EARNING & LEARNING

HOW SCHOOLS MATTER

SUSAN E. MAYER AND PAUL E. PETERSON, EDITORS

EARNING *and* LEARNING

How Schools Matter

SUSAN E. MAYER
and
PAUL E. PETERSON
Editors

BROOKINGS INSTITUTION PRESS
Washington, D.C.

RUSSELL SAGE FOUNDATION
New York, N.Y.

BROOKINGS INSTITUTION PRESS RUSSELL SAGE FOUNDATION

Earning and Learning: How Schools Matter may be ordered from :
Brookings Institution Press
1775 Massachusetts Avenue, N.W.
Washington, D.C. 20036
Tel. 1-800-275-1447 or (202) 797-6258
Fax: (202) 797-6004
www.brook.edu

Library of Congress Cataloging in Publication Data

Earning and learning : how schools matter / Susan E. Mayer and Paul E. Peterson, editors.
p. cm.
Includes bibliographical references and index.

ISBN 0-8157-5528-7 (cloth : perm. paper)
ISBN 0-8157-5529-5 (pbk. : perm. paper)
1. Education—Economic aspects—United States. 2. Academic achievement—Social aspects—United States. 3. Educational equalization—United States. 4. School improvement programs—United States. 5. Class size—United States. 6. School choice—United States. I. Mayer, Susan E. II. Peterson, Paul E., 1936- .
LC66 .E23 1999
306.43'2—dc21 99-6383
CIP

9 8 7 6 5 4 3 2 1

The paper used in this publication meets minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Typeset in Adobe Garamond

Composition by
R. Lynn Rivenbark
Macon, Georgia

Printed by
R.R. Donnelley and Sons
Harrisonburg, Virginia

Contents

PART TWO
Improving Schooling

ERIC A. HANUSHEK

7

The Evidence on Class Size

NO TOPIC IN EDUCATION has received such public and professional attention as class size. Calls for reductions in class sizes are rallying points for parents, teachers, and administrators across the nation, and politicians have rushed to claim credit for introducing policies aimed at reducing class size. The pupil-teacher ratio in a school district, for example, is frequently used as the fundamental metric for quality, and comparisons across districts become indexes of equity. A prime reason for the attention to class size is that it represents such an extremely convenient policy instrument, one amenable to general political action. A legislature or a court, wishing to alter student outcomes, can easily specify a change in class size, whereas other potential changes are much more difficult to effect.

The rediscovery and promotion of positive findings from experimental evidence has apparently provided sufficient scientific support that legislators can confidently pursue popular initiatives either mandating smaller classes or providing substantial fiscal incentives for reductions.[1] Yet the surprising fact is that the enormous amount of research devoted to studying class size has failed to make a very convincing case that reducing class size

I wish to thank Charles Achilles for comments on a previous version. Financial support was provided by the William H. Donner Foundation and the Smith Richardson Foundation.

1. See, for example, Mosteller (1995) and chapter 6 above; Krueger (1997).

is likely to improve overall student performance. It will increase costs dramatically, but performance is another matter.

Findings of the general ineffectiveness of reducing class size tend to be controversial if for no other reason than that they seem to defy common sense, conventional wisdom, and highly publicized accounts of the available scientific evidence. Unfortunately, in order to support calls for reductions in class size, there has been a tendency to pick and choose among available studies and evidence. Moreover, the close ties between policy and resources available to schools suggest a variety of conflicts of interest. Therefore, it is useful to review the existing evidence and to reconcile the varying conceptions of what might be expected from class size reductions.

I begin this chapter by examining what the aggregate data indicate about the effectiveness of class size policies. Teachers' resources have been used increasingly intensively throughout the twentieth century, so that the current push for smaller classes is more an extension of past policies than something new.[2] Over the period that student achievement data are available (roughly the past quarter century) there have been no discernible improvements in performance, even though there have been large and steady declines in pupil-teacher ratios.

I then review the international data. Across the world, countries run their school systems in surprisingly different ways, including very different pupil teacher ratios. When combined with data on student performance, however, the wide discrepancies in pupil-teacher ratios show little relationship to achievement.

Next, I summarize the extensive econometric evidence about the effectiveness of reducing class sizes. This evidence, which incorporates almost 300 different estimates of the effect of class size on achievement, gives no indication that general reductions in class size will yield any average improvement in student achievement. By separating out the influences of families and other school factors, these studies effectively eliminate the primary interpretative concerns raised with the aggregate data. The lack of evidence for significant class size effects that results from analysis of achievement differences across individual classrooms is particularly persuasive.

2. As discussed below, the only consistent data across time and across countries track changes in pupil teacher ratios, not individual classroom aggregations. In the United States, because of contractual arrangements and management practices, including the use of teachers in a variety of activities outside the regular classroom, pupil-teacher ratios tend to be below average class sizes. Nonetheless, trends in class size will tend to track trends in pupil-teacher ratios. See Lewit and Baker (1997).

I then consider the evidence developed in Project STAR, an experiment conducted by the State of Tennessee in the mid-1980s. This work involves direct comparisons of achievement by students randomly assigned to small classes (thirteen to seventeen students) and large classes (twenty-two to twenty-five students) in kindergarten through third grade. While there is some ambiguity, the overall findings suggest that small kindergarten and perhaps first grade classes might improve initial learning, but that additional resources in later grades did not have a significant influence on subsequent growth in student achievement. This work and related follow-up analyses in Tennessee form much of the scientific basis for the current political debates.[3] Unfortunately, most of the policy discussions go considerably beyond the experimental evidence. In doing so, they tend to ignore the concentration of results in the earliest grades, to generalize to class sizes outside the experiment, and to neglect any consideration of costs relative to potential gains.

In conclusion, I offer possible interpretations for the lack of any general support in the evidence for reducing class size, and then relate the evidence to prospective educational policies.

Basic Aggregate Data

It is common to hear it said that it is not surprising that achievement is what it is, given the large classes that teachers must face. In reality, it is the history of additional teachers without any commensurate increases in student achievement that makes a strong prima facie case about the ineffectiveness of class size policies.

There has been a consistent and dramatic decline in pupil-teacher ratios over most of the twentieth century. Figure 7-1 displays the pattern of pupil-teacher ratios in the United States for the period 1950–94. The overall pupil-teacher ratio fell 35 percent. This decline is the result of steady drops in the pupil-teacher ratio at both the elementary and the secondary levels. The obvious conclusion is that if there is a problem of class size today, there must have been a larger problem in the past.

One closely related trend is that spending per student has grown dramatically over this same period. Because the pupil-teacher ratio indicates the intensity with which school and classroom resources are applied to education, the greater intensity represented by lower instructional ratios translates

3. See Mosteller (1995) for a summary.

is likely to improve overall student performance. It will increase costs dramatically, but performance is another matter.

Findings of the general ineffectiveness of reducing class size tend to be controversial if for no other reason than that they seem to defy common sense, conventional wisdom, and highly publicized accounts of the available scientific evidence. Unfortunately, in order to support calls for reductions in class size, there has been a tendency to pick and choose among available studies and evidence. Moreover, the close ties between policy and resources available to schools suggest a variety of conflicts of interest. Therefore, it is useful to review the existing evidence and to reconcile the varying conceptions of what might be expected from class size reductions.

I begin this chapter by examining what the aggregate data indicate about the effectiveness of class size policies. Teachers' resources have been used increasingly intensively throughout the twentieth century, so that the current push for smaller classes is more an extension of past policies than something new.[2] Over the period that student achievement data are available (roughly the past quarter century) there have been no discernible improvements in performance, even though there have been large and steady declines in pupil-teacher ratios.

I then review the international data. Across the world, countries run their school systems in surprisingly different ways, including very different pupil teacher ratios. When combined with data on student performance, however, the wide discrepancies in pupil-teacher ratios show little relationship to achievement.

Next, I summarize the extensive econometric evidence about the effectiveness of reducing class sizes. This evidence, which incorporates almost 300 different estimates of the effect of class size on achievement, gives no indication that general reductions in class size will yield any average improvement in student achievement. By separating out the influences of families and other school factors, these studies effectively eliminate the primary interpretative concerns raised with the aggregate data. The lack of evidence for significant class size effects that results from analysis of achievement differences across individual classrooms is particularly persuasive.

2. As discussed below, the only consistent data across time and across countries track changes in pupil teacher ratios, not individual classroom aggregations. In the United States, because of contractual arrangements and management practices, including the use of teachers in a variety of activities outside the regular classroom, pupil teacher ratios tend to be below average class sizes. Hence the focus, to make no class size will tend to track trends in pupil teacher ratios. See Lewit and Baker (1995).

I then consider the evidence developed in Project STAR, an experiment conducted by the State of Tennessee in the mid-1980s. This work involves direct comparisons of achievement by students randomly assigned to small classes (thirteen to seventeen students) and large classes (twenty-two to twenty-five students) in kindergarten through third grade. While there is some ambiguity, the overall findings suggest that small kindergarten and perhaps first grade classes might improve initial learning, but that additional resources in later grades did not have a significant influence on subsequent growth in student achievement. This work and related follow-up analyses in Tennessee form much of the scientific basis for the current political debates.[3] Unfortunately, most of the policy discussions go considerably beyond the experimental evidence. In doing so, they tend to ignore the concentration of results in the earliest grades, to generalize to class sizes outside the experiment, and to neglect any consideration of costs relative to potential gains.

In conclusion, I offer possible interpretations for the lack of any general support in the evidence for reducing class size, and then relate the evidence to prospective educational policies.

Basic Aggregate Data

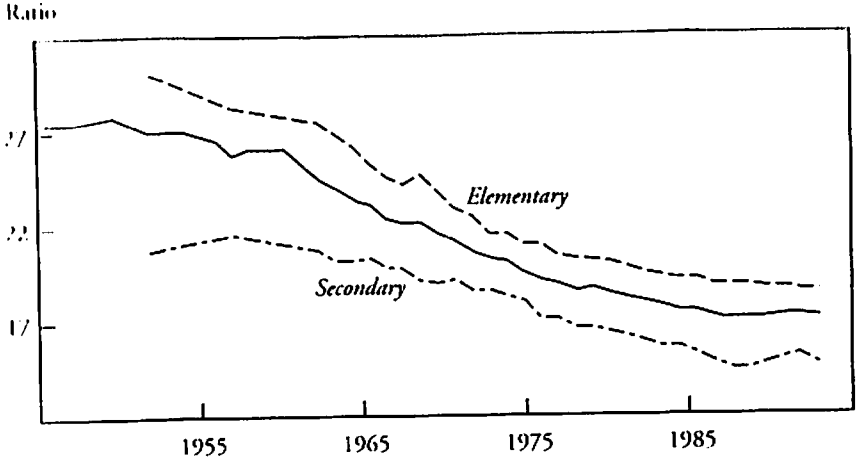
It is common to hear it said that it is not surprising that achievement is what it is, given the large classes that teachers must face. In reality, it is the history of additional teachers without any commensurate increases in student achievement that makes a strong prima facie case about the ineffectiveness of class size policies.

There has been a consistent and dramatic decline in pupil-teacher ratios over most of the twentieth century. Figure 7-1 displays the pattern of pupil-teacher ratios in the United States for the period 1950–94. The overall pupil-teacher ratio fell 35 percent. This decline is the result of steady drops in the pupil-teacher ratio at both the elementary and the secondary levels. The obvious conclusion is that if there is a problem of class size today, there must have been a larger problem in the past.

One closely related trend is that spending per student has grown dramatically over this same period. Because the pupil-teacher ratio indicates the intensity with which school and classroom resources are applied to education, the greater intensity represented by lower instructional ratios translates

3. See Mosteller (1995) for a summary.

Figure 7-1. *Pupil-Teacher Ratios*



Source: National Center for Education Statistics (1997).

directly into greater spending. Figure 7-2 displays real spending per student over the period 1890–1990. The growth in spending amounts to some 3 1/2 percent per year over the entire period, after adjusting for inflation.

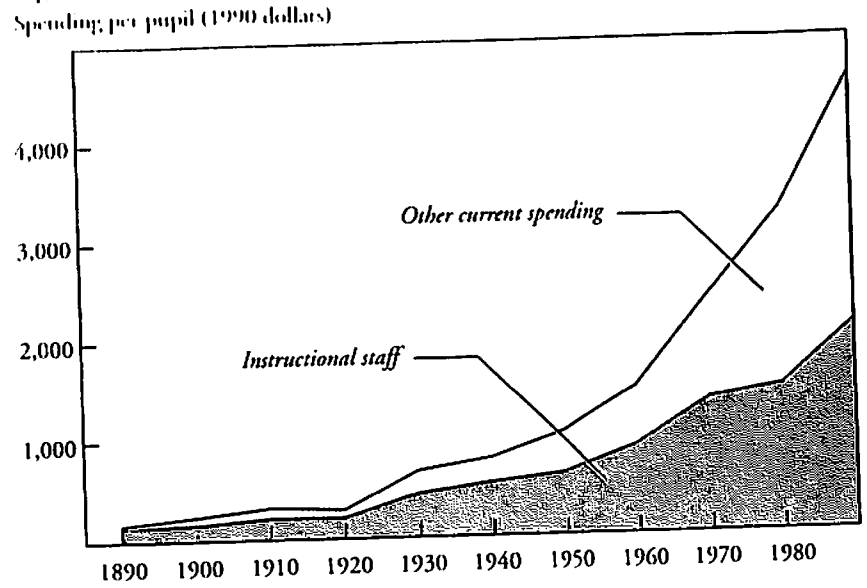
Figure 7-2 breaks down spending per pupil into that related to the salaries of instructional staff and all other spending. Looking at the growth in total instructional staff salaries, one finds that 20 percent of the growth over the century can be attributed to increased intensity of use of instructional staff.[4] This percentage rises in recent periods, reaching 85 percent in 1970–90.[5] In other words, the reductions in pupil-teacher ratios shown in figure 7-1 produce strong effects on spending.

While instructional staff salaries and other spending moved together over the long term, it is also clear from figure 7-2 that nonsalary spending has grown more rapidly in the past two decades. Thus the total growth in spending per pupil is not linked in any simple, mechanical way to pupil-

4. Hanushek and Rivkin (1997).

5. The change in spending over the period 1970–90 is complicated by fact that the school-age population actually declined from the mid-1970s through the mid-1980s. During this period, school systems tended to keep the same number of teachers, leading to declines in pupil-teacher ratios. With the recent increase in student population, however, there has been no tendency for the pupil-teacher ratio to increase. For 1980–90 alone, increased intensity of teachers accounted for 34 percent of the growth in total instructional staff spending (Hanushek and Rivkin, 1997).

Figure 7-2. *Instructional Staff and Other Expenditure, 1890–1990*



Source: Author's calculations.

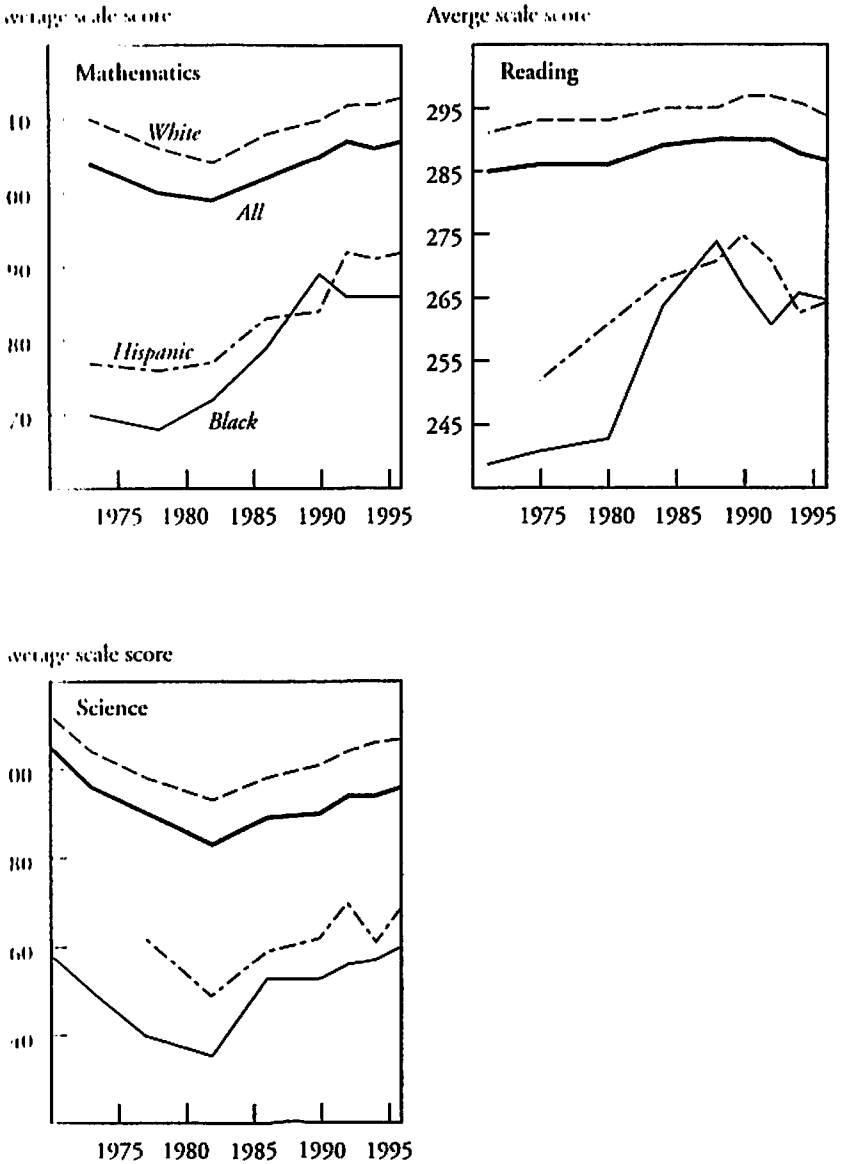
teacher ratios, even though increased intensity of use of instructional staff obviously is an important element.[6]

The other component of the basic aggregate picture is patterns of student performance. While representative student achievement data over the century are not available, the National Assessment of Educational Progress (NAEP) does provide data since the 1970s.[7] Figure 7-3-displays

6. While no systematic analysis is available, it seems plausible that the increased intensity of use of instructional personnel is directly related to parallel increases in noninstructional personnel. At least a portion of the increase in other costs is undoubtedly attributable to various legal changes, including mandates for special education (see below) and desegregation efforts.

7. A longer time series can be constructed from the Scholastic Aptitude Test (SAT), although using those data introduces some added interpretive issues. SAT scores fell dramatically from the mid-1960s until the end of the 1970s—suggesting that the achievement decline in the NAEP data neglects an earlier period of decline. The primary interpretive issue, however, revolves around the voluntary nature of the SAT and the increase in the proportion of high school seniors taking the test. The SAT is taken by a selective group of students who wish to enter competitive colleges and universities. As the proportion taking the test rises, so the hypothesis goes, an increasingly lower achieving group will be drawn into the test, leading to lower scores purely because of changes in test taking. While the exact magnitude of any such effects is uncertain, it seems clear that this change in selectivity has caused some, but not all, of the decline in SAT scores; see, for example, Wirtz (1977); Congressional Budget Office (1986).

Figure 7-3. *Achievement on the National Assessment of Educational Progress for Seventeen-Year-Olds, by Race and Ethnicity, 1970–96*



Source: Word and others (1990).

the patterns of NAEP scores for seventeen-year-olds in mathematics, science, and reading. Scores are shown for all students and by racial and ethnic group. Three aspects stand out. First, overall performance is approximately the same in 1970 as in the mid-1990s. While there are some differences in patterns across the subject areas, the composite picture is one of flat scores over the past quarter of a century. Second, there has been some convergence of scores between whites and either blacks or Hispanics. For the period up to 1990, the average black-white gap across subject areas narrowed by 0.4 standard deviation, even though the differences remain substantial.[8] Third, the convergence of scores across racial and ethnic group may have stopped during the 1990s. In fact, since 1990 there has been a noticeable widening of the racial and ethnic achievement gaps that is not captured in the calculations above.

The challenge is to reconcile the data on pupil-teacher ratios and resources with the data on student outcomes. On the surface, they suggest that increases in the intensity of use of teachers and the commensurate increases in spending have had minimal effect on student achievement. But a variety of alternative explanations have also been suggested.

Changes in the Student Population

One simple explanation for why added resources yield no apparent improvement in performance is that students are more poorly prepared or motivated for school than were students twenty-five years ago, so that the additional resources enable performance to stay even. To bolster this view, some point to the growing numbers of children living in single-parent families and the related increases in child poverty rates—both of which are hypothesized to lead to lower student achievement. Table 7-1 displays these major demographic changes. Between 1970 and 1990, the proportion of children living in poverty families rose from 14.9 to 19.9 percent, while the proportion of children living with both parents declined from 85 to 73 percent. But there have been other trends that appear to be positive forces for student achievement. Table 7-1 also shows that family sizes have fallen and levels of parental education have improved. Over the same period, the proportion of adults aged between twenty-five and twenty-nine with a high school diploma or greater level of schooling rose from 74 to 86 percent (up

8. Hauser and Huang (1996); Hedges and Nowell (1998).

Table 7-1. *Selected Family Characteristics, 1970–90*
Percent

Characteristic	*1970*	*1980*	*1990*
Children in poverty	14.9	17.9	19.9
Children living with both parents[a]	85	77	73
Families with 3 or more children	36.3	22.9	20.1
High school graduate or more[b]	73.8	84.5	85.7

Source: U.S. Bureau of the Census, *Statistical Abstract of the United States*, 1992.
a. Children under eighteen years old.
b. Population aged between twenty-five and twenty-nine.

from 61 percent in 1960). Moreover, among all families with children, the share with three or more children fell from 36 to 20 percent.

It is difficult to know how to net out these opposing trends with any accuracy. Extensive research, beginning with the Coleman report and continuing through today, has demonstrated that differences in families are very important for student achievement.[9] Most of these studies do not focus on families per se, however, and thus have not delved very far into the measurement and structure of any family influences, apparently willing to employ any available measure of family structure or socioeconomic status. Susan Mayer suggests that the direct causal impact of family income might be fairly small, and that previous studies have identified associations rather than true causal impacts.[10] That analysis, nonetheless, cannot conclusively indicate whether there have been trends in the underlying causal factors that are correlated in cross-sections with income. In earlier work I have found that family size may have particularly powerful effects on achievement, and indeed may be partly responsible for narrowing the black-white gap in achievement as indicated in figure 7-3, but again it is difficult to compare the influences of the various trends that have been identified.[11]

David Grissmer and coauthors attempt to sort out these factors.[12] They use econometric techniques to estimate how various family characteristics influence children's achievement. They then apply these cross-sectionally estimated regression coefficients as weights to the trended family background factors identified above. Their overall findings are that black students perform better over time than would be expected from the trends in black family factors. They attribute this performance to improvements in schools. By contrast, white students perform worse than would be expected over time, presumably leading to the opposite conclusion, that for the majority of white students schools got worse over time.

9. Coleman and others (1966); Hanushek (1997).
10. Mayer (1997).
11. Hanushek (1992).
12. Grissmer and others (1994).

Once again, there is reason to be skeptical about these results. First, they do not observe or measure differences in schools, but instead simply attribute unexplained residual differences in the predicted and observed trends to school factors. In reality, any factor that affects achievement, that is unmeasured, and that has changed over their analysis period would be mixed with any school effects. Second, in estimating the cross-sectional models that provide the weights for the trending family factors, no direct measures of school inputs are included. In the standard analysis of misspecified econometric models, this omission will lead to biased estimates of the influence of family factors if school factors are correlated with the included family factors in the cross-sectional data that underlie their estimation. For example, better educated parents might systematically tend to place their children in better schools. In this simple example, a portion of the effects of schools will be incorrectly attributed to the education of parents.[13] Such biased estimates will lead to inappropriate weights for the trended family inputs and will limit the ability to infer anything about the true changes in student inputs over time. Third, one must believe either, contrary to Mayer, that the factors identified are the true causal influences, or that they maintain a constant relationship with the true causal influences.[14]

In sum, a variety of changes in family inputs has occurred over time, making it possible that a portion of the increased school resources has gone to offset adverse family factors. The evidence is nonetheless quite inconclusive about the directions of any trend effects, let alone their magnitudes. At the same time, the only available quantitative estimates indicate that changing family effects are unable to offset the large observed changes in pupil-teacher ratios and school resources. Indeed, for the nation as a whole, these trends are estimated to have worked in the opposite direction, making the performance of schools appear better than it was. Thus the most

13. While it is sometimes possible to ascertain how such statistical misspecification affects the estimated results, in this case the complications—with multiple factors omitted from the modeling of achievement—make that impossible.
14. Mayer (1997).

frequent explanation for the perceived ineffectiveness of historic resource policies does not resolve the puzzle.

Special Education and the Changing Structure of Schools

The discussion until now has focused on pupil-teacher ratios, but these are not the same as class size. Data on pupil-teacher ratios reflect the total number of teachers and students at a given time, but not their utilization. To take a trivial example, consider a district that only has two teachers, one of whom spends all day in class with the students and the other of whom is department head and spends all day evaluating the lesson plans of the classroom teacher. In this case, the pupil-teacher ratio is half that of the class size experienced by the students. More to the point, if teachers are required to meet fewer classes during the day than the number of classes that each student takes, the pupil-teacher ratio will be less than the average class size. Some teachers are also assigned to various duties outside the regular classroom. Thus typical class sizes observed in schools tend to be larger than measured pupil-teacher ratios.

The only data that are consistently available over time reflect pupil-teacher ratios. This is not surprising, because reporting on actual class sizes requires surveying the assignment practices of individual districts. Moreover, class sizes will also be influenced by the range of choices given to students and the number of separate courses that individual students take. Measuring the actual class sizes faced by students requires a variety of decisions about which classes to count and which not to count. For example, should one count physical education and driver's education? Class size is generally best defined in the traditional elementary school grades, where a single teacher is responsible for a self-contained classroom, and the definition gets progressively more problematical as the instructional program becomes more complex.

It remains to be seen, however, how large the influence of any divergence of class size and pupil-teacher ratios might be on the aggregate trends discussed above. In order to influence the trends (as opposed to the observed level during any period), it must be the case that the relationship between pupil-teacher ratios and class sizes is changing over time. For example, Eugene Lewit and Linda Baker show that while class sizes reported in data from the National Education Association are greater than pupil-teacher ratios, they follow a common downward trend.

While the relationship between class size and pupil-teacher ratios could change for a variety of reasons, from altered work days for teachers to expanded curricular offerings, the increased emphasis on special education has received the most attention and deserves examination. The growth in the number of students identified as handicapped, coupled with legal requirements for providing them with educational services, has increased the size of the special education sector since the late 1970s. Since this sector is relatively staff-intensive, its expansion could reduce the overall pupil-teacher ratio without commensurate decreases in regular class sizes. To the extent that mandated programming for handicapped students is driving the fall in the pupil-teacher ratio, regular class sizes may not be declining and, by extension, one might not expect any improvement in measured student performance.[15] This discussion that follows draws on earlier work by Steven Rivkin and me and provides a simple analysis of the potential importance of special education in explaining the decline in pupil-teacher ratios and the commensurate increase in educational expenditure.[16]

Concerns about the education of children with both physical and mental disabilities were translated into federal law with the enactment of the Individuals with Disabilities Education Act in 1976.[17] The act prescribes a series of diagnostics, counseling activities, and services for handicapped students. To implement this and subsequent laws and regulations, school systems expanded staff and programs, in many cases developing entirely new administrative structures. The general thrust of the educational services has been to provide regular classroom instruction where possible ("mainstreaming"), along with specialized instruction to deal with specific needs. The existence of partial categorical funding from outside and of intensive instruction for individual students creates incentives both for school systems to expand the population of identified special education students and for parents to seek admission for their children to special

15. It is frequently asserted that special education students are not generally included in tests and other measures of performance. If so, it would be appropriate to link expenditure on regular-instruction students alone with their test performance. On the performance side, however, if a larger proportion of students is identified as needing special education, and if these are generally students who would perform poorly on tests, the shift to increased special education over time should lead to general increases in test scores, *ceteris paribus*. Direct analysis of special education in Texas indicates that many special education students do take standard achievement tests, although this varies widely by type of disability, due to various state and federal regulations (Hanushek, Kain, and Rivkin, 1998).

16. Hanushek and Rivkin (1997).

17. P.L. 94-142 was originally titled the Education for All Handicapped Children Act.

Table 7-2. *Special Education Student Population and Related School Personnel, 1978–90*
Units as indicated

Year	Disabled children[a] (thousands)	Percentage of elementary and secondary students	Special education personnel (thousands)		
			Teachers	Other instructional	Non-instructional
1978	3,777	8.7	195	140	32
1979	3,919	9.2	203	178	37
1980	4,036	9.7	221	159	56
1981	4,178	10.2	233	167	40
1982	4,233	10.6	235	168	46
1983	4,298	10.9	241	168	57
1984	4,341	11.1	248	173	53
1985	4,363	11.1	275	172	54
1986	4,370	11.1	292	183	47
1987	4,422	11.1	296	175	48
1988	4,494	11.2	301	192	49
1989	4,587	11.4	303	208	48
1990	4,688	11.6	308	220	53

Source: Data are from various annual reports on the Individuals with Disabilities Education Act of 1976. In particular, student numbers are from U.S. Department of Education (1991, p. 4).
a. Aged up to twenty-one years.

education programs.[18] The result has been growth of children classified as special education students, even as the total student population has been falling.

The aggregate changes in the population identified as disabled between 1978 and 1990 are shown in table 7-2.[19] Despite the fact that overall public school enrollment declined by over 1.5 million students between 1980 and 1990, the number of students classified as disabled increased from 4.0 million to 4.7 million. Therefore the percentage of students classified as disabled increased from 9.7 to 11.6 percent during this period. Almost all of this increase represents students classified as having "learning disabilities," a not very well defined category. Moreover, the number of special education teachers increased much more rapidly than the number of children classified as disabled. Table 7-2 shows that the number of special education teachers and other instructional staff increased by over 50 percent between 1978 and 1990: the number of special education teachers rose from 195,000 to 308,000, while the number of other special education instructional personnel (including teacher's aides) rose from 140,000 to 220,000.[20] The number of noninstructional special education staff grew before 1980, but remained roughly constant during the 1980s.

These numbers suggest that the previously noted decline in pupil-teacher and pupil-staff ratios during the 1980s might simply reflect the growth in the number of students receiving special education services and an increase in the intensity of special education—that is, a decrease in effective pupil-teacher ratios for special education. While it is not possible to calculate directly the intensity of special education, since many of the students classified as disabled attend regular classes for much of the day, the maximum impact of the special education changes on overall pupil-teacher and pupil-staff ratios can be estimated. Specifically, by assuming historic values for special education students, instructional staff, and classroom teachers, one can roughly approximate the impact of the growth in special education on the overall ratios. Results are reported in table 7-3.

As shown in the table, the actual pupil-teacher ratio, counting all students and teachers, falls from 19.1 in 1980 to 17.2 in 1990, a decline of 10 percent. The next column estimates what the overall pupil-teacher ratio would have been in 1990 if the observed special education pupil-teacher ratio had remained at its 1980 level instead of falling. The right-hand column shows what the 1990 pupil-teacher ratio would have been if, additionally, the proportion of students classified as disabled had remained at its 1980 level instead of climbing. The simulations indicate that most of the fall in the pupil-teacher ratio during this period was not caused by the expansion of special education. If the proportion of students classified as disabled and the observed pupil-teacher ratio for special education had remained constant, the aggregate pupil-teacher ratio would have fallen at

18. See Hartman (1980); Monk (1990).

19. Data on special education come from annual reports required under the Individuals with Disabilities Education Act of 1976. Prior to the act, there were no consistent data on handicapped students or their schooling.

20. Precise accounting for special education personnel is frequently difficult, suggesting that these data contain more error than the other aggregate data presented.

Table 7-3. *Estimated Effects of Changes in Special Education on Pupil-Teacher Ratios, 1980 and 1990*

		Overall pupil-teacher ratio if the following remained at 1980 levels:	
Year	*Actual pupil-teacher ratio*	*Special education pupil-teacher ratio*	*Special education pupil-teacher ratio and proportion of disabled students*
1980	19.1	19.1	19.1
1990	17.2	17.6	17.9

Source: Author's calculations as described in text.

least to 17.9. In other words, just over one-third of the fall in the pupil-teacher ratio could possibly be attributed to increases in special education. This calculation reflects an upper bound on effects. Importantly, the graphs of pupil-teacher ratios and survey information on class size reported by Lewit and Baker do not appear to show divergences in the two trends that could be attributable to the introduction of special education.[21]

The overall conclusion is that special education could have had some significant effect on pupil-teacher and pupil-staff ratios, but that much more has also been going on during recent times. In terms of the basic issue of flat student performance over recent decades, however, there have clearly been reductions in class size and increases in the resources available for regular classrooms.

Black-White Differentials

An alternative interpretation of the basic aggregate trends follows the observation that there has been a narrowing of the racial gap in NAEP performance, particularly during the 1980s. A variety of commentators have taken this as evidence that school resources have an important effect. They point, in particular, to the increase in federal compensatory programs during the 1970s and 1980s, including Title 1 and Head Start. Since these programs are aimed at disadvantaged students, and since blacks and Hispanics are disproportionately disadvantaged, the narrowing of the differential merely reflects the increased resources.

21. Lewit and Baker (1995).

One problem with this argument, however, is the magnitude of specific programs for the disadvantaged. At the federal level, compensatory education spending amounted to just $7 billion in 1995. These programs target poor students of all racial and ethnic groups, not merely minorities. Moreover, the amount is relatively small compared with total spending on elementary and secondary schools, which exceeds $300 billion.[22]

Michael Cook and William Evans analyze the black-white achievement differential using the panel of NAEP data.[23] They attempt to decompose the differences in performance into family, school, and other factors. Their analysis indicates that school resources and specific schoolwide factors cannot account for the narrowing of the gap. In related analysis, Jeffrey Grogger analyzes the effects of specific school resources on black-white differences in earnings.[24] He also concludes that school resources have not had a significant effect on these differences. Furthermore, there is no indication that pupil-teacher ratios have a significant effect on students' subsequent earnings.

Summary

The available evidence and data suggest some uncertainty about the underlying relationships among families, school organization, class size, and achievement. Allowing for changes in family background and in special education, however, it remains difficult to make a case for reduced class size from the aggregate data. Pupil-teacher ratios and, as best one can tell, average class sizes have been falling significantly for a long time. The increases in teacher intensity have been large enough that discernible impacts on average student performance might reasonably be expected. Nonetheless, overall achievement data do not suggest that it has been a productive policy. The aggregate data are, however, quite limited, restricted to a small number of performance observations over time, and it has been difficult to rule out conclusively other fundamental changes that might affect school success. Therefore it is useful to turn to other evidence, including more detailed school-level information.

22. The federal government has other relevant programs: Head Start added $3.5 billion and child nutrition was $7.6 billion. Such expenditures, even if included in the totals for elementary and secondary spending, still yield small relative total spending.

23. Cook and Evans (1996).

24. Grogger (1996).

International Evidence

Somewhat surprisingly, similar kinds of results concerning the relationship between pupil-teacher ratios and student performance are found across countries. While it is difficult to develop standardized data that control for the many differences in populations and schools across countries, international comparisons have some appeal. The international variations in class size and pupil-teacher ratios are larger than those found within the United States, offering some hope that the effects of alternative intensities of teacher usage can be better understood. Even given these wide differences, there is no evidence that lower pupil-teacher ratios systematically lead to increased student performance.

The Third International Mathematics and Science Study was conducted during 1995. A series of mathematics and science tests were given to students in voluntarily participating nations. As a simple exercise, the eighth grade math and science scores can be correlated with the primary school pupil-teacher ratio in each country.[25] For the seventeen nations with consistent data, there is a positive relationship between pupil-teacher ratios and test scores, and it is statistically significant at the 10 percent level for both tests. When Korea, the sampled country with the largest pupil-teacher ratio, is left out of the analysis, the statistical significance disappears but the positive result remains. Nonetheless, while this international evidence points to the surprising result that performance is better when there is less intensive use of teachers, there cannot be much confidence that such differences are more than statistical artifacts.

A more systematic attempt to investigate the relationship between student performance and pupil-teacher ratios draws on the six prior international tests in math or science, conducted between 1960 and 1990.[26] Using seventy country-test-specific observations of test performance, this analysis finds a positive but statistically insignificant effect of pupil-teacher ratios on performance, after allowing for differences in parental schooling. Again, while there are very large differences in pupil-teacher ratios, they do not appear significantly to influence student performance.[27]

25. Test scores are reported in Beaton and others (1996a, 1996b). Primary pupil-teacher ratios for public and private schools are found in Organization for Economic Cooperation and Development (1996).

26. Hanushek and Kim (1996).

27. At the same time, differences in test performance are extraordinarily important in determining differences in national growth rates (Hanushek and Kim, 1996).

Finally, while uniform data on class size differences are not available, some investigations have shown that class sizes vary more than pupil-teacher ratios across countries. Specifically, Japan and the United States have quite similar pupil-teacher ratios, but because of decisions about how to organize schools and to use teachers, Japanese classes are much larger than U.S. classes.[28] Student performance in Japan is, on average, much better than in the United States. There are many differences in the schooling and societies of the sampled nations, so it would be inappropriate to make too much of these results. They do, however, underscore the fact that common presumptions about the achievement effects of pupil-teacher ratios and class size are not supported by the evidence.

Econometric Evidence

The most extensive information about the effects of class size comes from attempts to estimate input-output relationships, or production functions, for schools. The investigation of the effects of school resources began in earnest in the late 1960s, after the publication of the Coleman report.[29] This congressionally mandated study by the U.S. Office of Education startled many by suggesting that schools do not exert a very powerful influence on student achievement. Subsequent attention was directed at both understanding the analysis of the Coleman report and additional evidence about the effects of resources.[30]

Over the past thirty years, a steady stream of research has built up a consistent picture of the educational process. The following summary concentrates on a set of published results available through 1994.[31] The basic

28. Stevenson and Stigler (1992).

29. Coleman and others (1966).

30. These analyses suggest serious flaws in the statistical methodology and interpretation of the Coleman report, but most of that discussion is not relevant for the present purposes; see Bowles and Levin (1968), Cain and Watts (1970), Hanushek and Kain(1972).

31. See Hanushek (1997) for greater detail. The tabulations do include results in Hanushek, Rivkin, and Taylor (1996), since this updating was conducted as part of that research. Some analyses have been published subsequently—for example, Betts (1995), Ehrenberg and Brewer (1995), Lamdin (1995), Staley and Blair (1995), Ferguson and Ladd (1996), Grogger (1996), and Wenglinsky (1997)—but their results will not affect the overall conclusions here (see Hanushek, 1997). While they have some unique features, these studies are conceptually similar to prior studies and can be considered within the context of prior results. Moreover, given the number of studies sampled, a few additional results could not affect the overall conclusions reported here even if they uniformly pointed in the same direction, which they do not.

studies include all available that meet minimal criteria for analytical design and reporting of results. Specifically, the studies must be published in a book or journal (to ensure a minimal quality standard), must include some measure of family background in addition to at least one measure of resources devoted to schools, and must provide information about statistical reliability of the estimate of the effect of resources on student performance.

The summary relies on all of the separate estimates of the effects of resources on student performance. For tabulation purposes, a "study" is a separate estimate of an educational production function found in the literature. Individual published analyses typically contain more than one set of estimates, distinguished by different measures of student performance, different grade levels, and frequently entirely different sampling designs. If, however, a publication includes estimates of alternative specifications employing the same sample and performance measures, only one of the alternative estimates is included.[32] Thus the ninety individual publications that form the basis for this analysis contain 377 separate estimates of production functions. While several of the studies were produced in more or less immediate reaction to the Coleman report, half of the available studies were published since 1985.

These econometric estimates relate class size or teacher intensity to measures of student performance, while also allowing for the influence of family and other inputs into education. The precise sampling, specification of the relationships, measurement of student performance, and estimation techniques differ across studies, but here I concentrate on the summary of a relationship across studies. To do this, studies are aggregated according to the estimated sign and statistical significance of the relationship.[33]

Table 7-4 summarizes the available results for estimates of the effects of teacher-pupil ratios on student outcomes. Of the total of 377 available

32. Some judgment is required in selecting among the alternative specifications. As a general rule, the tabulated results reflect the estimates that are emphasized by the authors of the underlying papers. In cases where this rule did not lead to a clear choice, the tabulation emphasizes statistically significant results among the alternatives preferred by the original author. An alternative approach is followed by Betts (1996), who aggregates the separate estimates of a common parameter in a given study.

33. More details about the methodology and the available studies can be found in Hanushek (1979, 1997). Some controversy also exists about the best way to summarize the results of different studies, but these issues have little bearing on the present discussions; see Greenwald, Hedges, and Laine (1996) and Hanushek (1996a, 1997). Other issues concerning estimation strategies are raised in Card and Krueger (1996), Heckman, Layne-Farrar, and Todd (1996), and Hanushek (1996b). These latter issues, while relevant, are very technical and in my opinion do not affect the policy conclusions presented below.

Table 7-4. *Distribution of Estimated Influence of Teacher-Pupil Ratios on Student Performance, by Level of Schooling*

Percent, except as indicated

School level	Number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
All	277	15	13	27	25	20
Elementary	136	13	20	25	20	23
Secondary	141	17	7	28	31	17

Source: Author's calculations. Sources of underlying data are detailed in Hanushek (1997).

econometric studies of the determinants of student performance, 277 consider teacher-pupil ratios. (Estimates of the effect of class size or pupil-teacher ratios are reversed in sign, so that conventional wisdom would call for a positive effect in all cases.) The first row of the table shows that just 15 percent of all studies find a positive and statistically significant relationship between teacher intensity and student performance. At the same time, 13 percent of all studies find negative and statistically significant relationships with student performance. Ignoring the statistical significance, or the confidence that there is any true relationship, the estimates are almost equally divided between those suggesting that small classes are better and those suggesting that they are worse.[34] This distribution of results is what one would expect if there was no systematic relationship between class size and student performance. Fully 85 percent of the studies suggest either that fewer teachers per student are better (that is, they yield negative estimates) or that there is little confidence of any relationship at all (that is, they are statistically insignificant).

Some people have suggested that the effect of class size may differ at different points in the schooling process (see, for example, the interpretations of Tennessee's Project STAR study discussed below). To consider this possibility, the overall estimates of the effects of teacher-pupil ratios are divided into estimates for the elementary and secondary school levels. Table 7-4 shows that there is little difference between the estimated effects in elementary and in secondary schools, but, if anything, there is less support for increasing teacher-pupil ratios at the elementary level. For elementary

34. Twenty percent of the studies do not report the sign of any estimated relationship. Instead, they simply note that the estimates were statistically insignificant.

schools, more estimated effects are negative, that is, indicate that smaller classes are worse. This is true for all studies and for those with statistically significant estimates. Unfortunately, there are too few studies to permit one to look at individual grades, as opposed to all elementary grades combined.

With these data, it is also possible to address explicitly the distinction between the pupil-teacher ratio and class size. While these two concepts differ, they are highly related, and it remains to be seen whether they provide similar evidence about the effectiveness of smaller classes. The estimates include both studies that measure class size and those that contain aggregate measures of teacher-pupil ratios for a school, district, or state. Studies that investigate performance within individual classrooms invariably measure class size, whereas those at higher levels of aggregation most often measure average teacher-pupil ratios. In particular, studies that are highly aggregated, such as those investigating performance across entire districts or states, are almost always forced to consider only the overall teacher-pupil ratio.

Table 7-5 displays the results of estimates according to the level of aggregation of the teacher-pupil measure. As the table shows, analyses conducted at the state or district levels are more likely to indicate that teacher-pupil ratios have a positive and statistically significant relationship to student performance. Nonetheless, while this pattern coincides with the less precise measure of class size at the classroom level, it is more likely to come from other fundamental analytical problems than from the pure measurement issues. The more aggregated analyses are subject to a series of specification problems (independent of the ones considered here) that are exacerbated by the aggregation of the analysis.[35] In particular, these analyses leave out all consideration of state-by-state differences in school policies, which appears to bias the results toward finding stronger effects of teacher-pupil ratios and school resources in general. At a minimum, one can conclude that the insignificance of the results appears real and is not just an artifact of measuring teacher-pupil ratios instead of actual class size. The best studies, with the most precise measurements of class size and school resources, arrive at the same general conclusions, indicating that the results are not easily explained away by poor data or research methods.

Statistical investigations that employ a value-added specification are generally regarded as conceptually superior and likely to provide the most reliable estimates of education production functions. These studies relate

35. See Hanushek, Rivkin, and Taylor (1996).

Table 7-5. *Distribution of Estimated Influence of Teacher-Pupil Ratios on Student Performance, by Aggregation of Resource Measure*

Percent, except as indicated

Aggregation	Number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
Total	277	15	13	27	25	20
Classroom	77	12	8	18	26	36
School	128	10	17	26	28	19
District	56	21	16	39	20	4
County	5	0	0	40	40	20
State	11	64	0	27	9	0

Source: Author's calculations. Sources of underlying data are detailed in Hanushek (1997).

an individual's current performance to his or her performance at some prior time and to school and family inputs during the intervening period. The superiority of this approach comes from the use of prior achievement to mitigate any problems arising from missing data about past school and family factors and from differences in the innate abilities of students.[36]

Table 7-6 provides a summary of value-added results, both for all seventy-eight estimates of class-size effects and for the twenty-three estimates that come from samples in a single state. On the one hand, there are many fewer such estimates than in the overall set, and thus any conclusions are subject to more uncertainty. On the other hand, because of the superiority of these analyses, each study deserves more weight than one of the general studies reviewed above. The restriction to samples within single states corrects for differences in school policies to avoid the biases previously discussed. From the results in table 7-6, there is little reason to believe that smaller classes systematically lead to improvements in student achievement. Of the best available studies (single-state, value-added studies of individual classroom achievement), only one out of twenty-three, or 4 percent,

36. For a more thorough discussion of estimation approaches, see Hanushek (1979). Krueger (1997) points out that value-added studies cannot identify any impact of resources prior to the period of study. While he argues from this that one should not use value-added models, the very essence of such models is to provide more precise estimates of the magnitude and timing of any resource effects on student performance. Statistical models of the level of student performance that do not include measures of prior achievement will tend to be biased by all time-varying historical factors, such as prior class size or teacher quality, that are not explicitly measured.

Table 7-6. *Distribution of Other Estimated Influences of Teacher-Pupil Ratios on Student Performance, Based on Value-Added Models*
Percent, except as indicated

School level	Number of estimates	Statistically significant		Statistically insignificant		
		Positive	Negative	Positive	Negative	Unknown sign
All	78	12	8	21	26	35
Within a single state	23	4	13	30	39	13

Source: Author's calculations. Sources of underlying data are detailed in Hanushek (1997).

finds that smaller classes have a statistically significant positive effect on student performance.

As noted by Alan Krueger, if the effects of class size on performance are small, a number of the reported econometric studies may not have adequate data to distinguish between "small effect" and "no effect," possibly leading to the pattern of statistically insignificant results reported.[37] Preliminary analysis of achievement data for several cohorts of students in the State of Texas provides partial support for this hypothesis.[38] From over 300,000 observations of gains in student performance across the schools of Texas, statistically significant positive results are found for smaller classes in fourth and fifth grade mathematics and reading performance.[39] Even with such large samples, however, class size is a statistically insignificant determinant of sixth grade performance in either subject. More important, the estimated magnitudes are very small. A class size reduction of ten students, which cuts average class size approximately in half and represents a 2 1/2 standard deviation movement, is never estimated to yield more than 0.11 standard deviation of improvement in student achievement for the results that are statistically significant. When one separates out the results for students eligible for free or reduced lunches, the performance of these disadvantaged students is found to be more sensitive to class size: a targeted

37. Krueger (1997).

38. Rivkin, Hanushek, and Kain (1998).

39. The analysis of student performance is based on complicated statistical analysis involving regression of achievement growth with individual fixed effects, a design employed to eliminate both unmeasured differences in individual ability and the potential effects of student selection into specific schools. The small results reported here are also consistent across alternative estimation strategies based on simple models of achievement growth.

class size reduction of ten students could yield as much as 0.18 standard deviation of improvement in fourth grade math performance. Estimated class size effects for students ineligible for free or reduced lunch, however, are less than half as big as those for disadvantaged students and are more frequently insignificant.

A final set of questions about the econometric studies of teacher-pupil ratios involves the underlying mechanism for establishing small and large classes. If, for example, a school district uses a subjective method of assigning "weaker" students to small classes and "stronger" students to large classes, the econometric methods might not provide an accurate assessment of the direct, causal influence of class size.[40]

The econometric evidence is clear. There is little reason to believe that smaller class size systematically yields higher student achievement. While some studies point in that direction, an almost equal number point in the opposite direction. Moreover, restricting attention to the best of these studies, including those with the most accurate measurement of individual class size, merely strengthens the overall conclusion.

Project STAR

In the mid-1980s, because of ambiguity about the effects of class size on student performance, the State of Tennessee launched a random assignment experiment on the effects of reducing class size on student achievement called Project STAR.[41] The design was heavily influenced by an early summary of research by Gene Glass and Mary Lee Smith.[42] Glass and

40. Hoxby (1998) employs other information about the source of decisions about class size in order to correct for such problems, but she finds that class size still has no consistent effect on student outcomes. Note, moreover, that this problem arises only when decisions are made on the basis of unmeasured student characteristics. If, for example, students are assigned to classes on the basis of their early test scores, and if these test scores are controlled for in the econometric analysis as in the value-added estimation, these problems do not arise. The statistical analysis in Rivkin, Hanushek, and Kain (1998) provides an alternative approach to the selection problems. In another attempt to correct for possible influences of school decisionmaking, Angrist and Lavy (1999) do find significant class size effects. Their study considers special features of Israeli law that permit alternative statistical approaches to identifying small class effects. The applicability to U.S. schools is unclear. An alternative instrumental variable approach is found in Akerhielm (1995). Taken together, these analyses do not suggest that the prior results are merely statistical artifacts.

41. STAR stands for student-teacher achievement ratio. See Word and others (1990); Finn and others (1990); Finn and Achilles (1990).

42. Glass and Smith (1979).

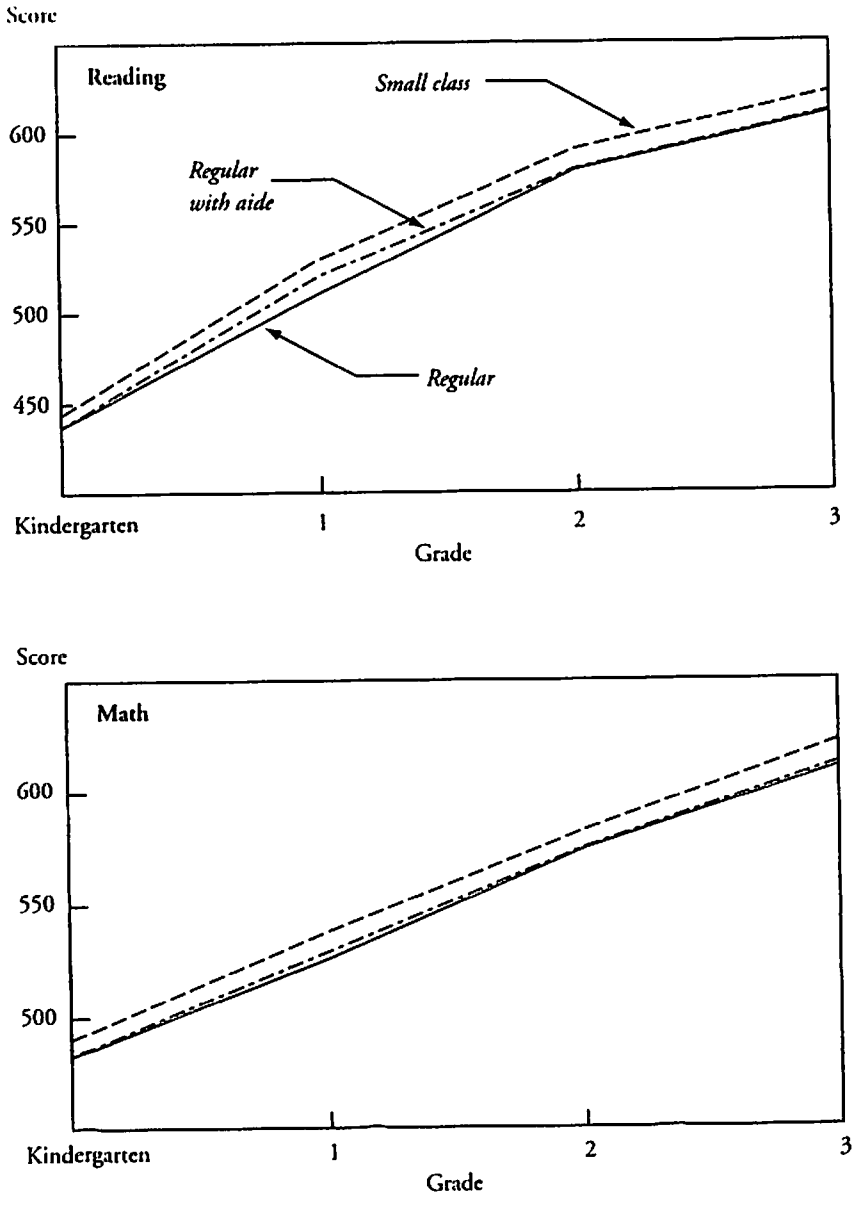
Smith combined evidence from different experimental studies and suggested that student achievement is roughly constant across class sizes until the pupil-teacher ratio gets down to approximately 15 to 1. Beyond this, reductions in class size appeared to yield significant gains in student performance. Project STAR is a large and complicated program. Beginning in 1985, a group of students from kindergarten through third grade was randomly assigned to either regular classes of twenty-two to twenty-five students or small classes of thirteen to seventeen students. The regular classes were divided into two groups, one with teacher's aides and one without. To be eligible to participate in the experiment, a volunteer school had to be large enough to ensure that there was at least one small and one large class of each type. The experimental design placed students in the same randomly assigned treatment group from kindergarten through third grade—although, as discussed below, attrition, experimental replacement, and change of treatment group were considerable over the course of the experiment. Over 11,000 students in seventy-nine schools eventually participated in the program; 48 percent of the 6,324 original students who started in kindergarten remained in the experiment for the full four-year period.

A variety of natural and design factors, discussed below, introduce uncertainty into the analysis and interpretation of the results of Project STAR. The most obvious issue is that many popular interpretations of the results are not supported by the basic data. Most important, even though the experiment ran through the third grade, any beneficial impact of small classes on achievement is confined to the first year of treatment.

The basic data from Project STAR indicate that, at best, very specific and limited achievement effects might follow from reductions in class size. Figure 7-4 illustrates the basic achievement results of the program. The two panels of the figure plot average reading and math scores from kindergarten through third grade for students randomly assigned to the three different classroom situations.[43] Three facts are immediately obvious from these graphs. First, for both reading and mathematics, students in small classes have significantly greater average achievement at the end of kindergarten. Second, on average, the performance of students in regular classes is virtually the same as that of students in regular classes with aides

43. For a more complete description of the experiment, see Word and others (1990). A series of different tests was given. Figure 7-4 reports results from the nationally normed Stanford Achievement Test. Using data from the alternative tests would not change the patterns or conclusions.

Figure 7-4. *Stanford Achievement Test Scores in Project STAR*



Source: Word and others (1990).

throughout the experiment (and students were rather freely reassigned across these treatments after the first year of the experiment). Third, the kindergarten gap between small and regular classrooms is maintained at essentially the same level through the first, second, and third grades.[44]

The original analyses of the experiment reported the differences in performance between the small and regular classes at each grade, giving the impression that the added classroom resources at each grade level led to significant gains in achievement in each grade. This interpretation is reinforced by the widely cited review of the study's findings by Harvard statistician Frederick Mosteller. He writes, for example, that "after four years, it was clear that smaller classes did produce substantial improvement in early learning and cognitive studies."[45] In reality, the differences appeared in the first year of the experiment and simply reappeared in subsequent years.

For policy purposes, the key to the interpretation of Project STAR involves expectations about student performance over time. Perhaps the most standard interpretation from learning theory begins with the view that education is a cumulative process, building on past achievement. If students learn certain skills in, say, the first grade, they tend to carry them over to later grades, even if with some depreciation, and to build on this base as they progress. According to this view, the basic evidence of Project STAR suggests that while smaller classes may be important at kindergarten, they have no average effect subsequently. Specifically, since the growth in achievement across experimental and control students is the same from first through third grades, the added resources of the smaller classes appear to add nothing to student performance. Early differences remain the same over time. If resources had a continuing impact, one should observe a greater disparity of achievement as more and more resources are applied. The achievement curves in figure 7-4 should fan out if smaller classes have an on-going, cumulative impact. Thus the cumulative effect learning model is rejected.

Alternatively, some have argued that the observed pattern could be consistent with small classes making a difference in all grades if students are

44. There is some ambiguity about the effect of small classes in first grade. The Project STAR results in Word and others (1990) indicate some faster growth in achievement during the first grade for those in small classes, and slower growth from first through third grades for students in small classes. But this finding is complicated by the lack of universal kindergarten in Tennessee and the introduction of new students into the experiment in the first grade. Krueger (1997) demonstrates that the overall results are consistent with small classes having an effect only in the first year of schooling.

45. Mosteller (1995, p. 113).

expected to fall back to a common mean performance each year. This is equivalent to saying that educational performance is not cumulative. Under this set of expectations, maintaining the difference in performance at the end of kindergarten requires the continued infusion of resources. Lowered class size might be effective if it stemmed the "inevitable" reversion of achievement to lower levels when resources were removed. Such interpretations are most common in discussions about the education of disadvantaged students, since they provide (a largely untested) way of explaining achievement reversions from initial gains in Head Start and other early childhood programs for the disadvantaged.[46] In this context, it is important to remember that Project STAR was a program for the broad spectrum of Tennessee children. Therefore interpretation of its results must be consistent with underlying notions of the learning process for all children, not just the disadvantaged.

In the presence of these alternative interpretations, the way to identify the effects of class size would be to assign randomly some of the experimental children to larger classes after they had been in small classes in the earliest grades. Unfortunately, this was not done systematically within the experiment (even though some movement across treatment groups did occur, in violation of the experimental protocol). However, follow-ups of Project STAR students after they had returned to regular class settings provide important information. The Lasting Benefits Study, which has traced students after Project STAR finished, shows that students from the small classes in kindergarten through third grade maintained most of the prior differences through the sixth grade.[47] Experimental comparisons of small versus regular classes yield effect sizes on norm-referenced third grade tests in reading and math of 0.24 standard deviation and 0.21 standard deviation, respectively.[48] In the sixth grade, three years after the end of any differential resources for the two groups, the effect sizes for comparisons of students previously in small versus regular classrooms were 0.21 standard deviation and 0.16 standard deviation for reading and math, respectively.[49] In other words, the differentials in performance found at kindergarten remain essentially unchanged by third grade after class size reductions of

46. See, for example, Barnett (1992).

47. Nye and others (1993). These results should be treated with caution, because the data have never been made available to other researchers; see Hanushek (1999).

48. Word and others (1990). Effect sizes indicate the differences in average performance for the two groups measured in units of standard deviations of the test; see Mosteller (1995).

49. Nye and others (1993).

one third were continuously applied (see figure 7-4), and they remain largely unchanged by sixth grade, after regular class sizes had been resumed for three years. This latter finding leads to rejection of the fall-back model and indicates that class size reductions after kindergarten have little potential effect on achievement.

A third interpretation is that small classes, particularly early in the schooling process, have a one-time effect on student performance that is not linked to the acquisition of cognitive skills per se. This one-time effect could reflect early training in the "activity of school." Students in small classes learn norms, behavior, and learning patterns that are useful in subsequent years, so that they are able to continue achieving at a higher level. In fact, this last interpretation is the one most consistent with the Project STAR data, ignoring the other possibilities of flaws in the underlying experimental design and data collection.[50] It provides a parsimonious explanation of why there is a one-time but lasting effect of class size reductions in kindergarten. It also has powerful implications for policy.

The most expansive conclusion that can be reached from Project STAR and the Lasting Benefits Study is that they might support an expectation of positive achievement effects from moving to small kindergartens, and maybe small first grades. None of the Project STAR data support a wholesale reduction in class size across grades. Moreover, the achievement results come from large reductions (one-third of the existing regular class size) that make the small classes quite small (fifteen students) compared with most existing classroom situations. The data do not provide evidence about what might happen with lesser changes that take class sizes down to levels above the Tennessee experiment, say, to between eighteen and twenty students. (Recall that the original motivation for Project STAR involved research results suggesting no effects for class sizes greater than fifteen students.) This policy interpretation is quite different from that commonly attributed to analysis of the Project STAR data, which many cite to justify almost any sort of reduction in class size at almost any school grade.

Project STAR and its related programs do support one aspect of the econometric results from Texas noted above: disadvantaged students appear more sensitive to variations in class size than the majority of students.[51] Again, however, disadvantaged students, on average, are not currently in larger classes than more advantaged students, and the effects appear small relative to the costs of programs and alternative policy approaches.

50. Krueger (1997) reanalyzes the Project STAR data and finds that there is a large first-year effect from reduced class size, but little continuing effect. Krueger incorporates the fact that some students begin small classes in the first grade, rather than in kindergarten. If students were not in small kindergartens, he suggests that they achieve gains from small classes in first grade, that is, in their first year with small classes. See also Hanushek (1999).

Up to this point the discussion has taken the reported results at face value, but some aspects of the experimental design and implementation of Project STAR introduce additional uncertainty into the analysis.[52] The most important concerns are as follows. First, not all students started the experiment at the same time, in part because kindergarten is not mandatory or universal in Tennessee, and in part through the replacement of children who dropped out of the experiment. Second, sizable attrition occurred over the course of the experiment, due to mobility and other factors, and this attrition was likely not random. Third, parents, teachers, and schools knew they were part of an experiment, and parental pressure led to part of the experiment being compromised by the reassignment of students: approximately 10 percent of the students in small classes in grades one through three had been in regular-sized classes the previous year.[53] Fourth, no student took an achievement test prior to entering the experiment, so it is difficult to analyze whether elements of the random assignment process contributed to any observed differences in achievement during the program.[54] Fifth, in any given year, up to 11 percent of students did

51. Mosteller (1995).

52. A full description and preliminary analysis of the effects of these issues is found in Hanushek (1999). See also Prais (1996); Goldstein and Blatchford (1998); Krueger (1997).

53. After the first year, students in the regular classes and the regular classes with aides were randomly reassigned. Preliminary analyses from kindergarten had indicated that these two treatments did not result in significantly different performance, but the reassignment made analysis of aide effects difficult. A much more serious problem is that some students in regular classes were moved to the small class treatment (and a smaller number were moved in the opposite direction). Such transfers were "intended to separate incompatible children and 'to achieve sexual and racial balance'"(Mosteller, 1995, p. 124). Such transfers potentially bias simple comparisons of small and regular classrooms, because treatments are no longer independent of student characteristics.

Virtually no attention has been given to how teachers responded to the experiment. Each teacher clearly knew whether they were part of the small or regular class portion of the experiment. And, teachers in general would prefer smaller to larger classes. Whether these factors influenced work effort or behavior is not known.

54. Considerable controversy exists about how early in schooling reliable achievement testing is possible, but few people suggest that it is either reliable or useful before kindergarten. Nonetheless, approximately half of the students who ever participated began the experiment in the first grade or later—that is, when tests were readily available within the experiment itself. Krueger (1997) demonstrates that there does appear to be random assignment based on key student characteristics, such as race or eligibility for free or reduced-price lunch, providing a prima facie case that kindergarten differences are not just the result of simple biases in treatment assignment.

not take the examinations, although there is little indication that this proportion differs by treatment group. Finally, there was some drift from the target class sizes of thirteen to seventeen and twenty-two to twenty-five students, so that there is actually a distribution of realized class size outcomes over time in both treatment groups. Each of these issues has been raised by the initial researchers and by later interpreters of the results, but the experimental data do not provide information that permits one to ascertain fully the effects of such possible problems.[55] One indication of bias comes from the smaller effects of class size for students in the experiment for the full four years, as compared with each sample taken separately, but this is not conclusive.[56]

It is particularly important to note that Project STAR has never been replicated. The contaminating conditions described above clearly suggest that further experimentation would be useful in reducing uncertainties arising from the original study. Indeed, practical problems of implementation suggest that no single experiment is likely to be entirely free of ambiguities.

Nonetheless, the power of random assignment experimentation should not be ignored. One of Mosteller's strongest messages is the ability of random assignment experiments to circumvent some of the difficulties of relying on statistical analyses of observations from natural outcomes of the schooling process.[57] It has been argued that there has been no replication or extension of the approach taken by Project STAR because experiments are expensive. Project STAR involved appropriations of about $3 million per year.[58] Yet the proper frame of reference is the cost of a full-scale program, such as the 1996 California class size initiative for kindergarten through third grade, which involves annual expenditures of over $1 billion. Proposed national programs go far beyond this. The potential costs of implementing an ineffective policy on that scale dwarf the costs of designing, implementing, and evaluating a series of extensive random assignment studies designed to investigate alternative policy proposals.[59] Further, when class size reductions are implemented for an entire state, they defy subsequent analysis of their effectiveness. In other words, it is unlikely that one will ever obtain reliable evaluations of whether the $1 billion spent annually in California is achieving any positive educational results. But if reductions in class size are viewed as policies with political rather than educational objectives, it is perhaps the case that policymakers do not want to know whether performance improves.

55. See, for example, Finn and Achilles (1990); Mosteller (1995); Krueger (1997).

56. See Hanushek (1999). Because the annual samples and the four-year sample have various selection biases, it is difficult to estimate the magnitude of any one type of bias. Nonetheless, if the potential biases have no effects on the treatment estimates, the four-year effects should be as high as or higher than the annual effects.

57. Mosteller (1995). See also Hanushek and others (1994).

58. Word and others (1990).

59. Hanushek and others (1994).

Interpretation and Conclusions

The extensive investigation of the effects of class size on student performance has produced a very consistent picture. There appears to be little systematic gain from general reductions in class size. This story comes through at the aggregate level, where pupil-teacher ratios and class sizes have fallen dramatically over the past three decades and student performance has remained virtually unchanged. It also comes through from international data, where one finds extraordinarily large differences in class sizes without commensurate differences in student performance. But since the aggregate analyses could be misleading for a variety of analytical reasons, more weight should be put on school-level analyses and on experimental data. From production function estimates, there is little reason to believe that overall reductions in class size will yield much in the way of positive achievement gains. Across several hundred separate estimates of the effects of reduced class size, positive and negative effects almost balance each other out, underscoring the ineffectiveness of overall class size policies such as those being currently advocated. Finally, the one major random assignment experiment, Tennessee's Project STAR, provides no support for widespread reductions in class size, although it holds out hope for gains from smaller kindergartens.

None of this says that smaller classes never matter. Indeed, the micro-level evidence, which shows instances where differences in teacher-pupil ratios appear to be important, suggests just the opposite. My own interpretation is that there are likely to be situations—defined in terms of specific teachers, specific groups of students, and specific subject matter—where small classes could be very beneficial for student achievement. At the same time, there are many other situations where reduced class size has no important effect on achievement, even though it always has very significant impacts on school costs. Thus, for example, across-the-board policies of class size reductions, such as those enacted in 1996 in California for elementary education

through grade three, are unlikely to have a beneficial effect on overall student achievement.[60]

The real difficulty is that we do not know how to describe, a priori, situations where reductions in class size will be beneficial. Thus it is not possible to legislate only good outcomes from the state capital, or to institute only good outcomes from the courtroom. Policies developed in these contexts can only expect average gains, which appear to be very small or nonexistent.

The California policy, which many other states are apparently on the verge of emulating, illustrates another aspect of the relationship between research evidence and policymaking. This program was designed to reduce class sizes close to the regular-sized classes in the Tennessee experiment. No evidence from Project STAR relates to the likely effects of such a policy change, as opposed to moving classes down to the level of fifteen students. Moreover, Glass and Smith's original analysis, on which Project STAR was based, casts serious doubts on the potential for any improvement in student performance from such a policy.[61]

Much of the case for reduced class size rests on common sense. With fewer students, teachers can devote more attention to each child and can tailor the material to an individual child's needs. But consider, for example, a change from classes of twenty-six students to classes of twenty-three. This represents an increase of over 10 percent in teacher costs alone (and most likely would raise other costs as well). It is relevant to ask whether teachers would in fact notice such a change and alter their approach. The observational information from teacher and classroom process effects of the one-third class size reductions in Project STAR suggest no noticeable changes.[62]

The policy issue is not defined exclusively by whether one should expect positive effects from reducing class size. Class size reduction is one of the most expensive propositions for change in the schools. Project STAR involved increasing the number of classroom teachers by one-third, a policy with massive expenditure implications if implemented on a widespread basis. The expense of such policies puts natural limits on what is feasible, so that many reductions are ultimately rather marginal. Marginal changes, however, are even less likely to lead to underlying changes in the behavior of teachers.[63]

60. In the short run, it is quite conceivable that the California program could have negative effects. Because it was introduced unexpectedly and with little lead time, many districts found themselves without the necessary classrooms or teachers to permit the desired reductions in class sizes, leading to transitional difficulties with consequences for a number of years. Unfortunately, it will never be possible to analyze the effects of this program because it was available to all school districts. Although some districts did not immediately lower class sizes, these surely were not otherwise comparable to the districts that did.

61. Glass and Smith (1979).

62. Finn and Achilles (1990).

But could reducing class size hurt? Many argue (correctly) that improving the quality of U.S. schools should be a very high priority for policy at the local, state, and national levels. Moreover, this argument is supplemented by concerns about equity and equality of opportunity that follow from significant disparities in outcomes across economic and racial groups. Faced with these real concerns and given the strong advocacy for reducing class sizes from teachers and parents, many politicians appear to find proposals to reduce class size simply irresistible. After all, some students will undoubtedly benefit from smaller classes even if all do not.

The primary argument against pursuing policies with little chance of success is that bad policy drives out good policy. There are clear limits to the amount of funds and attention available to education. Squandering public and political attention on policies that reinforce existing inefficiencies and promise little hope of success is likely to have long-term consequences, although these will not be apparent until some time in the future.

The available evidence indicates that the ultimate effect of any large-scale program to reduce class size will depend much more importantly on the quality of the new teachers hired than on the effects of smaller classes per se. Variations in teacher quality have been shown to be extraordinarily important for student achievement, and econometric studies indicate that these variations completely dominate any effects of altered class size.[64] Thus

63. The magnitude of achievement effects versus expenditure can be put into perspective from the analysis of Texas schools by Rivkin, Hanushek, and Kain (1998). In that study, an increase in spending of $1,000 per student would yield a reduction in class size by from three to six students, depending on how it was accomplished. Such an expenditure, which represents a 14 percent increase from the 1996 national average of $7,000 per student, is estimated to yield an increase of 0.05 to 0.10 standard deviation of performance for low-income students in the fourth grade and half that amount in fifth grade. For students not eligible for free or reduced-price lunches and for low-income students in later grades, there is no reason to expect any average improvement for such an increase. Reductions of the magnitude considered in Project STAR would cost even more than those considered here.

64. Rivkin, Hanushek, and Kain (1998) demonstrate that class size variation can explain just a very small portion of the variation in student achievement and that variations in teacher quality are much more significant. Hanushek (1992) estimates variations in total teacher differences (measured and unmeasured) and shows that the differences in student achievement with a good versus a bad teacher can represent more than one grade level within a single school year. See other references in Hanushek (1997).

if new hires resulting from a class size reduction policy are above the average quality of existing teachers, average student performance is likely to increase; if below the average teacher quality, average student performance is likely to fall with class size reductions. But from past experience there is little reason to believe that the quality of new teachers will be significantly different from that of existing teachers unless the incentives facing schools also change.[65]

To many in the system, one appeal of simple class size policies is that they maintain the existing structure of schools while simply adding more resources. Yet the existing organization and incentives have proved very ineffective at translating resources into student performance. Moreover, much of the expenditure growth in the past came from pursuing the very policies being proposed today. If such policies failed previously, why should the next round be any different?

The uncertainty about the specific circumstances that lead to desirable student performance through smaller classes and the ineffectiveness of current selection, hiring, and retention policies for teachers are exactly what lie behind calls for improved incentives in the schools.[66] The current school structure provides few incentives for improving student performance. In essence, nobody's job or economic reward depends on student performance. In such a situation, nobody really expends much effort to discover in which situations small classes will succeed. Decisions about class size are discussed in terms of fairness rather than student performance or cost control. Would it be fair to allow one group of students or teachers to have small classes while others must have large classes? Such logic, which totally ignores consideration of effectiveness, has almost certainly contributed to the growth in expenses and use of resources underlying the currently ineffective school operations.

Most discussions of class size reductions begin with the assertion that student performance will increase only if class sizes are reduced, a proposition generally shown to be erroneous. These discussions then move quickly to policies of large-scale reductions in pupil-teacher ratios that offer no direct linkage between specific decisions about class size and student performance. The situation in the schools and the resulting educational outcomes might change significantly if everyone had stronger incentives to use budgets wisely and to improve student performance.

65. Under some circumstances, such as the large unexpected hiring in 1996 for the California class size reduction program, one might expect the average quality to fall. In general, however, there is no shortage of trained teachers, and the real issue is selection from the substantial pool not currently employed in the schools. See Ballou and Podgursky (1997); Murnane and others (1991).

66. Hanushek and others (1994).

References

Akerhielm, Karen. "Does Class Size Matter?" 1995. *Economics of Education Review* 14(3): 229–41.

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2): 533–75.

Ballou, Dale, and Michael Podgursky. 1997. *Teacher Pay and Teacher Quality.* Kalamazoo, Mich.: Upjohn Institute for Employment Research.

Barnett, W. Steven. 1992. "Benefits of Compensatory Preschool Education." *Journal of Human Resources* 27(2): 279–312.

Beaton, Albert E., and others. 1996a. *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

———. 1996b. *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

Betts, Julian R. 1995. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth." *Review of Economics and Statistics* 77(2): 231–47.

———. 1996. "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by Gary Burtless, 141–91. Brookings.

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

Cain, Glen G., and Harold W. Watts. 1970. "Problems in Making Policy Inferences from the Coleman Report." *American Sociological Review* 35(2): 328–52.

Card, David, and Alan B. Krueger. 1996. "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. Brookings.

Coleman, James S., and others. 1966. *Equality of Educational Opportunity.* Government Printing Office.

Congressional Budget Office. 1986. *Trends in Educational Achievement.*

Cook, Michael D., and William N. Evans. 1996. "Families or Schools? Explaining the Convergence in White and Black Academic Performance." Unpublished paper.

Ehrenberg, Ronald G., and Dominic J. Brewer. 1995. "Did Teachers' Verbal Ability and Race Matter in the 1960s? Coleman Revisited." *Economics of Education Review* 14(1): 1–21.

Ferguson, Ronald F., and Helen F. Ladd. 1996. "How and Why Money Matters: An Analysis of Alabama Schools." In *Holding Schools Accountable: Performance-Based Reform in Education*, edited by Helen F. Ladd, 265–98. Brookings.

Finn, Jeremy D., and Charles M. Achilles. 1990. "Answers and Questions about Class Size: A Statewide Experiment." *American Educational Research Journal* 27(3): 557–77.

Finn, Jeremy D., and others. 1990. "Three Years in a Small Class." *Teaching and Teacher Education* 6(2): 127–36.

Glass, Gene V., and Mary Lee Smith. 1979. "Meta-Analysis of Research on Class Size and Achievement." *Educational Evaluation and Policy Analysis* 1(1): 2–16.

Goldstein, Harvey, and Peter Blatchford. 1998. "Class Size and Educational Achievement: A Review of Methodology with Particular Reference to Study Design." *British Educational Research Journal* 24(3): 255–68.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. 1996. "The Effect of School Resources on Student Achievement." *Review of Educational Research* 66(3): 361–96.

Grissmer, David W., and others. 1994. *Student Achievement and the Changing American Family.* Santa Monica, Calif.: Rand Corporation.

Grogger, Jeffrey T. 1996. "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics* 14(2): 231–53.

Hanushek, Eric A. 1979. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions." *Journal of Human Resources* 14(3): 351–88.

———. 1992. "The Trade-Off between Child Quantity and Quality." *Journal of Political Economy* 100(1): 84–117.

———. 1996a. "A More Complete Picture of School Resource Policies." *Review of Educational Research* 66(3): 397–409.

———. 1996b. "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. Brookings.

———. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19(2): 141–64.

———. 1999. "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects." *Educational Evaluation and Policy Analysis* 21(2) (forthcoming).

Hanushek, Eric A., and John F. Kain. 1972. "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. Random House.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 1998. "Does Special Education Raise Academic Achievement for Students with Disabilities?" Working Paper 6690. Cambridge, Mass.: National Bureau of Economic Research.

Hanushek, Eric A., and Dongwook Kim. 1996. "Schooling, Labor Force Quality, and the Growth of Nations." Unpublished paper. University of Rochester, Department of Economics (December).

Hanushek, Eric A., and Steven G. Rivkin. 1997. "Understanding the Twentieth-Century Growth in U.S. School Spending." *Journal of Human Resources* 32(1): 35–68.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. 1996. "Aggregation and the Estimated Effects of School Resources." *Review of Economics and Statistics* 78(4): 611–27.

Hanushek, Eric A., and others. 1994. *Making Schools Work: Improving Performance and Controlling Costs.* Brookings.

Hartman, William T. 1980. "Policy Effects of Special Education Funding Formulas." *Journal of Education Finance* 6(Fall): 135–59.

Hauser, Robert M., and Min-Hsiung Huang. 1996. "Trends in Black-White Test-Score Differentials." Discussion Paper 1110-96. University of Wisconsin, Institute for Research on Poverty (October).

Heckman, James S., Anne Layne-Farrar, and Petra Todd. 1996. "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by Gary Burtless, 192–289. Brookings.

Hedges, Larry V., and Amy Nowell. 1998. "Black-White Test Score Convergence since 1965." In *The Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips, 149–81. Brookings.

Hoxby, Caroline Minter. 1998. "The Effects of Class Size and Composition on Student Achievement: New Evidence from Natural Population Variation." Working Paper 6869. Cambridge, Mass.: National Bureau of Economic Research (December).

Krueger, Alan B. 1997. "Experimental Estimates of Education Production Functions." Working Paper 6051. Cambridge, Mass.: National Bureau of Economic Research (June).

Lamdin, Douglas J. 1995. "Testing for the Effect of School Size on Student Achievement within a School District." *Education Economics* 3(1): 33–42.

Lewit, Eugene M., and Linda Schurmann Baker. 1997. "Class Size." *Future of Children* 7(3): 112–21.

Mayer, Susan E. 1997. *What Money Can't Buy: Family Income and Children's Life Chances.* Harvard University Press.

Monk, David H. 1990. *Educational Finance: An Economic Approach.* McGraw-Hill.

Mosteller, Frederick. 1995. "The Tennessee Study of Class Size in the Early School Grades." *Future of Children* 5(2): 113–27.

Murnane, Richard J., and others. 1991. *Who Will Teach?* Harvard University Press.

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

Nye, Barbara A., and others. 1993. "The Lasting Benefits Study: A Continuing Analysis of the Effect of Small Class Size in Kindergarten through Third Grade on Student Achievement Test Scores in Subsequent Grade Levels: Sixth Grade Technical Report." Tennessee State University, Center of Excellence for Research in Basic Skills.

Organization for Economic Cooperation and Development. 1996. *Education at a Glance: OECD Indicators.* Paris.

Prais, S. J. 1996. "Class-Size and Learning: The Tennessee Experiment—What Follows?" *Oxford Review of Education* 22(4): 399–414.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 1998. "Teachers, Schools, and Academic Achievement." Working Paper 6691. Cambridge, Mass.: National Bureau of Economic Research.

Staley, Samuel R., and John P. Blair. 1995. "Institutions, Quality Competition and Public Service Provision: The Case of Public Education." *Constitutional Political Economy* 6: 21–33.

Stevenson, Harold W., and James W. Stigler. 1992. *The Learning Gap: Why Our Schools Are Failing and What We Can Learn from Japanese and Chinese Education.* Summit Books.

U.S. Department of Education. 1991. *To Assure the Free Appropriate Public Education of All Children with Disabilities.* Thirteenth annual report to the Congress on the implementation of the Individuals with Disabilities Education Act.

Wenglinsky, H. 1997. "When Money Matters: How Educational Expenditures Improve Student Performance and When They Don't." Unpublished paper. Princeton, N.J.: Educational Testing Service.

Wirtz, Willard. 1977. *On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline.* New York: College Entrance Examination Board.

Word, Elizabeth, and others. 1990. *Student/Teacher Achievement Ratio (STAR), Tennessee's K–3 Class Size Study: Final Summary Report, 1985–1990.* Nashville: Tennessee State Department of Education.