results are seldom jointly considered in contemplating policy.

The starting point of this discussion is a review of observed outcomes of U.S. schooling as they have evolved over time. Since the beginning of regular testing, focus has centered on the significant variations in student performance identified by race. Yet policy initiatives have appeared to have relatively minimal impact on the test variation.

Research has also provided somewhat disheartening findings, suggesting the limited impact of schools. The accumulated evidence has not provided much in the way of systematic findings that suggest obvious policies for the improvement of student achievement.

The goal of this work is assessing the leverage that public policy can have to change the current patterns of achievement disparities. The central focus is racial differences in achievement, although other dimensions are also considered. While the discussions are closely related to how overall performance can be improved, differences arise with distributional issues.

This discussion begins with a review of these different strands of research put into the context of a common database and school system. It is difficult to make clear statements about much of the prior discussion of factors that impact achievement because data and modeling issues become completely intertwined with the analytical results. This discussion does not attempt to provide a thorough discussion of relevant existing work—a Herculean task given the breadth and depth of research that has now developed. Instead, it focuses directly on analyses of student performance in Texas that permit a consistent evaluation of outcomes. Texas is itself interesting because it is a large and diverse state that permits a variety of detailed analyses of performance. More important, by focusing on a single state with a common methodological approach, the magnitudes of various achievement factors can be compared directly.

The analysis suggests that the common interpretation of the evidence is much too pessimistic because there are actions of schools that have considerable potential for change. Concentrating on the white-black achievement gap, the analysis suggests that improving the quality of teachers and their assignment and altering the peer composition in schools can produce noticeable changes in achievement gaps.

7.1 School Outcome Differences, Research, and Policy

Most school-policy discussions go back and forth between consideration of overall performance and consideration of the distribution of outcomes. The backdrop for this chapter—which emphasizes distributional, or equity, concerns—is what has been happening in terms of overall performance. If there is a concern about an equity-efficiency trade-off, it would be important to compare movements in one dimension with those in the other.

Figure 7.1 provides the overall performance of U.S. 17-year-olds on the National Assessment of Educational Progress (NAEP). NAEP provides a consistent national testing of a random sample of students in different subjects, so it is possible to observe any changes in performance over time.[1] The remarkable thing about this picture is that performance appears roughly flat for over three decades. This constancy is particularly remarkable given the effort expended (measured in terms of resources) to improve the performance.

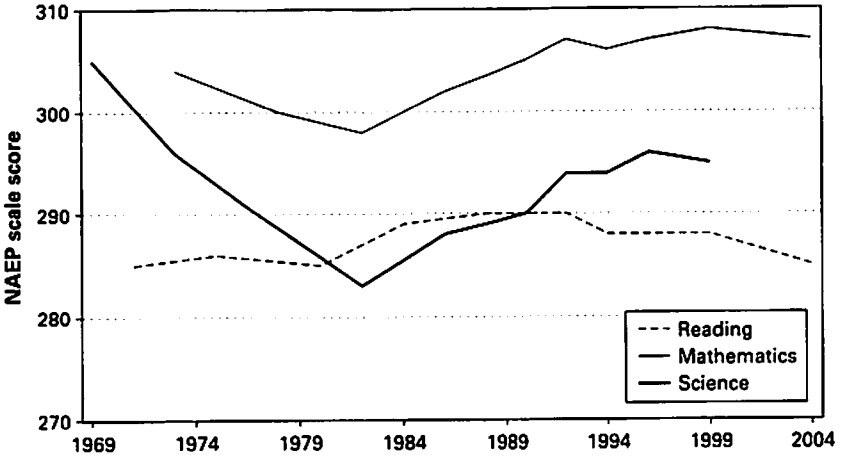

Figure 7.1
Trends in NAEP performance by subject, 17-year-olds

This flat achievement profile has been maintained in the face of substantial policy efforts to change it. Probably the most obvious policy change has been continued increases in the funding and resources of schools. The commonly discussed policy instruments—reducing pupil-teacher ratios, retaining more teachers, and having more educated teachers—have been systematically employed over the past decades. Between 1960 and 2000, U.S. pupil-teacher ratios fell by a third, teachers with a master's degree and over doubled to above 50 percent, and average experience increased (see Hanushek 2003). These actions are expensive, and real spending per pupil more than tripled between 1960 and 2000.

The simple picture thus is that school policy has not been directed primarily at overall student performance (at least as seen by outcomes). Thus, it is also useful to see what happened in terms of the distribution of outcomes. This discussion concentrates largely on racial differences in performance patterns, although income differences are also discussed below.

Over a long period, differences in school attainment by race and family background have been the subject of analysis. The large

discrepancies in quantity of schooling for blacks and whites are easily seen from decennial census data (e.g., Smith and Welch 1989; Jaynes and Williams 1989; Neal 2006). Analyses of the differences in schooling also pointed to potential quality differences, arising partly from segregated schools but also from differences in local schooling outside of states that had de jure segregation of schools. The evidence on such differences centered on data about such things as credentials of teachers, length of the school year, and spending differences among the schools attended by blacks and those attended by whites.

The attention to the quality issue was elevated, however, by a massive government report, *Equality of Educational Opportunity*, commonly referred to as the "Coleman Report" after its principal author (Coleman et al. 1966). This report was mandated by the Civil Rights Act of 1964, which instructed the U.S. Office of Education to report on the lack of educational opportunity by reason of race or ethnicity. To address this issue, the Coleman research team turned attention to school outcomes through testing some 600,000 students in the United States in 1965.

The analysis vividly underscored huge differences in the achievement of students by race and background. A simple summary of the magnitude of differences comes from equating test scores to grade-level equivalents. If white twelfth graders in the urban Northeast (in 1965) were the standard for the knowledge that a twelfth grader should have, black twelfth-grade students also in the urban Northeast were achieving at the ninth-grade level, and black twelfth-grade students in the rural South were achieving at the seventh-grade level. Surprisingly, however, the magnitude of these differences never received much attention, perhaps because most of the discussion revolved around their analysis of the determinants of achievement (below).

The achievement differences have been consistent across studies. For example, when disaggregated by race, the SAT tests showed differences of approximately one standard deviation. The SAT relied on voluntary test taking for a changing group of students, however, and thus the interpretation is somewhat ambiguous.

The clearest picture nonetheless again comes from the National Assessment of Educational Progress. Figure 7.2 displays the average performance gap between whites and blacks in the different subject areas at age 17. Across each of the tests there is a very consistent pattern: racial gaps tended to shrink noticeably during the 1980s and then to be flat or to widen somewhat during the 1990s. If anything, the white-

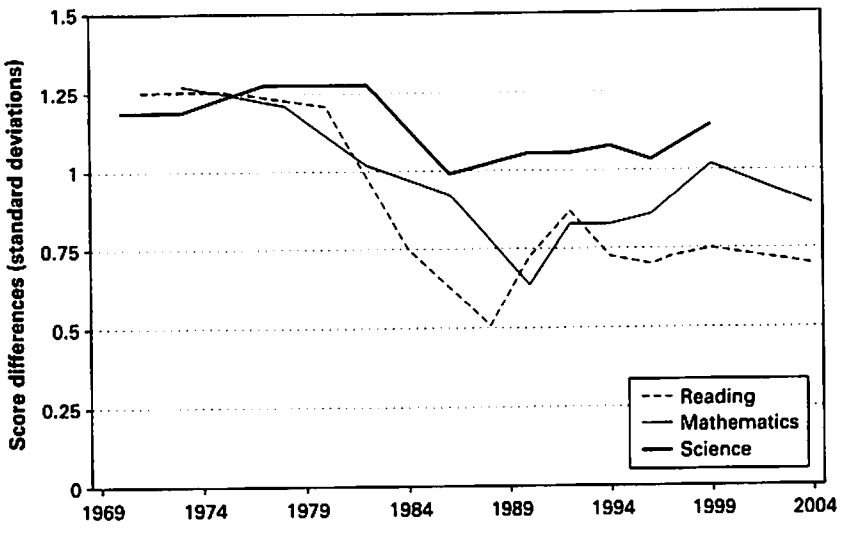

Figure 7.2
White-black differences in NAEP scores, 17-year-olds

black gap expanded some in the 1990s (even though the white-Hispanic gap, not shown, narrowed some).[2]

Much has been made of the narrowing of the black-white achievement gap including a widely cited conference book (Jencks and Phillips 1998). The one-time nature of the test-score convergence, however, was not anticipated and has received less attention than the significant closing of the gaps that occurred over a decade ago.

The resource patterns described previously were not explicitly directed at disadvantaged students or at the racial and ethnic gaps in performance. Nonetheless, throughout this period the level of performance overall did not increase (see figure 7.1), suggesting that it was not just resources going to majority students. Moreover, there was indeed a general tendency to focus money on disadvantaged students with spending on the schools of the disadvantaged (particularly inner-city schools) surpassing that others (National Center for Education Statistics 2004). During this period, the federal government also expanded its role in providing compensatory funds for disadvantaged students under the Elementary and Secondary School Act of 1965.[3] This targeted funding was also accompanied by federal support of preschool programs for disadvantaged students under the Head Start program.

The fact that substantial resources had been put into improving schools and specifically into raising the performance of disadvantaged students with no results has led to some discouragement about the efficacy of school programs to improve equity. Decades of attempts to add programs and improve the schools for disadvantaged students have shown little impact except perhaps for those in the late 1970s.

This aggregate situation was reinforced and extended by the analysis in the 1966 Coleman Report and subsequent work. The Coleman Report is commonly viewed as the first attempt to judge systematically the factors that affect student outcomes. The background is straightforward. While the U.S. Office of Education was instructed to report on inequality of educational opportunity, it did not have any common metric for assessing the importance of different resources that might enter into achievement differences. If, for example, it surveyed schools and found that one group had better science laboratories but its teachers had less experience than another group, which students were better off?

To deal with this issue, the Coleman team pursued a statistical analysis of the determinants of student performance—an introductory foray into what is now commonly referred to as *educational production function analysis*. The Coleman Report came out with the stunning conclusion that the most important factor in achievement was parents and that schools played a much less important role. In fact, in terms of impact, the ordering of influences was family, peers, and finally schools. This led to two very common statements in policy debates. First, by far the most important influence on achievement cannot be readily treated by public policy because we are not prepared to intervene in the family except in extreme circumstances. Second, schools do not make a difference.

The Coleman Report has been heavily criticized for its methodology.[4] Nonetheless, many of the basic findings of the Coleman Report have been confirmed—namely, many of the measured attributes of teachers and schools, following the approach of the report, have not been systematically related to student performance (see the review in Hanushek 2003). A wide variety of statistical analyses have failed to find descriptors of schools and teachers that are consistently associated with improved student achievement.

The interpretation of the results from the Coleman Report and subsequent work is very important and guides the remainder of this discussion. Specifically, finding that a series of measures of teacher

characteristics do not systematically influence performance is not the same as finding that teachers do not matter. Since the publication of the Coleman Report, there has been a continued confusion between measurement and effectiveness.

The issue of measurement pervades all of the discussions and is the heart of the various analyses that we have undertaken. In simplest terms, accurately identifying the influences of both schools and peers is highly dependent on having satisfactory measures of the range of various influences.

The policy leverage to deal with equity and performance issues in schools resides in altering the operations of schools and, perhaps, affecting the composition or peer groups. Therefore, it is crucial that these influences are accurately identified and estimated. Specifically, much of new work on achievement differences concentrates on issues of causal relationships. One concern with much of the past research into student performance is that it has not accurately identified factors that directly affect performance but instead has obtained biased estimates owing to misspecification of the underlying models. Thus, a key element in the ensuing work is to identify reliably factors that are causally related to achievement.

7.2 Texas Schools

The analysis here is based on the experiences in the state of Texas. It is useful to understand the nature of Texas and the schools in Texas. With some 3 million students, Texas is the second-largest state.[5] White and Hispanic students each make up slightly over 40 percent of the student population, with blacks being about 15 percent. The state combines both heavily urbanized areas and very rural areas; 15 districts are in the top 100 districts of the nation in terms of student population. Its spending in 2000 was $6,288 per student, or 91 percent of the national average. Performance on the NAEP tests in both math and reading is approximately the level of the national average.

The analysis here relies on state administrative records for student performance and school characteristics. The cornerstone of the analysis of teacher quality is the unique stacked panel dataset constructed by the Texas Schools Project of the University of Texas at Dallas. The data on students, teachers, schools, and other personnel come from administrative records on individual students and teachers collected by the Texas Education Agency (TEA) and follow several entire cohorts of

students. Each cohort contains some 200,000 students, and depending on the specific analysis, individual students are followed for up to five years. The student data contain a number of student, family, and program characteristics including race, ethnicity, gender, and eligibility for a free or reduced-price lunch (the measure of economic disadvantage) and Title I services. Students are also observed when they switch schools and can be followed across all public schools in Texas. Teacher and administrative personnel information include characteristics such as race/ethnicity, degrees earned, years of experience, certification test results, tenure with the current district, role, and campus.

Student performance is assessed by the Texas Assessment of Academic Skills (TAAS), which was administered each spring to eligible students enrolled in grades three through eight. These criterion-referenced tests evaluate student mastery of grade-specific subject matter in reading and mathematics.

The relative performance of students in Texas is seen from table 7.1, which provides performance by family income and race/ethnicity for the cohort of fifth-grade students in 1995. TAAS score are normalized for the state to have a mean of zero and a standard deviation of one. However, following most of the analyses employed here, students in special education or limited English proficiency (LEP) programs along

Table 7.1
Student performance in Texas by race/ethnicity/income

	Proportion of students	Mean fifth-grade TAAS math	Mean sixth-grade TAAS math
Total	1.00	0.14	0.20
By eligibility for free or reduced price lunch:			
Eligible	0.40	−0.16	−0.09
Not eligible	0.60	0.33	0.39
Race/ethnicity:			
White	0.58	0.34	0.41
Black	0.13	−0.37	−0.26
Hispanic	0.29	−0.08	−0.04

Notes: Data for fifth-grade students in 1995. Mean achievement scores are standardized to mean zero and variance one for the entire grade cohort in Texas. The data exclude special education and students identified as having limited English ability, thus yielding the nonzero mean for the population.

with students lacking information on gains over time are excluded, so that the average performance of the remaining students is 0.14. In this sample, math performance of white students exceeds that of blacks by 0.71 standard deviation and of Hispanics by 0.42. Students eligible for free or reduced-price lunch fall almost one-half standard deviation below those not eligible.

These data, following individual students over time and across schooling experiences, permit unique analyses of the determinants of achievement. The question is whether the gaps in performance are affected by public policy and, if so, by how much.

7.3 Teacher Quality

Since the Coleman Report, answers to questions about the impact of schools have been surrounded by a series of difficult methodological problems. To understand the basic nature of these, we begin with a simple description of student achievement and then proceed to consider ways of analyzing it.

Today's achievement is influenced not just by current family, school, and peer interactions but also by those of the past that establish the base for any current learning. This fundamental relationship is captured by equation (7.1) that describes achievement (A) for student i in grade G, in school s:

$$A_{iGs_t} = \underbrace{X_{iGs_t}\beta_G + S_{Gs_t}\delta_G + \bar{P}_{(-i)Gs_t}\lambda_G}_{\textit{current inputs}} + \underbrace{\sum_{\tau=1}^{G-1} X_{i,G-\tau,s_{t-\tau}}\beta_g + \sum_{\tau=1}^{G-1} S_{i,G-\tau,s_{t-\tau}}\delta_{G-\tau} + \sum_{\tau=1}^{G-1} \bar{P}_{(-i),G-\tau,s_{t-\tau}}\lambda_{G-\tau}}_{\textit{cumulative past inputs}} + \sum_{\tau=1}^{G} e_{i,G-\tau,s_{t-\tau}}, \quad (7.1)$$

where $\bar{P}$ measures peer behavior, X and S are vectors of relevant family background and school inputs, respectively, and the subscript $(-i)$ indicates that peer measures omit attributes of student i. Because it is useful for developing the estimation issues, this representation separates current and past influences.[6]

Clearly, simply estimating relationships between the current level of achievement and the current inputs has little chance of accurately separating the various influences on achievement. Almost certainly, current inputs are correlated with past inputs, leading to obvious problems.

The now standard approach of analyzing the growth in student achievement, as in equation (7.2), substantially reduces the problem, but not all concerns are eliminated:

$$\begin{aligned}\Delta A_{iGs_t} &= A_{iGs_t} - A_{iG-1s_{t-1}} \\ &= X_{iGs_t}\beta_G + S_{Gs_t}\delta_G + \bar{P}_{(-i)Gs_t}\lambda_G + e_{iGs_t}. \end{aligned} \tag{7.2}$$

One still needs good measures of the inputs (X, S, and $\bar{P}$). In the presence of either mismeasured or left-out inputs, the remainder of the estimation is going to be problematic.[7]

By far the most important issue is the specification of school and teacher inputs. The approach that we have pursued is the semiparametric estimation of teacher and school effects. In a simple formulation, consider

$$\Delta A_{iGs_t} = X_{iGs_t}\beta_G + S^*_{iGs_t}\delta_G + \bar{P}_{(-i)Gs_t}\lambda_G + \sum_{j=1}^{N} t_j T_{ijG} + e_{iGs_t}. \tag{7.3}$$

where $T_{ijG} = 1$ if student i has teacher j in grade G and $= 0$ otherwise. $S^*_{iGs_t}$ represents school factors other than individual teachers. In this, we include individual teacher fixed effects, and t_j is a natural measure of teacher quality that is based on effectiveness of individual teachers in raising student achievement.[8]

This formulation circumvents problems of identifying the separate components of teachers but does not necessarily provide unbiased estimates of teacher quality. First, several selection issues related to the matching of teachers and students are important. Because of the endogeneity of community and school choice for families and of administrator decisions for classroom placement, the unmeasured influences on achievement are potentially not orthogonal to teacher quality. In particular, students with family background and other factors conducive to higher achievement will tend to seek out better schools with higher-quality teachers. Administrative decisions regarding teacher and student classroom assignments may amplify or dampen the correlations introduced by such family choices. The matching of better students with higher-quality teachers would tend to increase the positive correlations produced by family decisions, while conscious efforts to place more effective teachers with struggling students would tend to reduce them.

Second, another source of correlation between teacher quality and student circumstances results from the matching of teachers with schools. Teacher preferences for better working conditions and higher-achieving, nonpoor, nonracial/ethnic minority students in addition to higher salaries potentially introduce a positive correlation between teacher quality and family contribution to learning (Hanushek, Kain, and Rivkin 2004b). Note, however, that failure to hire the best available candidates would reduce the magnitude of this relationship (see Ballou 1996). Within districts, the assignment practices tend to give the newest teachers the lowest priority in terms of deciding where to teach.

Our general approach to separating the effects of teachers, discussed with the empirical results, is to remove student, school, and school-by-grade fixed effects. This strategy, made possible by our stacked panel data on performance, provides a very general way of dealing with the severe selection and measurement issues.

7.3.1 Potential Influence

To provide bounds for the potential impact of teacher quality, we can look at two different general estimates of the impact of schools and teachers on student performance. These use very different approaches to estimate quality differences and thus the potential for using direct policy interventions.

Method 1 is the most conservative (Rivkin, Hanushek, and Kain 2005). It focuses entirely on the within-school variation in student performance that is related to teacher-quality differences. Looking just within schools eliminates the potential bias from school selection by students and their parents. By aggregating student-performance gains across classrooms within each grade, the potential impact of purposeful classroom placement is also circumvented. Finally, individual and school fixed effects are removed—allowing for the influences of family and school factors (other than teachers) in a very general manner. It then directly estimates the variance in teacher quality by considering the variance in average student outcomes over grades in each school and the ways that this relates to teacher turnover. Because this approach assumes, for example, no changes in teacher effectiveness across years, because it ignores any between-school variance in quality,

and because of its treatment of measurement error, this approach produces a lower bound on the variance of school quality.

The estimates of teacher quality for teachers of fifth, sixth, and seventh graders indicate that one standard deviation in teacher quality translates into at least a 0.10 standard deviation higher annual growth in student achievement.[9]

Method 2 relies on the direct matching of students and teachers for one large urban district in Texas (Hanushek et al. 2005). By following students and teachers over time, we can estimate the mean achievement gain of students in each classroom. These raw estimates tend to be overestimates because the variance in classroom gains will include a component of measurement error that is possibly amplified by the impact of remaining selection effects and of school organization and leadership.[10] However, by also investigating variations in teacher effectiveness within schools, it is possible to control for selection across schools by students and teachers. Direct estimates of teacher effectiveness come from extracting the common component of teacher differences across years.

Based on the learning across classrooms for teachers in grades four to eight within one large Texas district, we obtain estimates of the standard deviation of teacher quality of approximately 0.15.[11]

These differences in teacher quality ignore differences that might exist in the quality of teachers across districts. Fortunately, we have another method for developing the across-district differences in quality. Our analysis of student mobility (Hanushek, Kain, and Rivkin 2004a) identified the average gains in achievement that accrued from moving to a new district. If moves were generally predicated on seeking out improved schooling for children, the gains would indicate how teachers were distributed across schools. Our estimate of the difference in achievement (.025 standard deviation) is, however, an underestimate of the variation in average quality as many moves will reflect other purposes, such as job location or housing-quality choice.

Combining the within-school and between-school estimates of quality suggests a range of teacher-quality differences of 0.125 to 0.175 standard deviation. In other words, moving one standard deviation across the teacher-quality distribution—say, from the median to the 84th percentile—is associated with differences in annual student-achievement growth of 0.125 to 0.175 standard deviation. This provides an indication of how different teachers and schools can be in terms of annual achievement growth. For any student a run of good or bad teachers would clearly accumulate to yield substantial differences in the level of achievement—a point we return to below.

7.3.2 Current Distribution

The prior estimates characterize how much leverage exists if policies were put into place either to rearrange existing teachers or to alter the hiring of teachers. An alternative perspective is consideration of the magnitude of existing differences across race or income groups. In other words, can the distribution of existing teachers be part of the explanation for currently observed achievement differences?

The easiest summary comes from the large urban district that was employed in the prior estimation. While most discussions of differences in teacher and school quality point to differences across districts, a lower bound on differences would come from looking at just the within-district differences.

Unfortunately, our methodology does not permit reliable assessment of any teacher-quality differences between black and white students. Black students disproportionately have black teachers, so it would be interesting to look at differences in quality of black and white teachers. Yet it is difficult to distinguish between differences in preparation of students taught by black and white teachers and differences in the distribution of teachers across the district. Specifically, within our general methodology, inclusion of student fixed effects can correct for any sorting of students by teachers, but it also restricts comparisons to students who have been observed with both black and white teachers. Thus, there is inherent ambiguity about any quality among teachers by race.[12]

7.4 Peer Influences

The second avenue for schools to influence performance is through the impact of peers in the school. The neighborhood and school determine a circle of friends and acquaintances. If these other students influence attitudes and behaviors, they can directly affect schooling outcomes.

Analysis of peer influences is, nonetheless, very difficult. The difficulty in this is making sure that the observed relationship really reflects the causal impact of peers—and not just other factors that tend to coincide with differences in peers. Three general and significant issues arise in doing this analysis.

First, most studies of the effects of peers rely on data about student outcomes and peer groups that are naturally generated by schools. But these observations of schooling circumstances are the result of the choices of schools (and implicitly peers) that are made by individual families and, to some extent, by school administrators. Thinking initially of the choices of families, which often come through residential location choices, we can be quite certain that they are not random. These choices, while frequently motivated by a number of factors beyond schools, such as incomes or job locations, will reflect the preferences and opportunities facing individual families. This simple fact—that there is a purposeful element in the individual choices of families—implies that some of the outcomes for student performance may result from attributes of families that are unobserved while they enter into their decisions. For example, the parents most motivated about the schooling of their children may both provide the best family environment for learning and pay particular attention to their choices of school location. In such a case, it is frequently difficult to sort out the separate influences on student performance and to identify the impact of peers per se, particularly when parents at a school tend to make similar choices. Similarly, school administrators often make both resource decisions and classroom-composition decisions with some underlying purpose in mind. They might attempt to place their best teachers with students most in need or to group students according to an estimate of their entering abilities.

Second, the ability to distinguish the separate effects of individual and school factors from those of peers depends crucially on observing and measuring the significant inputs into student performance. The typical analysis, however, does not have perfect measures of either family background or of school inputs. For example, from the perspective of family inputs into achievement, researchers typically have just a few crude measures of background available—and often lack even basic characteristics like the education level of parents. Similarly, the details of school quality and school inputs may be known only imperfectly. On the other hand, the consistency of choice of schools across families implies that there is a strong tendency for similar parents to select a common school, and there is an additional likelihood that school quality affects peers in a similar way as the individual student. As a result, measures of peer backgrounds and performance may provide reasonably accurate surrogates for the individual's characteristics

(which are measured with error). Even when peers have no true impact, for example, they may appear significant just because the peer measurements effectively provide additional information about the individual student.

Finally, one must sort out causal influences. It is not sufficient to know that, say, peer characteristics are associated with individual characteristics and performance. One needs to know whether this association results from peer attributes and interactions causing the observed differences in student performance. The reason for this is also straightforward. To ascertain the impacts of peers and of possible alterations in the composition of peers, it is necessary to capture the amount of difference that the peers cause achievement differences as opposed simply to selecting peers with certain characteristics or to residing together because of common decision-making processes. This issue of causation pervades most analyses of student performance but is most acute when analyzing peers.[13] The inherent tendency for peers with similar attributes and motivations to cluster together makes associations of performance across peers very likely and builds in difficulties in inferring the causal aspects of the various associations.[14]

These issues are introduced to underscore the uncertainty that surrounds much of the discussion of peer influences. Our approach throughout this analysis is to exploit our stacked panel data to deal with the significant measurement issues. With the stacked panel data, we can generally remove individual fixed effects, allowing for very general background and ability factors through individual specific growth rates. We also quite generally remove school-by-grade differences in curriculum, leadership, student aging patterns, and so forth—things that might be correlated with the grade level.

7.4.1 Potential Influence

7.4.1.1 Race or Ethnicity The landmark legislatively mandated civil rights report on the *Equality of Educational Opportunity* (Coleman et al. 1966) and its offshoot (U.S. Commission on Civil Rights 1967) provide empirical evidence that racial isolation harms academic achievement. Subsequent work by Crain and Mahard (1978), Boozer, Krueger, and Wolkon (1992), and Grogger (1996) also finds that school racial composition affects academic, social, and economic outcomes. In contrast, Cook and Evans (2000) conclude that desegregation has little if any

effect on mathematics and reading achievement in elementary school, and Rivkin (2000) finds no evidence that exposure to whites increased academic attainment or earnings for black men or women in the high school class of 1982. Overall, there remains considerable disagreement about the nature and magnitude of benefits of desegregation efforts, let alone about their costs (see, for example, the reviews in Crain 1970; Armor 1995; and Schofield 1995).[15]

The contrasting findings and lack of consensus concerning the importance of school racial composition emanate in large part from the difficulty of isolating the causal impact of peer characteristics.

In Hanushek, Kain, and Rivkin (2002), we estimate the impact of racial composition on blacks, whites, and Hispanics in ways consistent with the previous modeling discussion. Specifically, we adopt the very general fixed-effect approach to eliminating the bias from mismeasured other inputs.

We find small and insignificant impacts of school racial composition on whites and Hispanics, but there are strong impacts of the black composition of schools on the performance of blacks. The magnitude of the proportion black coefficient for blacks of −0.25 suggests that a 10 percentage-point reduction in percentage black would raise annual achievement growth by 0.025 standard deviations. These estimated effects apply to the growth of annual achievement and thus accumulate across grades, implying a substantial role for school racial composition in the determination of the racial achievement gap.

7.4.1.2 Socioeconomic Status (SES) Much of the attention to socioeconomic status has concentrated on issues of neighborhood poverty and, particularly, the ways that concentrations of poverty affect individual outcomes. This discussion of neighborhood poverty emphasizes employment and crime outcomes, although some gets into schooling.[16] For example, Mayer (1991) finds that socioeconomic status (and racial composition) of the school affects high-school completion of both whites and blacks—but measures of characteristics of schools other than student-body composition are missing.

The direct analysis of achievement effects of low-income peers (Hanushek, Kain, Markman, and Rivkin 2003) does not indicate that poverty concentrations have a significant negative effect on student performance. While the income measure is relatively imprecise, these results suggest that prior estimation of the effect of poverty concentrations in schools have not uncovered causal influences.

7.4.1.3 Peer Ability The analysis of peer ability and achievement has been particularly problematic from a statistical viewpoint.[17] Students in a common classroom have many shared educational experiences, so that the quality of questions or the amount of disruption affects all of the students. From an analytical viewpoint, each student contributes to the classroom experience and is simultaneously affected by those same experiences. Moreover, common factors such as an impact of a particularly good teacher will heighten the common experiences and, if teacher quality is not well measured, lead to biases in understanding peer influences. These situations make it virtually impossible to separate out the effects of current classroom behavior on individual achievement. The import of this is largest when considering the influence of other students' ability and achievement on learning.

If we distinguish between the ability of peers and their current behavior, however, it is possible to gain some insights. By measuring peer ability by their prior achievement levels, any direct relationship of current interactions, teacher quality, and the like is broken, and it is possible to gain some insights into how the level achievement of other students influences individual performance.

Attempts to estimate peer effects on educational achievement in this way have been relatively limited. Hanushek (1972, 1992) finds no peer-achievement effects when looking at achievement growth in individual classrooms. On the other hand, Henderson, Mieszkowski, and Sauvageau (1976), Summers and Wolfe (1977), and Zimmer and Toma (2000) report positive influences of higher-achieving peers at least for some students. Summers and Wolfe (1977) find stronger effects of peers for low-income students. Consideration of ability tracking in schools likewise has yielded mixed results (e.g., see Oakes 1992; Argys, Rees, and Brewer 1996).

Our own attempt to investigate peer ability yields ambiguous results. Our initial work suggested that the level of achievement of others in the classroom has a small but significant influence on performance (Hanushek et al. 2003). It also suggested that any effect is relatively constant across achievement levels.[18] However, after developing a more detailed description of the racial composition of schools (Hanushek, Kain, and Rivkin 2002), we found no impact of student achievement. In part, our approach aggregates performance across classrooms in a grade—a necessity because of data availability but a useful approach for assessing selection effects. This aggregation may

be particularly important, however, in the case of ability differences, since classroom interactions likely to be a central issue.

In sum, our best estimates do not support a strong influence of peer achievement on learning, but difficulties in the estimation leave some uncertainty.

7.4.1.4 Student Mobility Student moves are associated with lower achievement, but the more interesting impact of mobility is the externality for other students. The relevance of this is that schools with higher mobility rates tend to have a less coherent structure of instruction. The possibility that turnover affects nonmovers as well as movers is raised by many, including Alexander, Entwisle, and Dauber (1996) and Kerbow (1996), though neither study attempts to estimate the turnover externality.

Our estimation again relies on our fixed-effects strategy, removing both school-by-grade and school-by-year terms and then observing how students react to varying amounts of annual mobility (Hanushek, Kain, and Rivkin 2004a). A high mobility rate lessens the amount of learning, even for students who themselves do not move. The magnitude of the coefficient for overall proportion new students in the more complete specifications suggests that a one standard-deviation increase in the proportion of students who are new to the school (an 11 percentage-point change) would reduce achievement by over 0.013 standard deviation. While a single-year effect of this magnitude is not large, the sum total of 10 or 12 years of high turnover will have a substantial cumulative effect on learning for those students who attend high-turnover schools year after year.

7.4.2 Current Distribution

The previous section describes a variety of potential influences on student achievement and, particularly, on racial or economic divisions and provide insights into some of the divisions in performance identified at the beginning. In terms of the peer effects, outcomes for different groups can diverge when there are different reactions to peer inputs or when the distribution of peers differs even with the same impact. This section discusses how the peer factors may or may not contribute to distributional impacts.

The magnitude of the black composition effects is significant and represents both fundamental forces behind peer impacts. First, black students react to racial composition to a much greater extent than whites and Hispanics. Second, the typical black student (regardless of achievement quartile) has 30 percent more black classmates than the typical white and has 25 percent more black classmates than would be obtained with a completely even distribution of blacks across the state. This difference combines with the race-specific impact of composition such that equalizing the black distribution throughout the entire state for just grade five would be consistent with an increase in black achievement growth of 0.06 standard deviation.[19]

School mobility provides the second significant example of peer influences that have direct distributional impacts. The income difference in school turnover rates is 1.5 percentage points, and the black/white difference is 6.2 percentage points. Higher school turnover reduces annual achievement gains for lower-income students by roughly 0.005 standard deviation relative to higher-income students; blacks lose roughly 0.015 standard deviation relative to whites. Hispanics, by a similar calculation, would lose 0.005 standard deviation relative to whites because they attend schools with higher student mobility rates. These annual differences would cumulate as blacks, Hispanic, and low-income students continue to attend high-mobility schools.

On the other hand, there is not much evidence that peer ability or the socioeconomic mixing of schools has much impact. Both of these investigations are, nonetheless, subject to greater uncertainty. The only available measure of socioeconomic status is the imprecise characterization of eligibility for free or reduced-price lunch. Nonetheless, it does not appear that the distribution of students by socioeconomic status has had much impact on the currently observed distribution of student outcomes. Similarly, while there is some uncertainty about the magnitude of any peer-ability impacts, our best estimates indicate that this aspect of peers is not having much influence on the distributional issues.

7.5 School Resources and Other Inputs

While the previous discussion has concentrated on issues of teacher quality as identified by student performance, the traditional perspective on both performance and distribution has focused more on characteristics of schools and teachers. Specifically, an enormous amount of policy attention has gone into analysis of the experience, degrees, and credentialing of teachers along with the class sizes that students face.

The discussion of school inputs has been controversial (see Hanushek 2003). Input characteristics nonetheless remain an important part of the debate for three reasons. First, they are the object of much policy consideration and debate. Second, relative to the distributional discussions here, a frequent hypothesis has been that disadvantaged students, variously defined by income or race, are more sensitive to variations in inputs. Thus, simply ensuring the same level of inputs could have beneficial effects for distributional outcomes. Third, teachers themselves have preferences for the schools at which they teach.[20] In a systematic way, teachers appear to seek out schools with higher-achieving students and fewer minority students.

Part of the controversy about school inputs has related to issues of causality and the possible contamination of unmeasured student and school characteristics. For example, if class sizes are set in a compensatory manner such that more educationally disadvantaged students are placed in smaller classes, one would see a positive correlation of the level of achievement and class sizes. Thus, commonly available estimates of the impact of class size might give a misleading view of the leverage that can be had.

Various approaches have been pursued to circumvent these problems, particularly in the area of class size. These include attempts to isolate exogenous variations in class size (Angrist and Lavy 1999; Hoxby 2000; Woessmann and West 2005). They also include the use of random assignment experiments (Word et al. 1990; Krueger 1999). In each instance, however, the efforts to isolate the causal impact have also been accompanied by other complications having to do with the quality of the underlying data, thus leading to uncertainty about the results.

An alternative within the framework of our work is to control more directly for various influences that might be correlated with the inputs of note to isolate their impact. Consider class-size policy and its potential interaction with estimation. If schools actively decide class size on the basis of student need and if student need is not accurately assessed in the analysis, standard estimation will yield significant bias. Our approach follows the development above. We investigate student-achievement growth, allow for individual specific growth rates through fixed effects, and incorporate generalized measures of school inputs with grade-by-school fixed effects. We then consider how the variations in class size that occur over and above these—largely through demographic variations across time—influence achievement.

Similarly, we investigate other measured teacher and school inputs after allowing for systematic variation in factors affecting achievement growth.

Our investigation of school performance in Texas confirms large parts of the past analyses of inputs but also sheds further light onto the distributional issues here. The analysis in Rivkin, Hanushek, and Kain (2005) suggests four important findings. First, among the traditional measured inputs, the most important is early career teaching experience. Teachers in their first few years of teaching do worse than those later in their careers, with the most important impact during the first year of teaching. In other words, regardless of subsequent performance, rookie teachers on average do more poorly in the classroom than they will later. Second, class size has a significant but very small impact on student performance. Third, there is no evidence that disadvantaged students, identified by parental income, are more sensitive to school inputs than more advantaged students. Finally, other common inputs including teacher degrees, scores on teacher-certification tests, and teacher certification in general do not have a systematic impact on student performance.

Perhaps the most important of these findings from a distributional perspective is the finding about early career performance of teachers. The impact of the initial year of experience appears to be approximately 0.1 standard deviation of student growth (that is, student growth is on average one-tenth of a standard deviation lower during a teacher's first year).[21] This impact is potentially important when put into the context of the mobility of teachers. Since teachers appear to seek out schools with higher achievement levels and lower percentages of disadvantaged and minority students (Hanushek, Kain, and Rivkin 2004b), there is a concern that this induces minority students to face more rookie teachers. In our samples, however, there is only a modest difference in the proportion of teachers with one to three years of experience (0.16 for whites, 0.19 for blacks, and 0.20 for Hispanics), and the net impact is just 0.001 to 0.002 s.d. on the gaps with blacks and Hispanics.

7.6 Distributional Policy

One way to draw together the previous evidence is to summarize the major factors identified both as having an impact on student performance and as potentially entering into the observed distribution of

outcomes. Table 7.2 provides the two dimensions of the key factors discussed above in the dimension of white-black achievement gaps. First, based on both the impact and the distribution of underlying characteristics, there is a rough calculation of how much the annual differences in underlying achievement factors contribute to the relatively higher performance of white students. Second, the potential impact reflects simply the estimate of how "distributionally sensitive" changes in each factor will affect the gap—that is, the potential strength of any policies aimed directly at improving distribution. The previous discussion is meant to highlight the various dimensions of policy choices that affect distributional issues. For example, while the importance of family background has been well understood since the Coleman Report, governmental intrusion into families has never been a substantial part of the policy agenda.[22] On the other hand, altering both the resources and organization of schools and the characteristics of student peers has been on the policy agenda. The suspicion has long been, however, that policy is impotent and that achieving a significant closing of the gaps through policy manipulations is not possible.

Table 7.2 provides the quantitative summary of the consistent estimation of various effects on the racial distribution of student outcomes.[23] The table, which directly follows the previous presentation in the text, gives an immediate picture of where leverage is greatest. Teacher quality can dramatically change student outcomes: a one standard-deviation improvement in teacher quality (measured in terms of variations in average classroom gains) can yield somewhere between 0.125 and 0.175 standard-deviation change in student achievement. (The range of estimates reflects the underlying approach to assessing differences in teacher quality, as described earlier.) Similarly, the racial composition of the school and the student mobility rate in a school have large impacts on distribution—reflecting the different impacts on black and white students.

The prior analysis focuses on the data for distributional considerations. The large and unmistakable variations in performance by race and income have been the object of an enormous amount of concern and policy attention. The full impact of policy interventions aimed at dealing with distributional issues depends on both the magnitude of any policy and its impact on different groups.

A distinguishing characteristic of policies aimed at distribution is the potential interaction with policies aimed at overall performance and efficiency. One class of policies considers simple redistribution of exist-

Table 7.2
Estimated current influence and potential influence on white-black achievement gaps

	Analytical sources	Annual contribution to current gap (white-black gap)	Potential annual impact
Teacher quality:			
Total quality[a]	Rivkin, Hanushek, and Kain (2005); Hanushek et al. (2005); Hanushek, Kain, and Rivkin (2004a)	−.025 to 0.08	0.125 to 0.175
Peer influences:			
Racial composition[b]	Hanushek, Kain, and Rivkin (2002)	0.038	−0.14 to −0.25
Peer SES	Hanushek et al. (2003)	—	—
Peer ability	Hanushek et al. (2003)	—	—
Student mobility[c]	Hanushek, Kain, and Rivkin (2004a)	0.06	−0.18 to −0.3
School resources/other inputs:			
Teacher experience[d]	Hanushek et al. (2005)	0.001	−0.10
Class size[e]	Rivkin, Hanushek, and Kain (2005)	—	0.0 to −0.01 per student
Teacher degree level	Hanushek et al. (2005)	—	—
Teacher certification	Hanushek et al. (2005)	—	—

a. Teacher quality is measured in terms of standard deviations of the teacher distribution, where, for example, 0.1 indicates that one standard deviation of teacher quality implies 0.1 s.d. higher growth in student achievement.
b. Racial composition indicates the impact of a higher proportion of black students on achievement growth of black students, where, for example, −0.14 indicates that 10% more black students (0.1) translates into 0.014 s.d. lower student annual achievement growth. The current gap is calculated as the impact of moving from the existing unequal distribution across schools to an equalized distribution.
c. School mobility indicates the impact of a higher proportion of student moves on student achievement growth, where, for example, −0.18 indicates that 10% higher student turnover (0.1) translates into 0.018 s.d. lower student annual achievement growth. The range on potential impact depends on the level of mobility before and during the school year.
d. Teacher experience is measured by teachers in their first three years of experience. The potential impact indicates that having a first-year teacher is associated with 0.1 s.d. lower student-achievement growth.
e. Class size effects give the change in performance predicted for a one student change in class size. Estimates vary by grade level with the largest impacts for grade four and smallest for grade seven.

ing resources. Thus, for example, if we take the current set of teachers as constant and simply redistribute them on the basis of student characteristics, it suggests the possibility of a zero-sum game: those who get higher achievement are offset by those who get lower achievement. Policies such as these might readily be justified if the existing distribution of resources, say, favors the otherwise advantaged group. They might also be justified if there is no existing inequity in distribution, but there is general agreement to weight the disadvantaged more heavily. The key element, however, is that actions to improve the distribution of outcomes—and the equity between groups—affect others, and thus it becomes a political question.

The example of the current distribution of teachers that favors higher-income white students is one obvious situation. Here policy aimed at achieving a more equitable distribution of teachers may have great political appeal.[24] But the evidence indicates substantial ability to alter the situation. The white-black gap of 0.7 s.d. (table 7.2) could, by our estimates, be eliminated if blacks systematically got teachers one standard deviation above the mean or at the 85th percentile for four to six years in a row. Having a teacher at the 70th percentile for this period would cut the gap in half. Clearly, these would imply substantial improvements in the quality of teachers within our urban district, but the results underscore the point that correcting the gaps is not impossible.[25]

Table 7.2 shows that the current gaps in performance do not result from differences in having inexperienced teachers. On the other hand, policies that simultaneously kept the good teachers in heavily disadvantaged schools and cut down on the necessity of hiring new teachers would be beneficial.

A potentially more fortuitous situation would be one where disadvantaged students were more sensitive to certain inputs than more advantaged students. For example, if disadvantaged students reacted more strongly to small class sizes, a policy of providing smaller class sizes for disadvantaged students would simultaneously meet two objectives—improving overall achievement by obtaining a more efficient distribution of inputs and working to reduce any distributional differences in outcomes. (If more advantaged students reacted more strongly, the distributional issues would be made even larger.)

This situation occurs in two places across the Texas schools: racial composition and school mobility. First, black achievement responds adversely to increased proportions of black students, but neither

whites nor Hispanics are similarly affected. The estimates in table 7.2 show that these effects are truly substantial. The difference in achievement growth given the current distribution of blacks compared to an equal distribution across the entire state is 0.038 standard deviation. This growth difference accumulates across time, suggesting that it is a direct contributor to the existing racial gap. At the same time, it is not entirely clear what can be done about the racial composition of schools from a policy standpoint. Most of the racial concentration in the schools results from black concentrations within certain districts. Within most districts, the distribution of the black population across schools is quite even—the result of school-desegregation actions following *Brown v. Board of Education* (Rivkin 2000). There is no legal basis for moving students across district boundaries (Armor 1995), and even if there were a basis, much of the distribution is also complicated by regional patterns of settlement in Texas. The possibility of opening up housing in suburban areas could accomplish part of this, although the policy consensus needed for such actions is difficult to achieve.

Second, blacks are more sensitive to mobility rates in their schools and also attend schools with higher mobility rates than whites. Therefore, if policies can stabilize the schools for black students, substantial gains could be possible. To date, few policies that try to affect either the level of mobility or the impact of mobility have been developed. Nonetheless, because of the magnitude of these effects, some increased attention would seem warranted.

The summary of this consideration of distributional issues is simple. The large gaps in performance by race and income can be affected by policies. The policies that might work, however, differ substantially from the existing set of common initiatives. Equalizing standard teacher inputs or reducing class sizes for disadvantaged students has little hope of lessening the observed achievement gaps simply because these factors do not systematically affect outcomes. On the other hand, substantial leverage exists through actions to alter the quality of teachers for disadvantaged students. Further, some peer aspects of schools—namely, the racial composition and the levels of student mobility—have substantial impacts on existing gaps and, if the effects could be lessened, offer another avenue for improving equity in the schools. Little policy attention has been given, however, to these aspects of peer composition.

The prior discussion has also taken a "benevolent-dictator" view of policy. If one actually wished to affect any of the changes discussed, it

would be necessary to consider the underlying politics of the situation. How could the changes be accomplished? These are large and truly important issues.

Additionally, the discussion has largely taken the current system—with its operations and possibilities—as given. In a variety of other analyses, however, it has been clear that substantial inefficiency exists (see the summary in Hanushek 2003). An alternative way to view the entire issue revolves around improving the entire system. If, in fact, something could be done to improve the overall performance of the system, policies that also improved the equity would be easier to accomplish. In other words, redistributing a larger pie is generally easier than redistributing a constant-size pie through zero-sum policies.

Although they go beyond the scope of this chapter, it is worth noting some of the choices that have a reasonable chance of improving the schools. The basic notion is to change the incentives that are relevant to the schools. If, contrary to the current situation, rewards are given for improving performance, it is much more likely that we will move in the direction of better results.

The two leading candidates for reform include a combination of improved accountability for school performance and enhanced parental choice of schools (see Hanushek and Raymond 2005; Peterson 2003). These options offer the possibility of spurring innovation and change that provide real improvements in student performance—and thus the possibility that a larger pie can also be used to improve the equity of the system.

Notes

1. Testing is conducted at ages 9, 13, and 17. The trend data employed here are designed to provide a direct summary of how performance changes through time.

2. Scores at age 17 are the product of schooling received over the prior 10 years. Looking at the achievement gaps for 13-year-olds shows that the gains seen during the 1980s for the oldest students have their antecedents in the 1970s. Most recently, the achievement gap for nine-year-olds narrowed in reading and math. Some popular statements have attributed this narrowing to increased national accountability and particularly the introduction of the federal No Child Left Behind Act of 2001. Nonetheless, no formal analyses have yet to be conducted.

3. This act, when most recently renewed, became the No Child Left Behind Act of 2001.

4. Bowles and Levin (1968); Cain and Watts (1970); Hanushek and Kain (1972).

5. Overall state data can be found in U.S. Department of Education (2003).

6. Presentation of achievement solely in terms of school experiences, ignoring preschool experiences, is done solely for expositional ease. Given our estimation strategy, it has no effect on the results.

7. Specifying the underlying achievement relationship in terms of the simple difference in achievement is one of several alternative forms (see Hanushek 1979). This formulation assumes that there is no depreciation of prior effects over time (Rivkin 2005). The primary alternative estimation puts lagged achievement on the right-hand side of the equation. A coefficient on lagged achievement of one indicates that the simple difference model in the text is correct, while a coefficient less than one indicates some depreciation. In estimation that relaxes the form in the text, the qualitative results shown here are very similar, although the precise quantitative results will vary (Hanushek and Rivkin 2006).

8. For previous analyses of this sort, see, among others, Hanushek (1971, 1992), Murnane (1975), Armor et al. (1976), Murnane and Phillips (1981), Aaronson, Barrow, and Sander (2007), and Rockoff (2004). Rivkin et al. (2005) address the various selection factors and provide a lower bound on the variations in teacher quality specified in this way.

9. In the specific estimates, while we concentrate most on math performance, we obtain an estimate of 0.09 s.d. for reading and 0.11 s.d. for math.

10. One important aspect of that analysis is making adjustments for characteristics of the student-achievement tests. The tests concentrate on performance at the lower end. Because of this, it is easier to get large changes in performance at the lower end of the test. For the analysis, achievement gains are standardized to the gains of others within each decile of the test score.

11. The estimation in Hanushek, Kain, O'Brien, and Rivkin (2005) considers estimates obtained from within-school and within-district comparisons. It also concentrates on standardized gains (see previous note). The bound on the estimates presented here translates gains into raw gains and uses the within-district estimates (which include variations across schools for the district).

12. We do find that there are positive effects to matching student and teacher race (Hanushek et al. 2005).

13. For example, it is common to employ income measures to proxy differences in family background that might be important for student learning or other outcomes, but there are serious questions about whether the relevant causal factor is income per se or some other attributes that are related to income (cf. Mayer 1997).

14. An additional problem, that we do not dwell on here, is the reciprocal relationship between the individual student and peers. The underlying idea behind peer influences is that the others in a classroom and school affect the character of learning. But if that is true, then it is natural to believe that the individual student also affects all of her classmates—implying that the direction of causation for any observed association is unclear. This problem, which is crucial in some kinds of analyses, proves to be difficult to deal with in many studies. This issue, sometimes referred to as the *reflection problem*, is described technically in Manski (1993) and Moffitt (2001).

15. The findings in areas other than achievement are even more difficult to characterize, in part because the quality of the underlying research is quite mixed. In reviewing reviews of desegregation effects on nonachievement outcomes, Schofield (1995, 607, 609) concludes that "desegregation has no clear-cut consistent impact" on African American

self-concept or self-esteem and that "the evidence taken as a whole suggests that desegregation has no clearly predictable impact on student intergroup attitudes". While each of these conclusions is heavily qualified, the research makes it clear that the currently available evidence does not indicate that these wider outcomes are places of systematic impact.

16. Discussions of a wide range of issues related to neighborhood-poverty concentrations can be found in Jencks and Peterson (1991), Jargowsky (1997), and O'Regan and Quigley (1999). More recent investigations relying on randomization of people who leave bad neighborhoods can be found in Rosenbaum (1995), Rosenbaum and Popkin (1991), Katz, Kling, and Liebman (2001), and Ludwig, Duncan, and Hirschfield (2001).

17. The chief problem has revolved around the simultaneous determination of achievement by all students in the classroom. Formal statements of the problem can be found in Manski (1993) and Moffitt (2001).

18. A common policy thread has been that low-achieving students benefit from being in classes with high-achieving students but that high-achieving students are unaffected by classroom composition. If this were the case, heterogeneous classroom groupings would provide the best policy because it would maximize performance of low achievers at no cost. This presumption has been challenged, however, suggesting that detracking or tracking is a zero-sum game where losers balance winners (Argys, Rees, and Brewer 1996).

19. When these results are translated into potential national effects, as measured by the national gaps on the National Assessment of Educational Progress (NAEP), it is estimated that past changes in racial composition of U.S. schools could account for a substantial portion—if not all—of the past closing of the racial achievement gap that occurred in the 1980s (Hanushek 2001).

20. Several early analyses suggest that teachers systematically search out schools with a more affluent population (Greenberg and McCall 1974; Murnane 1981). Those analyses motivate the general discussion here.

21. This estimate is obtained from two very different approaches. In the analysis of the lower bound on teacher quality in Rivkin, Hanushek, and Kain (2005), indirect estimation—through considering the impact on the variance in student achievement of teacher turnover and experience—is very consistent with the direct production function estimates. Also, after estimating the year-by-year performance on individual teachers in the large Texas district used in the quality estimation (Hanushek et al. 2005), virtually identical estimates are obtained.

22. At various times, some thought has been given to such ideas as improving the quality of parenting, although there is little evidence that any of these policy initiatives has been very successful.

23. A similar set of calculations using a different estimation approach is found in Hanushek and Rivkin (2006).

24. Even here, complications of alternative policy goals enter. In many U.S. urban areas, upper-income white families have moved out of the central city and into surrounding suburban areas. This movement has put fiscal pressure on cities as their tax bases erode and has led central cities to seek ways to make themselves attractive to middle-income families. Ensuring quality schools is often identified as the most important approach.

25. Note that the district has just 15 percent white students, so it is not feasible simply to move good teachers from whites to blacks. There are insufficient numbers of high-quality teachers currently with white students to yield the gains for blacks.

References

Aaronson, D., L. Barrow, and W. Sander. (2007). "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25(1): 95–135.

Alexander, K. L., D. R. Entwisle, and S. L. Dauber. (1996). "Children in Motion: School Transfers and Elementary School Performance." *Journal of Educational Research* 90(1): 3–12.

Angrist, J. D., and V. Lavy. (1999). "Using Maimondides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2): 533–575.

Argys, L. M., D. I. Rees, and D. J. Brewer. (1996). "Detracking America's Schools: Equity at Zero Cost?" *Journal of Policy Analysis and Management* 15(4): 623–645.

Armor, D. J. (1995). *Forced Justice: School Desegregation and the Law*. New York: Oxford University Press.

Armor, D. J., P. Conry-Oseguera, M. Cox, N. King, L. McDonnell, A. Pascal, E. Pauly, and G. Zellman. (1976). *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools*. Santa Monica, CA: Rand Corp.

Ballou, D. (1996). "Do Public Schools Hire the Best Applicants?" *Quarterly Journal of Economics* 111(1): 97–133.

Boozer, M. A., A. B. Krueger, and S. Wolkon. (1992). Race and School Quality since Brown v. Board of Education. *Brooking Papers: Microeconomics*. M. N. Baily and C. Winston. Washington, DC: Brookings Institution.

Bowles, S., and H. M. Levin. (1968). "The Determinants of Scholastic Achievement: An Appraisal of Some Recent Evidence." *Journal of Human Resources* 3(1): 3–24.

Cain, G. G., and H. W. Watts. (1970). "Problems in Making Policy Inferences from the Coleman Report." *American Sociological Review* 35(2): 328–352.

Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld, and R. L. York. (1966). *Equality of Educational Opportunity* [the Coleman Report]. Washington, DC: U.S. Government Printing Office.

Cook, M. D., and W. N. Evans. (2000). "Families or Schools? Explaining the Convergence in White and Black Academic Performance." *Journal of Labor Economics* 18(4): 729–754.

Crain, R. (1970). "School Integration and Occupational Achievement of Negroes." *American Journal of Sociology* 75(4, Part II): 593–606.

Crain, R. L., and R. E. Mahard. (1978). "Desegregation and Black Achievement: A Review of the Research." *Law and Contemporary Problems* 42(3): 17–53.

Greenberg, D., and J. McCall. (1974). "Teacher Mobility and Allocation." *Journal of Human Resources* 9(4): 480–502.

Grogger, J. T. (1996). "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics* 14(2): 231–253.

Hanushek, E. A. (1971). "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *American Economic Review* 60(2): 280–288.

Hanushek, E. A. (1972). *Education and Race: An Analysis of the Educational Production Process*. Lexington, MA: Lexington Books.

Hanushek, E. A. (1979). "Conceptual and Empirical Issues in the Estimation of Educational Production Functions." *Journal of Human Resources* 14(3): 351–388.

Hanushek, E. A. (1992). "The Trade-off between Child Quantity and Quality." *Journal of Political Economy* 100(1): 84–117.

Hanushek, E. A. (2001). "Black-White Achievement Differences and Governmental Interventions." *American Economic Review* 91(2): 24–28.

Hanushek, E. A. (2003). "The Failure of Input-Based Schooling Policies." *Economic Journal* 113(485): F64–F98.

Hanushek, E. A., and J. F. Kain. (1972). "On the Value of 'Equality of Educational Opportunity' as a Guide to Public Policy." In Frederick Mosteller and Daniel P. Moynihan (eds.), *On Equality of Educational Opportunity*. New York: Random House.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin. (2003). "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18(5): 527–544.

Hanushek, E. A., J. F. Kain, D. M. O'Brien, and S. G. Rivkin. (2005). "The Market for Teacher Quality." Working Paper No. 11154, National Bureau of Economic Research.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin. (2002). "New Evidence about *Brown v. Board of Education*: The Complex Effects of School Racial Composition on Achievement." Working Paper No. 8741, National Bureau of Economic Research.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin. (2004a). "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." *Journal of Public Economics* 88(9–10): 1721–1746.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin. (2004b). "Why Public Schools Lose Teachers." *Journal of Human Resources* 39(2): 326–354.

Hanushek, E. A., and M. E. Raymond. (2005). "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* 24(2): 297–327.

Hanushek, E. A., and S. G. Rivkin. (2006). School Quality and the Black-White Achievement Gap. Working Paper No. 12651, Cambridge, MA: National Bureau of Economic Research.

Henderson, V., P. Mieszkowski, and Y. Sauvageau. (1976). *Peer Group Effects and Educational Production Functions*. Ottawa: Economic Council of Canada.

Hoxby, C. M. (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115(3): 1239–1285.

Jargowsky, P. A. (1997). *Poverty and Place: Ghettos, Barrios, and the American City*. New York: Russell Sage Foundation.

Jaynes, G. D., and R. M. Williams. (1989). *A Common Destiny: Blacks and American Society*. Washington, DC: National Academy Press.

Jencks, C., and P. E. Peterson (eds.). (1991). *The Urban Underclass*. Washington, DC: Brookings Institution.

Jencks, C., and M. Phillips (eds.). (1998). *The Black-White Test Score Gap*. Washington, DC: Brookings Institution.

Katz, L. F., J. R. Kling, and J. B. Liebman. (2001). "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics* 116(2): 607–654.

Kerbow, D. (1996). "Patterns of Urban Student Mobility and Local School Reform." *Journal of Education for Students Placed at Risk* 1(2): 147–169.

Krueger, A. B. (1999). "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114(2): 497–532.

Ludwig, J., G. J. Duncan, and P. Hirschfield. (2001). "Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment." *Quarterly Journal of Economics* 116(2): 655–679.

Manski, C. F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60: 531–542.

Mayer, S. E. (1991). "How Much Does a High School's Racial and Socioeconomic Mix Affect Graduation and Teenage Fertility Rates?" In C. Jencks and P. E. Peterson (eds.), *The Urban Underclass*. Washington, DC: Brookings Institution.

Mayer, S. E. (1997). *What Money Can't Buy: Family Income and Children's Life Chances*. Cambridge, MA: Harvard University Press.

Moffitt, R. A. (2001). "Policy Interventions, Low-Level Equilibria, and Social Interactions." In S. D. and H. P. Y. (eds.), *Social Dynamics*. Cambridge, MA: MIT Press.

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

Murnane, R. J. (1981). "Teacher Mobility Revisited." *Journal of Human Resources* 16(1): 3–19.

Murnane, R. J., and B. Phillips. (1981). "What Do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research* 10(1): 83–100.

National Center for Education Statistics. (2004). *The Condition of Education 2004*. Washington, DC: U.S. Department of Education.

Neal, D. (2006). Why Has Black-White Skill Convergence Stopped? In E. A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Volume 1. Amsterdam: Elsevier.

O'Regan, K. M., and J. M. Quigley. (1999). "Accessibility and Economic Opportunity." In C. Winston, J. A. Gomez-Ibanez, and W. Tye (eds.), *Essays in Transportation Economics*. Washington, DC: Brookings Institution.

Oakes, J. (1992). "Can Tracking Research Inform Practice? Technical, Normative, and Political Considerations." *Educational Researcher* 21(4): 12–21.

Peterson, P. E. (ed.). (2003). *Our Schools and Our Future: Are We Still at Risk?* Stanford, CA: Hoover Institution Press.

Rivkin, S. G. (2000). "School Desegregation, Academic Attainment, and Earnings." *Journal of Human Resources* 35(2): 333–346.

Rivkin, S. G. (2005). Cumulative Nature of Learning and Specification Bias in Education Research, Amherst College.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain. (2005). "Teachers, Schools, and Academic Achievement." *Econometrica* 73(2): 417–458.

Rockoff, J. E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94(2): 247–252.

Rosenbaum, J. E. (1995). "Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program." *Housing Policy Debate* 6(1): 231–269.

Rosenbaum, J. E., and S. J. Popkin. (1991). "Employment and Earnings of Low-Income Blacks Who Move to Middle-Class Suburbs." In C. Jencks and P. E. Peterson (eds.), *The Urban Underclass*. Washington, DC: Brookings Institution.

Schofield, J. W. (1995). "Review of Research on School Desegregation's Impact on Elementary and Secondary School Students." In J. A. Banks and C. A. McGee Banks (eds.), *Handbook of Research on Multicultural Education*. New York: Macmillan.

Smith, J. P., and F. Welch. (1989). "Black Economic Progress after Myrdal." *Journal of Economic Literature* 27(2): 519–564.

Summers, A. A., and B. L. Wolfe. (1977). "Do Schools Make a Difference?" *American Economic Review* 67(4): 639–652.

U.S. Commission on Civil Rights. (1967). *Racial Isolation in the Public Schools*. Washington, DC: Government Printing Office.

U.S. Department of Education. (2003). *Digest of Education Statistics, 2002*. Washington, DC: National Center for Education Statistics.

Word, E., J. Johnston, H. P. Bain, B. D. Fulton, J. B. Zaharies, M. N. Lintz, C. M. Achilles, J. Folger, and C. Breda. (1990). *Student/Teacher Achievement Ratio (STAR), Tennessee's K–3 Class Size Study: Final Summary Report, 1985–1990*. Nashville: Tennessee State Department of Education.

Woessmann, L., and M. R. West. (2005). "Class-Size Effects in School Systems around the World: Evidence from between-Grade Variation in TIMSS." *European Economic Review* 50(3): 695–736.

Zimmer, R. W., and E. F. Toma. (2000). "Peer Effects in Private and Public Schools across Countries." *Journal of Policy Analysis and Management* 19(1): 75–92.