The Economics of Schooling: Production and Efficiency in Public Schools

Eric A. Hanushek

Journal of Economic Literature, Volume 24, Issue 3 (Sep., 1986), 1141-1177.

Stable URL:
http://links.jstor.org/sici?sici=0022-0515%28198609%2924%3A3%3C1141%3ATEOSPA%3E2.0.CO%3B2-X

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Journal of Economic Literature is published by American Economic Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aea.html.

Journal of Economic Literature
©1986 American Economic Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR

http://www.jstor.org/
Fri Mar 22 14:28:27 2002

Journal of Economic Literature
Vol. XXIV (September 1986), pp. 1141–1177

The Economics of Schooling: Production and Efficiency in Public Schools

By ERIC A. HANUSHEK
University of Rochester

Large parts of this article were written while I was at the Congressional Budget Office. Helpful comments and suggestions were provided by Stanley Engerman, Daniel Koretz, Richard Murnane, John Pencavel, Rudolph Penner, John Quigley, and a referee. Paul Houts provided excellent editorial suggestions. Responsibility for the interpretation of data, of course, rests solely with me.

I. *Introduction*

IN RECENT YEARS, public and professional interest in schools has been heightened by a spate of reports, many of them critical of current school policy.[1] These policy documents have added to persistent and long-standing concerns about the cost, effectiveness, and fairness of the current school structure, and have made schooling once again a serious public issue. As in the past, however, any renewed interest in education is likely to be short-lived, doomed to dissipate as frustration over the inability of policy to improve school practice sets in.

This frustration about school policy relates directly to knowledge about the educational production process and in turn to underlying research on schools. Although the educational process has been extensively researched, clear policy prescriptions flowing from this research have been difficult to derive.[2]

There exists, however, a consistency to the research findings that does have an immediate application to school policy: Schools differ dramatically in "quality,"

[1] During a two-month period in the spring of 1983, no fewer than five notable reports on the nation's schools appeared: National Commission on Excellence in Education 1983; Aerospace Research Center 1983; Business-Higher Education Forum 1983; Education Commission of the States 1983; and Twentieth Century Fund 1983. The Department of Education has responded with two major reports: U.S. Department of Education 1984, 1986. Two more recent reports presenting extensive reform plans are Task Force on Teaching (1986) and The Holmes Group (1986).

[2] Education, being a more recent subject of economists' attention, has been analyzed more extensively by researchers in other disciplines: psychology, sociology, and political science. Much of this work focuses on subjects outside those of interest to economists. However, there are very important points of overlap in measuring scholastic performance, in analyzing the educational production process, and in formulating educational policy. Indeed, although not usually found in economics journals, this related research is an important ingredient in the material discussed here.

but not because of the rudimentary factors that many researchers (and policy makers) have looked to for explanation of these differences. For example, differences in quality do not seem to reflect variations in expenditures, class sizes, or other commonly measured attributes of schools and teachers. Instead, they appear to result from differences in teacher "skills" that defy detailed description, but that possibly can be observed directly. This interpretation of research findings has clear implications for school policy.

This essay reviews existing analyses of the educational process from several different perspectives, one of which is the relevance of the research for school policy. The economics research on schooling is empirical in nature, and an understanding of its findings must begin with an underlying conceptual model of the educational process. A natural starting point is economic models of production theory and firm behavior. Unfortunately, standard textbook formulations or typical industry and aggregate production function specifications provide little direct guidance in educational analysis, because they seldom are designed to deal with the detailed policy questions that have been central to investigations of schooling. Indeed, after modifying the standard framework to accommodate the policy purposes, the measurement issues, the incentive structure of schools, and so forth, the resultant models may be sufficiently different that a new nomenclature is useful. The most important modification involves interpretations of economic efficiency—a concept that has a very clear meaning in textbook analyses of the theory of the firm but that becomes quite cloudy in the world of public schools.

The empirical formulations developed in the research on schooling do provide insights that appear applicable to other micro policy areas where complicated production relationships for services are central. The results of this review also have immediate implications for other areas of economic study. A variety of public finance investigations, urban housing and location studies, and labor economics analyses include at least tangentially some consideration of school quality and performance—but generally these studies do not incorporate the results of direct analyses of schooling.

A. *Limits of the Study*

This study examines the research on the economics of education and schooling and explores what has been learned and where major gaps remain, focusing on production and efficiency aspects of schools as opposed to the ultimate uses of education. Because there are excellent reviews of "human capital" (Jacob Mincer 1970; Sherwin Rosen 1977), this area is specifically downplayed, even though human capital investment and the economics of education are at times treated as being synonymous.[3] This review also concentrates on public education, for lack of comparable research on the private sector,[4] and on the United States to avoid the problems of drawing inferences from cross-country data where basic educational patterns differ substantially.[5]

[3] An additional reason for emphasizing production and efficiency aspects is that, although work on human capital ostensibly deals with investment behavior in schooling, the real focus frequently tends to be on income determination, or schooling as an input to the wage determination process.

[4] Recent work on private schooling, while generating considerable interest, has not looked explicitly at production relationships in private schools. Instead it has stopped at contrasting mean performance in public and private schools. See section IV, below.

[5] There have been a number of studies of schooling in developing countries, much of it emanating from the World Bank. See, for example, Stephen Heyneman and William Loxley (1983), Richard Kollodge and Robin Horn (1985), and Bruce Fuller (1985). These studies frequently involve a much wider range of inputs—such as teachers' education levels ranging from the third grade through college—and therefore are better able to identify and to estimate the effects

One other prefatory remark may be useful. At least from an economics perspective, distinctions between elementary and secondary schooling and postsecondary schooling seem small. While the private postsecondary schooling sector is somewhat larger, both segments of the educational system are dominated by public supply; the technologies appear very similar, at least on the surface; and, most frequently a year of schooling at any level is treated as being equivalent in the sense that years enter linearly into some other activity or behavior that is being modeled. However, the research and indeed the focus of policy attention in the two sectors have differed markedly. Economic studies of elementary and secondary schooling have concentrated on production processes, public finance questions about governmental support, and, to a lesser extent, labor markets for teachers, cost-benefit analyses of specific programs, and public-private choices. Economic studies of higher education have been largely concerned with distributional questions related to access and costs faced by different groups, with governmental subsidy policies, and with attendance decisions; virtually no attention has been given to production processes or the analysis of specific programs.[6]

of fundamental school inputs. On the other hand, the relevant range for policies in developed countries may not even be included in the sample data. The classic study for England is Central Advisory Council for Education 1967.

[6] Exceptions include David Breneman 1976; Lewis Perl 1976; and Timothy Hogan 1981. These studies have tended to concentrate on quantitative variations (for example, numbers of PhDs produced) instead of qualitative variations. Hogan's study is unique in measuring qualitative differences (through subsequent publication records) among PhDs produced.

The division by level of schooling might well be explained by the traditions of other disciplines; those divisions reflect in part differences in cognitive processes with age and in part organizational variations and perspectives. Much of the economic analysis of education has been rather recent and has built upon that of other disciplines.

TABLE 1
EXPENDITURES AND SOURCE OF FUNDING: 1960–83
(ALL ELEMENTARY AND SECONDARY SCHOOLS)

	1960	1970	1980	1983
Expenditures (Billion $)	18.0	45.7	108.6	132.9
Percent GNP	3.6	4.6	4.1	4.0
	Source of Funds (percent)			
Federal	3.9	7.4	8.7	6.8
State	31.1	34.6	41.5	43.3
Local	52.8	47.5	38.2	38.1
Private	12.3	10.5	11.5	11.8

Source: U.S. Statistical Abstract, 1985.

B. *The Elementary and Secondary School Sector*

Before discussing the direct analyses of schooling, it is useful to understand the overall dimensions of the sector. The size of the sector and the changes that have taken place frequently are not well understood.[7] Yet the kinds of policies behind these changes relate directly to the character of production in the public schools and the substance of economists' analyses of schools.

Expenditures. The total spending on elementary and secondary schooling in the United States, as shown in Table 1, is currently equal to about 4 percent of gross national product. The largest fluctuations in its relative size reflect simply total enrollments in schools, which peaked in 1970. A steady rise in per pupil expenditures, however, has pushed upward the resources going into elementary and secondary schools.

There have been two major changes over the past 25 years in the source of funding for schools. First, as displayed in Table 1, federal funding jumped during

[7] A more detailed analysis of schooling at all levels that also includes data since 1940 can be found in Dave O'Neill and Peter Sepielli (1985).

TABLE 2
ELEMENTARY AND SECONDARY SCHOOL PUPILS, STAFFING, AND TYPE OF CONTROL: 1960–80

	1960	1965	1970	1975	1980
Enrollment (thousands)					
Total	42,181	48,473	51,272	49,791	45,949
Elementary	29,150	31,570	31,553	29,340	27,779
Secondary	13,031	16,904	19,719	20,451	18,170
Classroom Teachers (thousands)					
Total	1,600	1,933	2,288	2,451	2,439
Elementary	991	1,112	1,281	1,352	1,365
Secondary	609	822	1,007	1,099	1,074
Private School Enrollment (Percent of Total Enrollment)					
Total	14.0	13.0	10.5	10.0	10.8
Elementary	16.5	15.5	12.8	12.6	13.0
Secondary	8.4	8.3	6.6	6.4	7.4
Catholic School Enrollment (Percent of Private Enrollment)					
Total	89.0	88.5	81.4	68.3	62.6
Elementary	91.1	91.7	82.9	68.2	62.6
Secondary	80.0	77.3	76.9	68.5	62.5

Source: U.S. Statistical Abstract, 1985.

the 1960s. This was followed by a slow growth in federal share during the 1970s and a decline during the 1980s. Second, the financing of local schools was altered extensively during the 1970s by a series of legal and legislative challenges to the use of local property taxes as the principal funding source. This resulted in the steady increase in the level of support from state revenue sources with a commensurate decline in the support of schools from local revenues. Direct private support for schools almost entirely represents expenditures on private schooling; there is a minuscule amount of governmental support for private schooling, and there is a minuscule amount of nongovernmental support for the public schools.

Enrollments. Currently, about 45 million students are enrolled in schools. The peak in elementary school enrollments (grades 1 through 8) occurred in the late 1960s, while high schools peaked in the mid-1970s (Table 2). While student enrollment fell by over 10 percent between 1970 and 1980, the number of classroom teachers actually increased by 7 percent over the same period.[8]

Enrollment in private schools declined in the 1960s and, since then, has remained roughly constant as a proportion of total enrollment. The private school decline largely reflects the decline in Catholic

[8] It is difficult to get total employment figures for elementary and secondary schools, because much of the governmental employment is not separated in the data by level of schooling. Classroom teachers make up 88 percent of the total instructional staff, which includes principals, librarians, and so forth.

Part of the increase in teachers may reflect the requirements of laws related to handicapped students. Federal legislation in 1975 (P.L. 94-142, The Education of All Handicapped Children Act) has been particularly important because of its specific requirements dealing with administrative and school processes.

TABLE 3
SCHOOL COMPLETION: 1950–80

	1950	1960	1965	1970	1975	1980
School Retention						
Percent graduating from high school	50.5	62.1	73.2[a]	75.0	74.3	74.4
Percent entering college	20.5	32.8	38.4[a]	46.1	45.2	46.3
School Completion						
(Population age 25–29)						
Median years completed	12.0	12.3	12.4	12.6	12.8	12.9
Percent 4 years high school or more	52.8	60.7	70.3	73.8	83.1	84.5

Source: U.S. Statistical Abstract, 1970, 1982, 1985.
[a] Data for 1966.

school enrollment. While Catholic schools made up almost 90 percent of private enrollment in 1960, this was down to 63 percent in 1980. Schools affiliated with other religions made up an additional 21 percent of the private school enrollment in 1980, leaving 16 percent of the private school instruction in private schools with no religious affiliation. Private schools remain more important at the elementary school level than at the secondary level.

Performance. In terms of graduation rates and continuation into college, there has been remarkably little change since the mid-1960s. As seen in Table 3, recent data show that an almost constant three-quarters of each age cohort graduates from high school at the normal time, and, with some fluctuations, about 45 percent of each age cohort will enter college immediately. The school completion data for the population age 25 to 29 give a similar view, only the timing is different because of the ages considered. The median years of school completed for the population age 25 to 29 has crept up from 12.3 years in 1960 to 12.9 in 1980. Further, reflecting the increased school attendance of the 1950s and 1960s, the calculated percentage of this age group completing 4 or more years of high school shows a steady rise, reaching 84.5 percent by 1980. Many people note the steady increases in educational attainment of the workforce without realizing that the graduation and college attendance behavior have been steady since before 1970.

There does remain some disparity between high school graduation rates and the completion percentages; the percentage of 25–29 year olds reporting 4 or more years of high school is 10 percent higher than the estimated graduation rate. This may reflect an increasing tendency to complete high school at later ages. Or, the recall data on school completion may simply be inaccurate. Unfortunately, it is not possible to differentiate among these alternative explanations.

Most of the attention given to schools relates to performance on standardized tests and, more specifically, on the Scholastic Aptitude Test (SAT). (For an excellent current review of evidence on test score declines, see Congressional Budget Office 1986a.) Figure 1 displays the history of average test scores on the verbal and math portions of the SAT. As is well known, beginning in 1963, test scores began a steady decline. Verbal scores fell about one half of a stardard deviation before bottoming out in 1979; math scores followed the same time pattern, although the magnitude of decline was not as large.

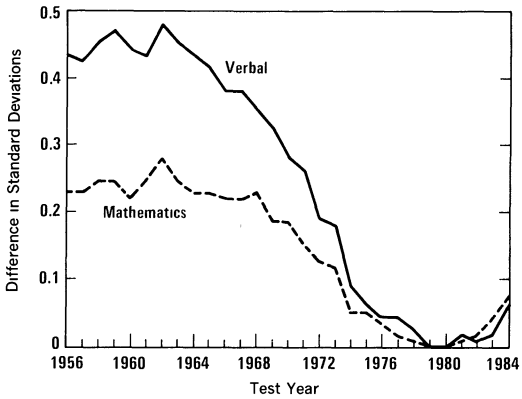

Figure 1. Average SAT Scores, by Subject, Differences from Lowest Year

Source: Congressional Budget Office, 1986a.
Note: Average test scores in each year are measured in standard deviations of student performance from mean scores in 1979, the year of lowest average SAT performance. Comparisons in terms of percentile position of the means across years can be calculated approximately from the normal distribution.

Because absolute scores have little meaning, the comparisons are made in terms of standard deviations of student performance, which can be translated into percentile comparisons using the normal distribution. Thus, a fall in mean performance of .48 standard deviations (verbal scores) implies that mean performance at the trough was approximately equivalent to performance at the 32nd percentile in 1963. Similarly, a drop of .28 standard deviations (math scores) implies that mean performance in 1979 was roughly equivalent to performance at the 39th percentile in 1963.

Performance on other tests, however, is much less known. A wide range of different tests, ones designed with different purposes and validated in a variety of different ways, show declines beginning about the same time. There was a pervasive decline at all grade levels, not restricted just to graduating students. Moreover, as described in Table 4, test scores at lower grades appear to have begun a recovery before the SAT scores. The time patterns of performance on the Iowa tests for different grade levels is shown in Figure 2. Making such intertemporal comparisons is frequently difficult, but the consistency of findings suggests improvements that began in the mid-1970s. The crude evidence points to declines closely related to specific years of schooling or birth cohorts.[9]

Over the past 15 years there has been a consistent narrowing of the gap in test scores between blacks and nonminority students (Congressional budget Office 1986a). This trend appears on virtually all tests, including the SATs. Nevertheless, gaps between minority and nonminority students remain sizable.

Public School Inputs. Dramatic changes in the operations of schools have come along with these changes in student performance. Most notable has been the increase in expenditures per pupil shown in Table 5. The 1983 spending for current services of $2,960 per public school student in attendance was 135 percent *in real terms* above that in 1960. This corre-

TABLE 4
ONSET AND END OF THE ACHIEVEMENT DECLINE, SELECTED TESTS

	Onset	End
Scholastic Aptitude Test (SAT)	1963	1979
American College Testing Program (ACT)	1966	1975
Iowa Tests of Basic Skills—grade 5	1966	1974
Iowa Tests of Basic Skills—grade 8	1966	1976
Iowa Tests of Educational Development—grade 12	1968	1979
Minnesota Scholastic Aptitude Test	1967	na

Source: Congressional Budget Office, 1986a.

[9] See Congressional Budget Office (1986b) for a discussion of alternative hypotheses about this time pattern.

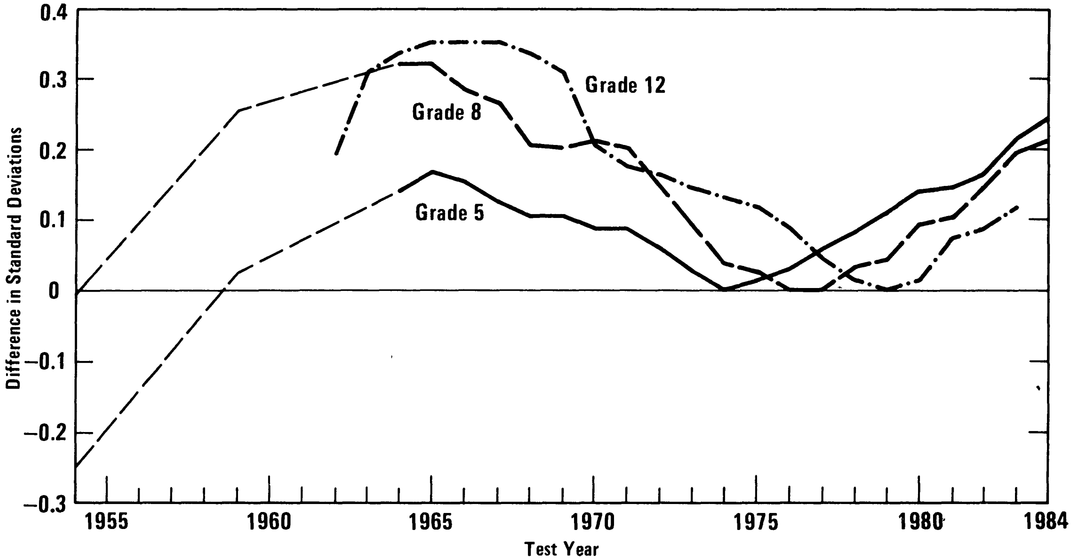

Figure 2. Iowa Average Test Scores, Grades 5, 8, and 12, Differences from Post-1964 Low Point

Source: Congressional Budget Office, 1986a.
Note: See explanation, Figure 1.

sponds to a compound annual growth rate of real expenditures of 3.8 percent. Total expenditures, which include capital expenditures and interest on debt, showed somewhat lower growth, because capital spending was a decreasing portion of the total. By 1983, total spending per student had reached an average of $3,261.

A very significant component of this growth in per pupil expenditures is the overall fall in pupil-teacher ratios (Table 6). These declines, which were previously seen in the increases in classroom teachers during a period of falling school enrollments, have an enormous effect on expenditures per pupil. In the public schools, pupil-teacher ratios fell over 25 percent between 1960 and 1980, with the decline being somewhat higher in elementary schools. By way of comparison, private schools have also had large falls in pupil-teacher ratios, although some of this may

TABLE 5
PUBLIC SCHOOL EXPENDITURES PER STUDENT IN AVERAGE DAILY ATTENDANCE (ADA): 1960–83

	1960	1966	1970	1975	1980	1983
Current expenditure/ADA	$375	$537	$816	$1,286	$2,230	$2,960
1983 Dollars	$1,262	$1,696	$2,094	$2,381	$2,696	$2,960
Total Expenditures/ADA	$472	$654	$955	$1,503	$2,502	$3,261
1983 Dollars	$1,598	$2,066	$2,451	$2,783	$3,025	$3,261

Source: U.S. Statistical Abstract, 1970, 1981, 1985.

TABLE 6
PUPIL-TEACHER RATIOS: 1960–80

	1960	1965	1970	1975	1980
Public Schools					
Total	25.8	24.6	22.3	20.4	19.0
Elementary	28.4	27.6	24.4	21.7	20.5
Secondary	21.7	20.8	19.9	18.8	17.1
Private Schools					
Total	30.7	28.3	23.0	19.6	17.9
Elementary	36.1	33.3	26.5	21.5	19.3
Secondary	18.6	18.4	16.4	15.7	15.0

Source: U.S. Statistical Abstract, 1982, 1985.

reflect the changing composition of private school enrollment noted above. While it is sometimes asserted that the falls in pupil-teacher ratios simply reflect an attempt to maintain overall teacher employment in the face of declining enrollment, this seems inconsistent with the fact that class size declines begin before total enrollment declines.

The characteristics of teachers have also changed dramatically over time. Perhaps the most dramatic change, shown in Table 7, is the aging of the current teacher force. While one-third of public school teachers in the mid-1960s were in their first four years of teaching, that fell to one-twelfth by 1983. The median experience of teachers reached 13 years of experience, from a low of 8 years.

Reflecting in part this stability in the teacher force and in part state regulations and financial incentives of teacher salary schedules, the percent of all teachers with a master's degree or more doubled between 1966 and 1983. By 1983, over half of all public school teachers held at least a master's degree.

Table 7 also shows that the picture of teachers' salaries is, however, different. Average salaries rose dramatically during the 1960s, but subsequently fell in real terms throughout the 1970s. This fall in average salaries is more dramatic when combined with the increase in experience and amount of graduate training of teachers, because both of those factors will increase salaries. Data on entry level salaries show a much steeper decline (in real terms). On the other hand, workers in the entire economy lost ground during the inflationary period of the last decade, so that teacher salaries hold up reasonably well against the salaries of private nonagricultural workers. (In Table 7, average teacher salaries are compared to annualized values of average weekly earnings.)

The Puzzle. The data on the schooling sector suggest a number of puzzles. The most important one—and the subject of most of this review—is that the constantly rising costs and "quality" of the inputs of schools appear to be unmatched by improvement in the performance of students. It appears from the aggregate data that there is at best an ambiguous relationship and at worst a negative relationship between student performance and the inputs supplied by schools. Such conclusions cannot, however, be made on the basis of just the aggregate data.

C. *Overview*

Studies of educational production functions (also referred to as "input-output" analyses or "cost-quality" studies) examine the relationship among the different inputs into and outcomes of the educational process. These studies are systematic, quantitative investigations relying on econometric, as opposed to experimental, methods to separate the various factors influencing students' performance.

The standard textbook treatment of production functions considers only the most stylized examples—say, for example, how much capital and labor to employ in producing some specific output. Knowledge of the production function and the prices for each of the inputs allows a straightforward solution of the "least cost"

TABLE 7
CHARACTERISTICS OF PUBLIC SCHOOL TEACHERS: 1960–83

	1960	1966	1971	1976	1981	1983
Teacher Experience						
1–4 years (percent)	na	32.2	32.3	27.1	14.1	8.3
Greater than 20 years (percent)	na	21.6	18.5	14.3	21.8	25.0
Median (years)	na	9	8	8	11	13
Education						
Master's degree or more (percent)	na	26.1	27.2	36.7	48.9	53.0
Salaries						
Average salary[a]	$5,174	$6,935	$9,470[b]	$12,448	$18,321	$21,790
1983 dollars	$17,406	$21,290	$23,296	$21,786	$20,070	$21,790
Ratio average worker[c]	1.23	1.35	1.43	1.36	1.38	1.49

Source: U.S. Statistical Abstract, 1982, 1985.
[a] Average salary of all instructional personnel (i.e., teachers, principals, guidance counselors, librarians and others associated with instruction).
[b] Estimated.
[c] Ratio of average salary to annualized average weekly earnings in the U.S.

set of inputs—that is, the combination of inputs that would produce any given output at minimum cost. The concept of a production function is a powerful pedagogical tool, and, in its basic form, appears applicable to a wide range of industries—from petrochemicals to education.

In an intermediate microeconomics classroom, production functions are generally assumed to be known precisely by decision makers, to involve only a few inputs that are measured perfectly, and to be characterized by a deterministic relationship between inputs and outputs (that is, a given set of inputs always produces exactly the same amount of output). Furthermore, it is assumed that all inputs can be varied freely.

The realities of education (and virtually all other areas for that matter) differ considerably from such pedagogical assumptions. Indeed, the production function is unknown (to both decision makers and researchers) and must be estimated using imperfect data; some important inputs cannot be changed by the decision maker; and any estimates of the production function will be subject to considerable uncertainty.

Perhaps the largest difference between applying production functions to education and to other industries, however, has been in its immediate application to policy considerations. Statistical estimates of educational production functions have entered into a variety of judicial and legislative proceedings and have formed the basis for a number of intense policy debates.

The history of educational production function analysis is typically traced to *Equality of Educational Opportunity*, or, more commonly, the "Coleman Report" (James Coleman et al. 1966). The Coleman Report was mandated by the Civil Rights Act of 1964 and was conceived as a study of the distribution of educational resources within the United States by race or ethnic background. The study, however, went far beyond simply producing an inventory of school resources. It created a massive statistical base containing survey information for more than one half million students found in some 3,000 separate schools that was employed to ascertain which of the various inputs into the educational process were most important in determining the achievement of students.

Although not the first such study, it is both the best known and the most controversial. In simplest terms, the Coleman Report appeared to demonstrate that differences in schools had little to do with differences in students' performance. Instead, family background and the characteristics of other students in the school seemed much more important. The report's findings generated extensive critiques, policy discussions, and further research (see, for example, Eric Hanushek and John Kain 1972; Samuel Bowles and Henry Levin 1968; and Glen Cain and Harold Watts 1970). Today, even though it remains the most cited analysis of schools, the Coleman Report is commonly held to be seriously flawed, and its importance is more in terms of intellectual history than insights into schools and the educational process.

The production function approach, which began in earnest with the Coleman Report, has not been universally accepted, particularly among educational decision makers. In part, criticism of the approach appears to be a reaction against the specific results. (For example, as described below, these studies tend to suggest that schools are very inefficient in their use of resources.) In part it appears to reflect a general reaction against doing any quantitative evaluation of education and schools. But, it also reflects concern about legitimate analytical problems or misinterpretation of the results of specific studies.

II. *Conceptual and Specification Issues*

The underlying model guiding most analysis is very straightforward. The output of the educational process—that is, the achievement of individual students—is directly related to a series of inputs. Some of these inputs—the characteristics of schools, teachers, curricula, and so forth—are directly controlled by policy makers. Other inputs—those of families and friends plus the innate endowments or learning capacities of the students—are generally not controlled. Further, while achievement may be measured at discrete points in time, the educational process is cumulative; inputs applied sometime in the past affect students' current levels of achievement.[10]

A. *Specification and Measurement of Output*

Clearly, to analyze school production it is essential to employ adequate measures of outcomes. But measuring outputs is not simple. While economic theory concentrates on varying quantities of a homogeneous output, this is not easily translated into an educational equivalent. Education is a service that transforms fixed quantities of inputs (that is, individuals) into individuals with different qualities. Educational studies concentrate—as they should—on "quality" differences.

A majority of studies into educational production relationships measure output by standardized achievement test scores, although significant numbers have employed other quantitative measures such as student attitudes, school attendance rates, and college continuation or dropout rates. The measures used, however, are generally proxies (with varying degrees of validation) for more fundamental outcomes. Some people, including many school practitioners, simply reject this line of research entirely because they believe that educational outcomes are not or cannot be adequately quantified.

[10] For further discussion of this model, see Hanushek (1972, 1979).

Interest in and concern about the performance of schools relates directly to the perceived importance of schooling in affecting the ability of students to perform in and cope with society after they leave school. While seldom fully articulated, the theory is that more schooling makes people more productive in the labor market, better able to participate in democracy, better consumers, and so forth—in other words, healthy, wealthy, and wise. Economists, sociologists, and political scientists have conducted a broad range of investigations into postschooling outcomes. In general, empirical studies confirm the correlation between higher levels of schooling and positive attributes after schooling. Indeed, it is commonplace for individual level investigations of behavior to include schooling more or less automatically as a conditioning variable regardless of the topic under investigation.

The analytic problem is that postschooling outcomes cannot be contemporaneously observed with the schooling. Of course, this kind of problem arises elsewhere—for example, in the analysis of environmental effects on health or of changes in social security law on retirement behavior, and a variety of approaches are employed for gleaning information from existing data. By far the most common approach in education is to analyze cross-sectional variations in measures that can serve as proxies for future performance. A natural starting point, thus, is an investigation of how schooling affects labor market performance and other postschooling activities.

From the standpoint of production function analyses, there are two fundamental difficulties with existing research into postschooling outcomes. First, the concentration on *quantity differences*, or pure time spent in schooling activities, as opposed to *quality differences* makes it difficult to relate the analyses directly. Treating all time spent in schooling activities equally neglects the possibility that time in some school settings might very well have different value from that spent in other settings; yet the differential effectiveness of schools is the heart of production function studies. This concentration on quantity of schooling, which is perfectly explicable in terms of the availability of data, holds equally for the labor market studies generally pursued by economists and for the nonlabor market studies pursued more frequently by researchers in other disciplines. Second, the conceptual underpinnings of the presumed improved performance of the more educated remain unclear. This complicates attempts to measure directly any quality differences among students, because there is little guidance on just what to look for.

The most extensive analyses by economists have related wages of workers to number of years of schooling completed (see, for example, the reviews by Mincer 1970 and Rosen 1977). To be sure, the theoretical modeling behind this work does not restrict attention merely to quantity and, in fact, in many instances can be interpreted as incorporating both quantity and quality differences in schooling. Nevertheless, when it comes to empirical implementation, data shortcomings frequently demand that exclusive attention be given to quantity. This is even the case in models of "human capital production functions" (see Yoram Ben-Porath 1970).

Attempts to incorporate qualitative measures of schooling into labor market studies have been severely limited by availability of data, by the necessity of using fairly peculiar samples, and by reliance on stringent assumptions about school operations. One approach has been to include individual test score information, but this sort of data exists only in rare instances and is usually not representative; see, for example, Zvi Griliches and William Mason 1972; John Hause 1972;

Hanushek 1973, 1978; Orley Ashenfelter and Joseph Mooney 1968; Daniel Rogers 1969; Randall Weiss 1970; Paul Taubman and Terence Wales 1974; W. Lee Hansen, Burton Weisbrod, and William Scanlon 1970.

These studies produce a wide range of estimates of the test score/earnings relationships; they range from finding no relationship to finding one that dominates any measure of quantity of schooling. In most studies, however, years of schooling and measures of cognitive ability exhibit independent effects on earnings.

Another general line of inquiry has been to incorporate measures of the characteristics of individuals' schools directly into earnings functions. One class of such studies includes average school expenditure data (for example, Paul Wachtel 1976; George Johnson and Frank Stafford 1973; Thomas Ribich and James Murphy 1975; Charles Link and Edward Ratledge 1975; John Akin and Irwin Garfinkel 1977). A second class includes measures of specific school resources or characteristics of teachers in the earnings model (for example, Finis Welch 1966, 1973; Christopher Jencks and Marsha Brown 1975; Jere Behrman and Nancy Birdsall 1983). But such analyses must *assume* that differences in expenditures or in the specific resources provide an index of differences in quality. This is an important question to be addressed through the analysis of educational production functions. Moreover, unless the models also include measures of other inputs into the educational process—such as the family backgrounds or characteristics of other students in the schools, they will obtain biased estimates of the effects of differences in schools.[11] Perhaps for these reasons, the results of these studies on the effects of quality differences have been inconclusive.

Although the relationship of schooling and labor market performance is central to many policy questions, it is not the only area of interest; see, for example, reviews by Robert Michael (1982) and Robert Haveman and Barbara Wolfe (1984). Studies have examined the role of education in increasing job satisfaction, in maintaining personal health (Michael Grossman 1975), and in increasing the productivity of mothers engaged in household production (Arleen Leibowitz 1974), as well as the effects of the mother's education on the learning of young children. Other studies have considered the effect of education on political socialization and voting behavior (Richard Niemi and Barbara Sobieszek 1977), the relationship between education and criminality (Isaac Ehrlich 1975), the contribution of education to economic growth (Edward Denison 1974), and the effect of education on marriage and divorce (Gary Becker, Elizabeth Landes, and Robert Michael 1977). While these studies have suggested some gross effects of the quantity of schooling on other life outcomes, they virtually have never addressed the question under consideration here: How do such outcomes vary in response to differences in school programs and operations?

In summary, the literature about the relationship between measures of schooling quality and subsequent attainment is ambiguous. The analyses available are often crude empirical forays that are difficult to replicate and to evaluate in a definitive manner. While these studies offer an important perspective on how to observe educational outcomes, they do not currently provide firm guidance about appropriate contemporaneous measures of quality that might be used in production function analysis. (This is not, of course, the primary purpose of such studies.)

As a general research strategy, one might think of approaching the issue in

[11] Education occurs both at home and in the schools, and characteristics of families (such as income levels) and characteristics of schools tend to be positively correlated. These correlations, discussed below, imply biased estimates in the analysis of earnings discussed here.

a different way—by considering what attributes of schooling (or individuals) were important for subsequent success and then developing direct measures that could be obtained during the same time period with the schooling. Yet, a fundamental shortcoming of this strategy is the superficiality of the conceptual notions of the mechanisms by which education affects productivity and later experiences. As measured through various standardized tests, cognitive skills are probably the chief contemporaneous measure of educational quality currently available. But this may not be the only, let alone the most important, outcome of schooling in determining the future success of students. One might think that more educated individuals can accomplish given tasks better or more swiftly, but surely this holds for only certain types of jobs. Less education may even be an advantage in jobs requiring manual skills or jobs that are very repetitive.

One rather commonly held presumption is that better educated individuals are able to perform more complicated tasks or are able to adapt to changing conditions and tasks (see Welch 1970; Richard Nelson and Edmund Phelps 1966). This hypothesis has important implications for studying the productivity and outputs of schools, because it provides some rationale for favoring measures of analytical ability. Outside of this area, however, similar conceptual views of the important elements of schooling are even harder to find.

The uncertainty about the source of schooling-earnings relationships is also highlighted by recent attention to "screening" aspects of schooling. Schools may not improve the skills of students but may simply identify the more able. The latter view has been the subject of both theoretical and empirical treatment by economists and sociologists (for example, Ivar Berg 1970; A. Michael Spence 1973, 1974; Kenneth Wolpin 1977; John Riley 1979a, 1979b; Richard Layard and George Psacharopoulos 1974; Andrew Weiss 1983). Screening implies that the social value of schooling may be considerably less than the private value if schools are merely identifying the more able instead of actually changing their skills.

The screening model also has direct implications for measuring educational outcomes and analyzing educational production relationships. In a screening model, the output of schools is information about the *relative* abilities of students. This would suggest that more attention should be directed toward the distribution of observed educational outcomes (instead of simply the means) and their relationship to the distribution of underlying abilities. Further, it might radically alter the interpretation of some studies, such as those of school dropout rates, because schools with higher dropout rates might actually be providing better information (higher output) than those with lower rates. (This is clearly an interpretation that is very different from that of the authors of these studies.)

Production and screening, however, are not the only models explaining subsequent performance. For example, Jencks et al. (1972) argue that luck and personal characteristics (unaffected by schooling) are the most important determinants of earnings differences. Bowles and Herbert Gintis (1976) believe that differences in earnings arise chiefly from the existing social structure and that schools adjust to rather than determine subsequent outcomes. While these last two theories are not completely convincing, there is not enough available evidence to determine conclusively which, if any, of these four divergent views are valid.

B. *Standardized Test Scores*

At this point, it may be useful to consider standardized test scores more specifically because they are the most commonly used measure in investigating the educational process. As previously mentioned, considerable uncertainty exists about the appropriateness of using test scores as outcome measures. Studies of lifetime outcomes, while conceptually very relevant to measuring school outputs, have not been particularly illuminating; existing empirical evidence is inconclusive about the strength of the link between test scores and subsequent achievement outside of schools. Because cognitive tests given in schools are generally not designed to measure subsequent performance, it is not particularly surprising that they are imperfect.[12]

Nevertheless, performance on tests is being used to evaluate educational programs, and even to allocate funds, and there are some pragmatic arguments for the use of test scores as output measures. Besides their common availability, one argument is that test scores appear to be valued in and of themselves. To a large extent, educators tend to believe that they are important, albeit incomplete, measures of education. Further, parents and decision makers appear to value higher test scores—at least in the absence of evidence that they are unimportant. In fact, the use of scores on standardized tests as criteria for high school graduation (usually referred to as minimum competency tests) has increased dramatically in recent years and now is mandated by many states.

A more persuasive argument for the use of test scores relates to continuation in schooling. Almost all studies of earnings that include both quantity of schooling and achievement differences find significant effects of quantity that are independent of achievement differences. This implies that measured differences in achievement do not adequately measure all skill differences. At the same time, however, test scores have an important use in selecting individuals for further schooling and thus may relate directly to the "real" outputs through the selection mechanism (cf. Dennis Dugan 1976). The use of tests for predicting future school performance and for selection is also a central issue in the study by Willard Wirtz et al. (1977), which reviews the decline in Scholastic Aptitude Test scores.

Finally, the variety of potential outcomes of schooling suggests that the educational process may have multiple outputs, some of which are very poorly measured by test scores. Moreover, how effective test scores are in measuring the contribution of schooling to subsequent performance probably varies at different points in the schooling process. Specifically, test scores might be more appropriate in the earlier grades, where the emphasis tends to be more on basic cognitive skills—reading and arithmetic—than in the later grades. (Note that virtually all production function studies have been conducted for elementary and secondary

[12] Standardized tests employed by schools lack external validation in terms of labor market skills or other subsequent outcomes. This is not particularly surprising, however, given the primary motivations behind their construction. Most tests are designed to: examine students on specific knowledge; rank students in terms of skills or knowledge; or predict performance. The performance of interest, however, is often future success in schooling. This, for example, is motivation behind the SAT tests. See Congressional Budget Office (1986a, Chapter 2) for an excellent discussion of standardized tests and their use.

The efforts to validate tests are often quite extensive (see, for example, Hunter Breland (1979) on the SAT test), but they are frequently concerned with such things are consistency across tests or correlations with other measures. Alternative standards for validation are described in American Educational Research Association (1985).

Reliability of tests is a second major concern. Does a given test produce the same score if taken at different times by the same individual, and do slightly different wordings of questions covering the same concept yield the same results? None of these relates directly to whether or not tests cover material, knowledge, or skills valued by society.

Finally, although they have come under considerable criticism, a variety of employment tests have been designed and validated with labor market performance. The attacks on these have concentrated on both their inaccuracies and their potential for discriminatory results. Data on performance on these tests have never been available to school researchers.

schools. In postsecondary education, few people believe that test scores adequately measure outputs.)[13]

The objective in measuring outputs of education is to find a quantitative measure that is both readily available and related to long-run goals of schooling. At the individual level, test scores related to ability or achievement have obvious appeal, even though available research provides little guidance about specific kinds of tests or different possible dimensions.[14] When analysis is conducted at the aggregate school level, other possibilities such as school continuation rates or college attendance rates also are available and provide a direct link to differences in quantities of schooling.

C. *Empirical Formulation*

Somewhat ironically, even though educational studies have attempted to provide much more detail about input differences, they have still been faced with extensive criticism about the specification of the inputs. Part of this criticism arises because the choice of inputs is guided, sometimes quite explicitly, by the availability of data more than by any notions of how the study is best conceived. But most of the criticism undoubtedly stems from the desire to apply findings to actual policy decisions—something not found in more academic investigations of production relations.

The general conceptual model depicts the achievement of a given student at a particular point in time as a function of the cumulative inputs of family, peers or other students, and schools and teachers. These inputs also interact with each other and with the innate abilities, or "learning potential," of the student. Two points deserve emphasis: The inputs should be relevant to the students being analyzed; and the educational process should be viewed as cumulative—past inputs have some lasting effect, although their value in explaining output may diminish over time. Failure to recognize these points has probably caused the greatest problems in interpreting individual studies.

Empirical specifications have varied widely in details, but they have also had much in common. Family inputs tend to be measured by sociodemographic characteristics of the families, such as parental education, income, and family size. Peer inputs, when included, are typically aggregate summaries of the sociodemographic characteristics of other students in the school. School inputs include measures of the teachers (education level, experience, sex, race, and so forth), of the school organization (class sizes, facilities, administrative expenditures, and so forth), and of district or community factors (for example, average expenditure levels). Except for the original Coleman Report, most empirical work has relied on data constructed for other purposes, such as the normal administrative records of

[13] One exception is found in work on economics education where students' knowledge of economics has been investigated in a variety of contexts. See, for example, John Seigfried and Rendigs Fels 1979; John Chizmar and Thomas Zak 1983.

[14] A few miscellaneous issues about output measurement deserve passing attention. First, if one does use test score measurements, there are a number of choices, related simply to the scaling of scores. Tests are often available in "grade level" equivalent, percentile ranking, or raw score forms, all of which provide the same ordinal ranking (except for the possibility of some compression of the rankings). Yet, for most statistical work, one wants a scale that indicates how different individuals function rather than one that simply ranks them. The choice really depends upon the relationship of these estimates of output to the subsequent outcomes and is best seen as a special case of more general questions about the functional form of production functions. Second, there is some movement toward criterion-references tests—tests that relate to some set of educational goals. The crucial issue is the development of goals. The previous discussion argues for goals that relate to performance outside of schools, but it is not obvious that these goals guide much of the current test development work. See also Congressional Budget Office (1986a) for a discussion of different kinds of tests and of validation techniques.

schools that might be supplemented in some manner.

As in most areas of empirical analysis, a wide variety of approaches to estimation exist. Some have concentrated on variations in individual student achievement (for example, Hanushek 1971; Richard Murnane 1975; Anita Summers and Barbara Wolfe 1977), while others have looked at aggregate performance across school buildings or districts (for example, Herbert Keisling 1967; Jesse Burkhead 1967; Byron Brown and Daniel Saks 1975; Frederick Sebold and William Dato 1981). Similarly, studies have both concentrated on variations within a single system (for example, David Armor et al. 1976; Stephen Michelson 1972; Donald Winkler 1975) and on variations across districts (for example, Marshall Smith 1972; Jencks and Brown 1975; John Heim and Lewis Perl 1974). Estimation has largely been done by single equation regression, but a number of studies have gone into simultaneous equation estimation (for example, Anthony Boardman, Otto Davis, and Peggy Sanday 1977; Elchanan Cohn 1975; and Henry Levin 1970).

Each of these approaches has both strengths and weaknesses, and each is helpful in answering some questions but not others. Because the details of these specifications are discussed and critiqued elsewhere (Hanushek 1979), the focus in this discussion is on two fundamental options in analysis. The first is whether estimation is conducted in "level" form or in "value-added" form; the second is whether teacher differences are measured explicitly or implicitly.

Two pervasive problems arise when an achievement measure is simply regressed on a series of available inputs. First, adequate measures of innate abilities have never been available. Second, while education is cumulative, frequently only contemporaneous measures of inputs are available, leading to measurement and specification errors. Each of these problems leads to biases in the estimated effects of educational inputs.[15] The latter problem, the imprecise characterization of the stream of educational inputs, is probably the more severe one in terms of biased estimation of school policy factors, but both are potentially important.[16]

[15] There are of course other data and estimation problems, but they are more idiosyncratic both in their appearance and their solution. Perhaps the most common issue not discussed here is the imprecise measurement of the specific school resources relevant to individual students at a given point in time. This problem, which is most severe in individual versus aggregate school estimation, occurs because schools are quite heterogeneous institutions offering a diversity of inputs to specific students, and the exact provision for individuals is often not recorded or available. The answer is straightforward (one should measure inputs more precisely) if unappealing in specific research situations (see Hanushek 1979).

[16] In a regression framework, the effect of omitting an important variable is bias in the estimated regression coefficients with the size of bias being related both to the influence of the variable on achievement and the correlation of the omitted variable with other included variables in the model; see Hanushek and John Jackson 1977. Because it is reasonable to assume that innate abilities are positively correlated with family background (both through genetics and environment), omission of innate abilities probably biases upward the estimated impact of family background on achievement. Yet, because the correlations between innate abilities and school attributes, after allowing for family background factors, is likely to be small, biases there are probably much less.

In terms of historical school inputs, because students regularly change teachers and schools, current inputs are frequently very inaccurate indicators of past inputs. This is also a problem with measuring peer inputs, particularly in the case of integration and the racial composition of schools. Because of student migration, abrupt changes in racial composition through court or administratively ordered desegregation, and other similar factors, the current racial mix may not indicate history.

One attempt to analyze the effect of historical errors in measurement induced by using purely cross sectional measures of school characteristics can be found in Daniel Luecke and Noel McGinn 1975. Most of their evidence pertains, unfortunately, to the estimation of simple correlations or the analysis of variance of the type included in the Coleman Report. By their simple simulations, which are not replicated across different samples, estimated regression coefficients are reasonably close to the theoretically correct values.

Both problems are also helped if one uses the "value-added" versus "level" form in estimation. That is, if the achievement relationship holds at different points in time, it is possible to concentrate on exactly what happens educationally between those points when outcomes are measured. For example, we could consider just the difference in achievement between two different years. This difference in achievement can then be related to the specific inputs over the same, more limited period.[17] Similarly, the importance of these omitted factors (such as innate abilities) is lessened if the model is estimated in value-added form, because any "level" effects have already been included through entering achievement and only "growth" effects of innate abilities have been omitted. (See Boardman and Murnane, 1979, for a discussion of potential biases in alternative specifications.)

For the most part, value-added estimation has been possible only when outputs have been measured by standardized test scores. This results simply from data availability, because a one-shot data collection effort using school records can still yield intertemporal information through the history contained in normal records (see Hanushek 1971; Murnane 1975; Armor et al. 1976; Summers and Wolfe 1977; Murnane and Phillips 1981).

The second "strategic" issue in estimation is how to characterize teacher and school inputs. In many ways, the most natural approach is to identify a parsimonious set of variables depicting the central inputs, and policy decisions, in the schools. Plausible descriptors of schools include such things as class sizes, backgrounds and experiences of teachers, use of particular curricula, expenditures on administration, and so forth. Indeed, this has been the mode of analysis for the vast majority of studies. It does, however, face a potentially severe problem, although one quite common: If the choice of inputs does not include the most important ones or if the inputs have an inconsistent effect on performance,[18] the regression estimates will be difficult to interpret. But education differs from most other areas in that an alternative is available that provides direct information about the two potential problems.

With large data samples that provide multiple observations of students with the same teacher, it is possible to estimate teacher effects *implicitly, instead of explicitly.* In particular, if one had a sample of "otherwise identical" students who differed only in the teachers that they had, a direct estimate of the effectiveness of

[17] In actual analyses, however, it is generally preferable to include the initial achievement measure as one of the inputs, instead of differencing the dependent variable. There are three reasons for doing this: (1) empirically, output measurements, particularly test scores in different grades, may be scaled differently; (2) levels of starting achievement may influence achievement gain; and (3) correlated errors in achievement measurement may suggest such a formulation (Lee Cronbach and Lita Furby 1970). However, the latter argument suggests that further corrections for errors in the exogenous variables—probably based upon test reliability measures—are also needed because such errors, even if they have zero means, will yield inconsistent estimates (see Hanushek 1986). The relationship between model specification and errors in measurement is discussed extensively by David Rogosa and John Willett (1985).

This general formulation of the "value-added" specification lessens the data requirements, but it does so at the expense of some additional assumptions about the relationships. This approach would suffer if prior inputs had a lasting effect over and above any effect on initial achievement levels. This is, for example, one interpretation that could be given to some of the analyses of preschool programs where persistent and long lasting outcome differences are observed even though early IQ effects of preschool disappear. The evidence is, however, very indirect; see Richard Darlington et al. (1980) and John Berrueta-Clement (1984).

[18] Some work in education suggests an inconsistency of effect arising from interactions among different factors. For example, if teachers with a particular background are effective in suburban schools but ineffective in urban schools, simple linear specifications that force common effects across different circumstances might yield very misleading results. Other similar examples, or hypotheses, abound in the educational literature.

each teacher would be the average performance of all the students each teacher taught. Obtaining samples of identical students is clearly impossible, but statistical analysis can be used to adjust for differences among students.

Consider test score performance. The idea is that a teacher can be judged on the basis of the average test scores of her students, but only after "correcting" for differences in the achievement of the students that occurred *before* the teacher had the particular group of students and only after correcting for differences in education that occur *outside* the classroom. This can be done by estimating a regression model that includes measures of prior achievement of students, family backgrounds, and so forth and that also includes a separate intercept for all students with a specific teacher. Such teacher-specific intercepts, which can be estimated by including a dummy variable for each teacher, are interpreted as the mean achievement of the students of a given teacher after allowing for other differences among the students.[19] This approach allows the implicit evaluation of the effectiveness of teachers while avoiding the requirement of providing a detailed specification of the separate characteristics of teachers that are important.

The approach, which we will call the study of "total teacher effects," does present its own problems. This estimation provides fundamentally less information than a completely specified explicit model, because it is not possible to characterize the kinds of teachers or teaching techniques that are most effective. It also presents sizable data requirements, which are only infrequently met. Estimation must be conducted in a value-added form to insure that the estimates provide information about teacher effectiveness as opposed to classroom assignment of students or other, nonteacher aspects of education. Further, if all students for a given teacher are together in the same class (such as in the case of a traditional elementary school where students stay with the same teacher for all subjects), the estimates indicate the combined effect of the teacher and the specific classroom composition. Therefore, interpretation of such estimates as just the effectiveness of teachers requires additional information or estimation work.[20] It is also difficult to provide interdistrict estimates, making this approach less suited to addressing any district level policy matters.[21] Nevertheless, in those studies where the approach has been applied (Hanushek 1971; Murnane 1975; Armor et al. 1976; and Murnane and Phillips 1981), important new information has resulted (as described in the following section).

As in other areas of empirical research, compromises are frequently necessary be-

[19] Actual estimation can be done in a variety of ways such as through a general covariance program or by differencing all variables from their teacher-specific mean. Some care is required, however, because these estimation techniques are frequently developed for balanced designs, that is, equal numbers of students for all teachers. See Hanushek and John Quigley 1985.

[20] It is possible to include characteristics of the students in the classrooms in the estimation, as long as one can find explicit measures of the characteristics. One attempt at doing so (Hanushek 1971) confirmed that teacher differences were much more important than any measured differences in classroom composition. Another way to disentangle teachers from other classroom characteristics would be to consider the stability of estimated teacher effects over time and across classrooms. Unfortunately, little such work is available. See Hanushek 1986 and, from a different perspective, Rosenshine 1970.

[21] Because of the extensive data requirements, such estimation across districts has not been possible. Even with the required data, the estimation would have to measure any important interdistrict aspects and, without a sizable number of different districts, would not be able to provide very reliable estimates of their independent effect on student achievement. It would, nonetheless, be useful to attempt such a study because it would help validate the results; these value-added studies, while much better than level studies in terms of complete data, still are subject to overall concerns about model specification.

tween what is conceptually desirable and availability of data. Because analysis proceeds on the basis of statistical investigations of "natural experiments," the precise specification and statistical methodology can directly affect the results, and controversy over the interpretation of results, such as with the Coleman Report, must therefore be put within the context of the underlying conceptual model. Frequently, educational production functions are interpreted as if the variables included are conceptually correct and accurately measured, when in fact this may not be the case. The severity of such problems, however, differs significantly from study to study and clearly explains part of the apparent inconsistencies in specific findings.

III. *Results*

Since the Coleman Report in 1966, some 147 separately estimated educational production functions have appeared in the published literature. While varying in focus, in methodology, and ultimately in quality, these estimates provide a number of insights into schools and school policy.[22]

A. *Do Teachers Differ?*

Since the publication of *Equality of Educational Opportunity,* the Coleman Report, intense debate has surrounded the fundamental question of whether schools and teachers are important to the educational performance of students. This debate follows naturally from the Coleman Report, which is commonly interpreted as finding that variations in school resources explain a negligible portion of the variation in students' achievement.[23] If true, this would indicate that it did not matter which particular teacher a student had—something most parents at least would have a difficult time accepting.

A number of studies provide direct analyses of this overall question of differential effectiveness of teachers through the estimation of total teacher effects (described above). The findings of these studies (Hanushek 1971, 1986; Murnane 1975; Armor et al. 1976; and Murnane and Phillips 1981) are unequivocal: *Teachers and schools differ dramatically in their effectiveness.*

While a number of implications and refinements of that work still need addressing, this conclusion is very firm. It also gives a very different impression from that left by the Coleman Report and indeed by a number of subsequent studies. These faulty impressions have primarily resulted from a confusion between the difficulty in explicitly measuring components of effectiveness and true effectiveness. In other words, existing measures of characteristics of teachers and schools are seriously flawed and thus are poor indicators of the true effects of schools; when these measurement errors are corrected, schools are seen to have important effects on student performance.

B. *Summary of Expenditure Relationships*

While it is important to confirm that teachers differ in effectiveness, it would be more desirable to be able to identify the specific aspects and characteristics of

[22] Other reviews and interpretations of this work can be found in James Guthrie et al. 1971; Harvey Averch et al. 1974; R. Gary Bridges, Charles Judd, and Peter Moock 1979; Murnane 1981b; Naftaly Glasman and Israel Biniaminov 1981; and Murnane and Nelson 1984.

[23] The Coleman Report concentrates on explained variance in student achievement. Its conclusions about school effects come directly from noting that the increase in explained variance (R^2) is small when school variables are added to a regression equation already containing other educational inputs. Such results are obviously sensitive to the order in which various inputs are added to the equation (Hanushek and Kain 1972).

teachers and schools that are important. In approaching this question, scholars have disagreed about the factors that should be explicitly measured and included as inputs into the educational production process. However, there is a "core" set of factors—those that determine basic expenditures—that is almost universally investigated. Instructional expenditures make up about two-thirds of total school expenditures. Given the number of students in a school district, instructional expenditures are in turn determined mostly by teacher salaries and class sizes. Finally, most teacher salaries are directly related to years of teaching experience and educational levels completed by the teacher. Thus the basic determinants of instructional expenditures in a district are teacher experience, teacher education, and class size, and most studies, regardless of what other descriptors of schools might be included, will analyze the effect of these factors on outcomes. (These are also the factors most likely to be found in any given data set, especially if the data come from standard administrative records.)

Because of this commonality in specification, it is possible to tabulate easily the effects of these expenditure parameters. An (attempted) exhaustive search uncovers 147 separate "qualified studies" found in 33 separate published articles or books.[24] These studies, while restricted just to public schools, cover all regions of the country, different grade levels, different measures of performance, and different analytical and statistical approaches.[25] While some of these factors could lead to differences in results, they are ignored in the overall tabulations of results.[26]

Table 8 presents overall tabulations for the 147 studies. Because not all studies include each of the expenditure parameters, the first column in Table 8 presents the total number of studies for which an input can be tabulated—for example, 112 (of the 147) studies provide information about the relationship between teacher-student ratio and student performance. The available studies provide regression estimates of the partial effect of given inputs, holding constant family background and other inputs. These estimated coefficients have been tabulated according to two pieces of information: the sign and the statistical significance (5 percent level) of the estimated relationship.

According to conventional wisdom, each tabulated factor should have a positive effect on student achievement. More education and more experience on the part of the teacher both cost more and are presumed to be beneficial; smaller classes (more teachers per student) should also improve individual student learning.[27] Having the "correct" sign in a pro-

[24] A qualified study is defined as a production function estimate: (1) published in a book or refereed journal; (2) relating some objective measure of student output to characteristics of the family and the schools attended; and (3) providing information about the statistical significance of estimated relationships. A given publication can contain more than one estimated production function by considering different measures of output, different grade levels, or different samples of students (but different specifications of the same basic sample and outcome measure are not duplicated). This is an expanded version of tabulations in Hanushek 1981.

[25] The studies are almost evenly divided between studies of individual student performance and aggregate performance in schools or districts. Ninety-six of the 147 studies measure output by score on some standardized test. Approximately 40 percent are based upon variations in performance within single districts while the remainder look across districts. Three-fifths look at secondary performance (grades 7–12) with the rest concentrating on elementary student performance. Added descriptive information about the universe of studies can be found in Hanushek 1981.

[26] Subsequent analysis does not suggest any bias from looking at all of the studies together. While there are obvious limits to the possible stratifications of the separate studies, further analyses that grouped studies by grade level, by whether individual or aggregate data are used, by measure of output, and so forth yield the same qualitative conclusions.

[27] Tabulated results are adjusted for variables being measured in the opposite direction; for example, the sign for estimated relationships including student-teacher ratios is reversed.

TABLE 8
SUMMARY OF ESTIMATED EXPENDITURE PARAMETER COEFFICIENTS FROM 147 STUDIES OF EDUCATIONAL PRODUCTION FUNCTIONS

Input	Number of Studies	Statistically Significant +	Statistically Significant −	Statistically Insignificant Total	Statistically Insignificant +	Statistically Insignificant −	Statistically Insignificant Unknown Sign
Teacher/pupil ratio	112	9	14	89	25	43	21
Teacher Education	106	6	5	95	26	32	37
Teacher Experience	109	33	7	69	32	22	15
Teacher Salary	60	9	1	50	15	11	24
Expenditures/pupil	65	13	3	49	25	13	11

Sources: Armor et al. 1976; Richard Beiker and Kurt Anschek 1973; Boardman, Davis, and Sanday 1977; Bowles 1970; Brown and Saks 1975; Burkhead 1967; Cohn 1968, 1975; Eberts and Stone 1984; Hanushek 1971, 1972; Heim and Perl 1974; Henderson, Mieszkowski, and Sauvageau 1976; Jencks and Brown 1975; Katzman 1971; Keisling 1967; Levin 1970, 1976; Link and Ratledge 1979; Maynard and Crawford 1976; Michelson 1970, 1972; Murnane 1975; Murnane and Phillips 1981; Perl 1973; Raymond 1968; Ribich and Murphy 1975; Sebold and Dato 1981; Smith 1972; Strauss and Sawyer 1986; Summers and Wolfe 1977; Tuckman 1971; Winkler 1975. See, Hanushek 1981 for further description of studies.

duction function is clearly a minimal requirement for justifying purchases of a given input, but quantitative magnitudes of estimated relationships are ignored here.[28]

Of the 112 estimates of the effects of class size, only 23 are statistically significant, and *only* 9 show a statistically significant relationship of the expected positive sign.[29] Fourteen display a statistically significant *negative* relationship. An additional 89 are not significant at the 5 percent level. Nor does ignoring statistical significance help to confirm benefits of small classes, because the insignificant coefficients have the "wrong" sign by a 43 to 25 margin.[30]

The entries for teacher education and teacher experience in Table 8 tell much the same story. In a majority of cases, the estimated coefficients are statistically insignificant. Forgetting about statistical significance and just looking at estimated signs does not make much of a case for the importance of these factors either.

The one possible exception—teacher experience—at least has a clear majority of estimated coefficients pointing in the expected direction, and almost 30 percent of the estimated coefficients are statistically significant by conventional standards. If experience is really a powerful factor in teaching, however, these results

[28] It would be extremely difficult to provide information of quantitative differences in the coefficients because the units of measure of both inputs and outputs differ radically from one study to another. One attempt to provide quantitative estimates of varying class sizes is Gene Glass and Mary Lee Smith (1979). This work, however, has been subjected to considerable criticism, largely because of the ultimate difficulties in doing such analyses.

[29] Teacher/pupil ratios are treated here as being synonymous with class sizes. This is not strictly the case and, in fact, could be misleading today. Several changes in schools, most prominently the introduction of extensive requirements for dealing with handicapped children in the mid-1970s, have led to new instructional personnel without large changes in typical classes. Because much of the evidence here refers to the situation prior to such legislation and restrictions, it is reasonable to interpret the evidence as relating to class sizes.

[30] Note that not all studies report the sign of insignificant coefficients. For example, 21 studies report insignificant estimated coefficients for teacher-student ratios but do not report any further information.

are hardly overwhelming. Moreover, because of possible selection effects, they are subject to additional interpretive questions. Specifically, these positive correlations may result from more senior teachers having the ability to select schools and classrooms with better students. In other words, causation may run from achievement to experience and not the other way around.[31]

The results are startlingly consistent in finding no strong evidence that teacher-student ratios, teacher education, or teacher experience have an expected positive effect on student achievement. According to the available evidence, one cannot be confident that hiring more educated teachers or having smaller classes will improve student performance. Teacher experience appears only marginally stronger in its relationship.

The final two rows in Table 8 include summary expenditure information, teacher salaries, and expenditures per student.[32] While less frequently available, these measures—not surprisingly—provide no separate indication of a relationship between expenditures and achievement.[33] Most data do show a strongly positive simple correlation between school expenditures and achievement, but the strength of this relationship disappears when differences in family background are controlled for.

Without systematic tabulation of the results of the various studies, it would be easy to conclude that the findings of the studies are inconsistent. But there is a consistency to the results: *There appears to be no strong or systematic relationship between school expenditures and student performance.* This is the case when expenditures are decomposed into underlying determinants and when expenditures are considered in the aggregate.[34]

There are several obvious reasons for being cautious in interpreting this evidence. For any individual study, incomplete information, poor quality data, or faulty research could distort a study's sta-

[31] David Greenberg and John McCall (1974) analyzed a single urban school system in the early 1970s and concluded that race and socioeconomic background of students were systematically related to the selection and transfer of teachers with different education and experience levels. However, Murnane (1981a) suggests, from analysis of a different school system, that declining enrollments and the subsequent surplus of teachers have led to a much greater reliance on institutional rules and much less on individual teacher preferences (which was the hypothesized mechanism in Greenberg and McCall 1974).

Nevertheless, the potential problems arise from achievement affecting selection, and not from family background, race, or other factors that are included on the right-hand side of the estimated model affecting selection. Clearly the severity of the problem is related to the structure of the model estimated and in many instances is serious only in the presence of fairly subtle selection mechanisms (particularly in a "value-added" specification).

[32] Information on salaries and expenditures is less frequently available. Importantly, because expenditures per student are generally measured for districts, any of the 60 analyses for individual districts would find no variation in this input and thus could not include it. Further, the interpretation of both of these measures is sometimes clouded by including them *in addition to* teacher experience, education, and/or class size.

[33] The expenditure and salary estimates are generally more difficult to interpret than the other, real resource measures. Because the prices can vary across the samples in the separate studies, it is sometimes difficult to interpret the dollar measures. Are they indicators of quality differences? of price differences? of costs that vary with the characteristics of the city and students (that is, of "compensating differentials" for various undesirable characteristics)?

In the expenditure estimates, 8 of 13 significant positive results also come from the different estimates of Sebold and Dato (1981). These estimates involve aggregate school districts in California and, importantly, involve very imprecise measures of family backgrounds. For lower grades, a socioeconomic index compiled by teachers is employed; for the higher grades (8 out of 10 separate estimates) the percentage of families on Aid for Dependent Children (AFDC) is the only measure available. With this imprecise measurement, school expenditures may in fact be a proxy for family background.

[34] This also holds up when the sample of available studies is divided along different dimensions: the measurement of outcomes (i.e., test score versus other measures); elementary versus secondary; single system versus multiple systems; value-added versus level; and so forth. While the precise tabulations obviously change with the smaller subsamples, the overall picture remains.

tistical results. Even without such problems, the actions of school administrators could mask any relationship. For example, if the most difficult to teach students were consistently put in smaller classes, any independent effect of class size could be difficult to disentangle from mismeasurement of the characteristics of the students. Finally, statistical insignificance of any estimates can reflect no relationship, but it also can reflect a variety of data problems—those above and others such as high correlations among the different measured inputs. In other words, as in most research efforts, virtually any of the studies is open to some sort of challenge.

Just such uncertainties about individual results led to this tabulation of estimates. If these specific factors were in fact central to variations in student achievement, the tabulations would almost certainly show more of a pattern in the expected direction. The reasons for caution listed above are clearly more important in some circumstances than others, and the inconsistency across these very different studies is still striking. Furthermore, given the general biases toward publication of statistically significant estimates, the paucity of statistically significant results is quite notable. While individual studies are affected by specific analytical problems, the aggregate data provided by the 147 separate estimates seem most consistent with the conclusion that the expenditure parameters are unrelated to student performance (after family backgrounds and other educational inputs are considered).

C. *Other Results*

In the course of these analyses, a wide variety of other school and nonschool factors have been investigated. First, family background is clearly very important in explaining differences in achievement. Virtually regardless of how measured, more educated and more wealthy parents have children who perform better on average.[35] One particularly interesting subset of these analyses, however, involves investigating more detailed aspects of family structure and size. The large changes in birth rates and divorce rates of the past two decades have created a concern about their potential effects on learning and achievement. Analyses of these issues unfortunately have not been undertaken in any systematic manner within the context of educational production functions.[36]

Second, considerable attention has been given to the characteristics of peers or other students within schools. This line of inquiry was pressed by the Coleman Report and pursued by a number of subsequent studies.[37] This question is especially important in considering school desegregation where the issues revolve around the racial compositions of schools. The educational effect of differing student bodies has also been important in the debate about public versus private schooling, as discussed below. Nevertheless, the findings are ambiguous.

Finally, a wide range of additional measures of schools and teachers has been pursued in the different existing studies. Various studies have included indicators of

[35] There have been vast quantities of studies concentrating on the effects of family background. Unfortunately, few such studies include measures of school factors. Exceptions are Murnane, Rebecca Maynard, and James Ohls (1981) and Hanushek (1986).

[36] General discussions and reviews of the issues can be found in Richard Easterlin (1978) and Samuel Preston (1984). For the most part, these ignore influences of schools on achievement, although it may not be too problematical in a time series context. A preliminary investigation of family factors based upon simple time allocation models can be found in Hanushek (1986).

[37] See, in particular, Hanushek 1972; Winkler 1975; Summers and Wolfe 1977; and Vernon Henderson, Peter Mieszkowski, and Yvon Sauvageau 1976. Part of the ambiguity about the results arises from the possibility of confusing measures of peers with the influence of family background through measurement errors in family characteristics; see Hanushek and Kain 1972. In terms of the public-private school debate, see Coleman, Thomas Hoffer, and Sally Kilgore 1982; Murnane 1984.

organizational aspects of schools, of specific curricula or educational process choices, and of such things as time spent by students working at different subject matters. Others have included very detailed information on teachers—their cognitive abilities, family backgrounds, where they went to school, what their majors were, their attitudes about education or different kinds of students, and so forth. Similarly detailed information has been included about school facilities and school administrators and other personnel. The closest thing to a consistent finding among the studies is that "smarter" teachers, ones who perform well on verbal ability tests, do better in the classroom, but even for that the evidence is not very strong (Hanushek 1981).[38]

D. *Teacher Skill Differences*

In the typical study of production relationships outside of education, measures of organization and process are seen as irrelevant in estimation. Production functions are interpreted as the relationship between inputs and outputs *mutatis mutandis*. Information about production possibilities is viewed as being publicly available in the form of scientific and engineering knowledge, and production processes are reproducible through blueprints and machinery. The possibility of the actors in production making dynamic choices about process is not considered, and the choice of "best" process is assumed to be automatically made after the selection of inputs. While the appropriateness of this framework is open to question in a wide number of instances, it is particularly questionable in the case of education.

Some aspects of the educational process are inherently difficult to disentangle from the characteristics of individual teachers (such as classroom management, methods of presenting abstract ideas, communication skills, and so forth). This creates serious problems both in applying the general conceptual model of production theory and in interpreting any estimated effects. Many educational decisions are "micro" ones, made mainly by teachers, and they are difficult to observe and measure and, quite possibly, not easy to reproduce. Further, these decisions interact with the characteristics and abilities of the individual teacher. As a shorthand description, these factors will be referred to simply as "skill" differences.[39]

Once the possibility of skill differences in introduced, it is difficult to define just what "maximum possible output" might mean because it is difficult to specify what the "homogeneous" inputs are. In other words it is difficult if not impossible to specify a few objective or subjective characteristics of teachers that capture the systematic differences of both backgrounds of teachers and their idiosyncratic choices of teaching style and methods.[40] This

[38] Many states currently require standardized testing of teachers, either for initial or continuing employment. There is little evidence that the commonly used teacher examinations provide much evidence about effectiveness at teaching. See, however, Strauss and Sawyer 1986. Further, if one thought of routinely using test information, such as scores on the verbal ability tests available to researchers, to determine hiring and salary, teachers would most likely concentrate more on the tests, thus lessening any correlations between test performance and teaching skill.

[39] A formal model that captures many of these ideas is presented in Anthony Lima 1981. The concept of skill in production also appears in Richard Nelson and Sidney Winter 1982.

This kind of notion also appears in the explanation for not finding any systematic relationship between process and organizational choices of schools and achievement. The explanation of the apparent insignificance of macro process variables in Armor et al. (1976) is the great variation in implementation of overall process decisions at the classroom level. This is also brought out in detailed analysis of the implementation of innovative techniques at the classroom level (see Paul Berman and Milbrey McLaughlin 1975).

[40] Part of the general specification issue can be found in other situations. For example, measurement of capital stocks of varying vintages clearly aggregates over heterogeneous inputs and therefore intro-

raises questions, discussed below, about how much of standard production theory is usable without some modification.

The empirical implications are that individual variables describing certain partial aspects of teacher skill are unlikely to display systematic relationships with student performance (which is our measure of the performance of teachers). This is just the interpretation of the previously presented results. Individual teacher skill differences are quite important, as estimated implicitly and discussed above. But, teacher skill is not systematically correlated with the explicit measures of teacher characteristics that have been available.[41] Again, the consequences of not measuring teacher inputs explicitly should not be mistaken for the ineffectiveness of teachers.

An important sidelight of such investigations is that decision makers might be able to identify with fair accuracy underlying differences in skills among teachers. Murnane (1975) and Armor et al. (1976) find that principals' evaluations of teachers were highly correlated with estimates of total effectiveness (that is, adjusted mean gains in achievement by the students of each teacher). For many purposes, this is almost as good as the ability to identify differences among teachers ex ante.

Recognition of skill differences does alter the interpretation of teacher and school inputs. It is still reasonable to consider the impact of measured attributes of teachers, because many school decisions such as hiring and salary are based on a set of these characteristics. The estimated impacts of these measured attributes, however, indicate the inability either to predict or develop more skilled teachers according to the attributes identified. Consider, for example, the almost universal finding that graduate education of teachers bears no systematic relationship to achievement, which can be interpreted as indicating that current teacher training institutions do not, on average, change the skills of teachers. This is somewhat different from saying that *everything* else being equal, more education for teachers has no effect. Similarly, the frequent finding that class size doesn't affect achievement may arise from complicated (and unobserved) interactions with the processes and instructional methods that teachers choose. Therefore, while it is possible that smaller classes could be beneficial in specific circumstances, it is also true that, in the context of typical school and teacher operations, there is no apparent gain.

The concept of teacher skill differences has clear implications for research. At least a part of past research development can be characterized as a search for "the" factor or specification that unifies other results or that at least explains the apparent inconsistencies for specific factors. But if teaching skill involves mixing different objective and subjective characteristics together, sometimes in very different ways across individuals, the search for a simply articulated and measured description of

duces error into any estimation of production relationships. However, in such a case one could at least conceptually provide more detailed measurements of the capital stock and eliminate the problems. The situation considered here is more complicated, because the inputs (teachers) are also the managers of the classroom, deciding how to organize the educational experience, how to employ their own education and experiences, and so on. There are perhaps similarities to labor inputs into other production processes, but the argument here is that there is a difference in degree of autonomy and choice exercised by the teacher. There is also a difference in the specificity of the analysis of education, which takes this analysis to a deeper level than most aggregate production functions.

[41] In individual studies, it appears that roughly only half of total teacher performance, estimated as described above by adjusted average performance of a teacher's students, can be explained by any combination of measured teacher and classroom attributes. (Such studies include more extensive measures of teachers than just the expenditure parameters found in Table 8.) See in particular Hanushek (1972) and Murnane and Phillips (1981).

effective teachers or schools is likely to fail.

E. *Efficiency in Schools*

If we think of schools as maximizing student achievement, the preceding evidence indicates that schools are *economically inefficient*, because they pay for attributes that are not systematically related to achievement. This statement, of course, presumes that schools are attempting to maximize student performance. While such motivation seems reasonable to assume, complicated objectives on the part of school officials would lead to tempering this judgment.[42]

A suggestion of inefficiency on the part of public schools of course does not come as a great surprise to many for two reasons. First, educational decision makers apparently not guided by incentives to maximize profits or to conserve on costs.[43] Second, they may not understand the production process and therefore cannot be expected to be on the production frontier. In other words, much of the optimization part of the theory of the firm and competitive markets is questionable in the case of governmental supply in quasi-monopoly situations.[44]

While few people would go so far as to say that school expenditures could not have an important effect on performance, it is at the same time possible to conclude that expenditures are unrelated to school performance as schools are currently operated. The fact that a school spends a lot on each of its students simply gives us little information on whether or not it does well in terms of value added to students.

It is, however, useful to be clear about the issues of efficiency and what can be inferred from the data on schools. Past education discussions have blurred any distinction between economic efficiency (the correct choice of input mix given the prices of inputs and the production function) and technical efficiency (operating on the production frontier). The previous evidence relates directly to economic efficiency. The consideration of technical efficiency is more complicated.

The standard conceptual framework indicates that, if two production processes are using the same inputs, any systematic difference in outputs reflects technical

[42] The studies reviewed previously do consider a wide range of measures of student performance. Therefore, a simple objection to test scores as representing the focus of attention by school officials does not suffice to overturn this conclusion.

There are two aspects of "nonmaximizing behavior" that have been analyzed within the context of educational production functions. First, a number of researchers have considered multiple objectives of school officials and have analyzed simultaneous equations models of production. These analyses do not come to qualitatively different conclusions from those presented. Second, a few have attempted to evaluate explicitly the impact of school preferences for specific outcomes. Michelson (1970) considers preferences for different outcomes that vary across schools and suggests that this could obscure relationships estimated for single dimensions. Brown and Saks (1975) consider a model where schools are interested in both the mean and variance of student achievement, although all schools have the same objective function. They suggest again that analyses of just mean test score performance could be biased by the unobserved preferences of districts. Unfortunately, because little information is available about preferences other than performance maximization on the part of schools, it is very difficult to evaluate their influence on the measured efficiency of schools.

[43] It should be pointed out that similar analyses of production functions for private, profit-making industries are not readily available. We are prone to accept without real evidence that for-profit firms are optimizing such that a tabulation of results for competitive firms would look different from Table 8. We at least know that for-profit firms that are not maximizing are more likely to go out of existence than a public enterprise that is not maximizing.

[44] It must be noted that economic inefficiency does not preclude estimation of production functions. Indeed, such inefficiency aids estimation (at least when done directly and not through cost functions) because it provides observations of the technical relationships under different input mixes. If all schools faced the same prices and operated efficiently, there would be no variation in the data, and estimation would not be possible.

inefficiency.[45] The concept of skill differences, however, simply recognizes that individuals with the same measured characteristics make a series of important production decisions (reflected in behavior, process choices, and so forth) that are difficult to identify, measure, and model. Therefore, it is not surprising that the same measured inputs yield variations in output; at the same time, it is difficult to label such observed variation differences inefficiency.[46] In part the argument is one of semantics: How much of economic theory should implicitly be brought along in analyses of production functions? There are, however, obvious implications for policy and research in terms of the interpretation of findings and the ability to operate on achievement by changing the observed attributes of teachers and schools.[47]

IV. *Some Policy Implications*

The conclusion that schools are operated in an economically inefficient manner has obvious implications for school policy. The clearest one is simply that increased expenditures by themselves offer no overall promise for improving education. Further, the components of these expenditures offer little promise. Thus, a simple recommendation: Stop requiring and paying for things that do not matter.

There is little apparent merit for schools to pursue their ubiquitous quest for lowered class sizes.[48] Nor should teachers be required, as they are in many states, to pursue graduate courses merely to meet tenure requirements or to get an additional salary increment. More teacher experience by itself does not seem to have much value.

Each of these statements also has its limits. The evidence for them comes from the current operations of public schools. Yet, policies that take schools outside the bounds observed could lead to different results. For example, class sizes between 15 and 40 students fall well within the data; classes of 2 students or 300 students do not—and they may show significant relationships with achievement. The evidence is also limited to overall, systematic relationships. Quite clearly, small classes might be very beneficial in certain circumstances, depending on the teachers and the subject matter; if there is specific evidence of this, one should clearly act differently. The point is that we have no evidence of this universally, and thus we have no mandate for making massive changes just to be doing something.

Pay of teachers offers another set of pol-

[45] Concern about technical inefficiency has led to some, basically nonstatistical, estimation (cf. Dennis Aigner and S. Chu 1968) of the production frontier. Different applications to educational production can be found in Levin (1976) and, with a somewhat different approach, Robert Klitgaard and George Hall (1975). Besides assuming accurate measures of both inputs and outputs, these analyses appear internally inconsistent: They are motivated by the perceived uncertainty about the production process, yet assume that the researcher knows and measures all of the inputs to the production process; see Hanushek (1976). Further, the possibility of nonreproducible skill differences is totally neglected.

[46] The importance of embodied process differences leads Murnane and Nelson (1984) to argue that the whole concept of production functions may not have much usefulness in education and other areas where the actors tacitly make many production decisions. The standard treatment of production functions is clearly strained by the necessity to observe and measure choices of classroom presentation, organization of materials, interactions with students, and so forth; that is, things that constitute how real inputs of teacher's knowledge, experiences, and other characteristics are put together in the production process. One could of course expand the simple notion of production functions to include such matters, but, since our current ability to identify and measure such expanded inputs is quite poor, this would not provide much guidance to empirical analysis.

[47] Note that this discussion is quite different from the consideration of "X-inefficiency" (Harvey Leibenstein 1966). That discussion is best interpreted as simply omitting an important factor that might simply be labeled entrepreneurial ability. This discussion goes deeper into the measurement and specification of production functions as a generic model.

[48] One pervasive and extremely expensive trend in American education has been the progressive lowering of student-teacher ratios. See Table 6, above, for the recent history of declines.

icy issues. Two issues receive constant attention: the level of pay and the distribution of pay.[49] We begin with the second issue, the one most directly addressed by the available research. In most school systems, salary schedules are rigidly linked to the education levels completed by the teacher and years of teaching experience. Salary is unrelated to specialty—math teachers are paid the same as English teachers—or to grade level. Is there an alternative, given that this structure does not appear to correlate very closely with productivity? Recent commission reports have increasingly called for instituting "merit pay," an idea that has been around for decades but that has defied widespread implementation.[50] Relating pay to performance is a key element of the comprehensive reforms suggested by the Task Force on Teaching (1986).

The previous evidence suggests that a merit pay system would make sense. It is clear that significant differences exist among teachers. And, while not conclusive, direct tests that correlate estimates of specific teachers' value added[51] with principals' evaluations of the same teachers suggest that principals do reasonably well at identifying good and bad teachers (Murnane 1975; Armor et al. 1976). Thus, the essential elements of merit pay schemes seem present.

The main argument used against merit pay is that objective evaluation is difficult and thus there is always the possibility that political and other influences may creep into pay determination. There is little direct evidence from schools related to this possibility. Of course, the pay of most other workers in the economy is at least partially determined by their supervisors, and there are not obvious reasons to believe that employment relationships in schools are unique.

The more difficult problem is to introduce such a system and get it working. First, the current pay system might be a classic illustration of the inflexible rules that are said to characterize internal labor markets, and they certainly have the effect of reducing any direct competition among teachers. This in turn promotes collaboration among teachers, which might suffer if teachers perceived themselves to be in competition. Second, principals seem to be able to identify good teachers when nothing is at stake, but whether they would make such judgments if their evaluations mattered is unknown. Third, a restructuring of pay would lead to direct conflict with teachers' unions. With little experience and analysis of these issues, however, there is no way to judge their importance.

The second aspect of pay is its overall level. Many people have argued that the rewards of teaching are so low that it is little wonder that the best graduates are not attracted to teaching. Others have used evidence of shortages of particular kinds of teachers—for example math and science teachers—to argue for general pay increases to teachers.

There is clearly no absolute standard for setting teachers' pay, although there is fre-

[49] When considering the distribution of pay, it is useful to hold average level constant so as to avoid direct incentive effects on choice of teaching as a profession. For example, if teacher experience is not systematically related to performance, it is not efficient to pay more for greater experience; this does not mean, however, that all teachers should receive the salary of a beginning teacher because that would dramatically lower the lifetime earnings of the average teacher. Clearly, it is not possible to separate these issues completely, but it is useful to deal with the components of level and distribution separately.

[50] The idea of merit pay enjoyed a brief period of national discussion after it was recommended by the National Commission on Excellence in Education in *A Nation at Risk* and after this idea was promoted by the President Reagan. See David Cohen and Murnane (1985, 1986) for a recent discussion of issues in the implementation of merit pay systems. Many school systems have tried "merit pay" systems, but most attempts have not lasted very long.

[51] Value added is measured, as described above, by estimates of the average gain in student performance (adjusted for factors unrelated to the teacher) that are associated with specific teachers.

quently an appeal to some notion of "comparable worth." Average real salaries of teachers rose during the 1960s to the mid-1970s. Indeed they rose faster than average salaries in other parts of the economy. After that, they slid back as did the real income of the average worker. By 1983 the average (nine month) salary for teachers was \$20,700; in the same year the median income for a male (female) year-round, full-time worker with four or more years of college was \$31,800 (\$20,251).[52] Salaries in schools have remained between 15 and 20 percent above the earnings of the average full-time employee in the economy over time. Whether this is too high or too low is difficult to judge.

Raising all salaries would almost certainly attract more able people into teaching. But three factors must be borne in mind. First, the ability to alter the teaching force is constrained by vacancies at schools. Somewhat less than 3.5 percent of all teachers in 1983 were in either their first or second year of teaching.[53] If there is a lag between choosing a profession in college and becoming trained for it (cf. Richard Freeman 1971) and if future turnover remains at current levels, it would be many years after changes in overall salaries took place before any significant change in the teacher force could be discerned. (Because of changing age patterns in the teaching profession, retirements and thus turnover will undoubtedly increase over the coming decade. Therefore, current rates probably underestimate the potential for change. See Table 7, above.)

Second, as argued by Murnane (1985) current restraints imposed by state certification requirements, inhibit the flow of new people into the profession. These requirements, frequently stated in terms of specific courses, practice teaching time, and so forth, act as an entry barrier to many potential teachers who find that their course of study would be noticeably distorted. (Radical changes in both teacher training and teacher certification are central to reform proposals in Task Force in Teaching, 1986, and the Holmes Group 1986.)

Third, if the salary structure takes into account no information about competing demands for specialties (of which math and science have received the most attention for the past two decades), considerable inefficiency must always be present: Either people in "low demand" areas will be overpaid when compared to what is needed to insure sufficient supply into teaching, or teachers in "high demand" specialties will tend to be of lower quality than those in low demand specialties. (This observation has, of course, been made previously; see Joseph Kershaw and Roland McKean 1962.)

The entire area of state certification and educational regulations is open to considerable question, particularly given the evidence above. While there is wide variation in the specifics, states tend to require teachers to pursue graduate degrees—a dubious restriction given the evidence about lack of effectiveness and an expensive one because school systems then pay these teachers more. By 1983, over half of all teachers had a master's degree or more, up from less than a third only a

[52] The information on salaries in Table 7 included all instructional personnel, in part because principals, guidance counselors, and other school people outside the classroom often start out as classroom teachers. (Classroom teachers make up about 90 percent of all instructional personnel.) Salaries of this larger group seem relevant to someone contemplating a teaching career.

Note that the median incomes include more than just wages and salary. U.S. Bureau of the Census 1985, p. 141; U.S. Bureau of the Census 1985, pp. 158, 162.

[53] The teacher force has gone through large growth and depression periods, related to the demographics of the school-age population. Based on the numbers of children under five (that is, already born but not in schools), there will be some growth in the school-age population, but not an enormous amount.

decade before.[54] States also set tenure rules, with tenure coming as early as the third year of teaching. And, in a number of programs, states either set explicit class size maximums or provide monetary incentives to have smaller class sizes. None of these practices seems very useful from a public policy view related to student achievement. Instead their primary justification must come in terms of compensating teachers or restricting the supply of teachers.

Many restrictions on hiring, promotions, and so forth also are found in contracts and local regulations (see Lorraine McDonnell and Anthony Pascal 1979). These have a similar inhibiting effect on schools, although it seems possible to eliminate the more harmful ones through the bargaining process. The effects of unions on salaries, expenditures, and other employment conditions is, however, incompletely understood; see, for example, the review by Freeman (1986) and analyses by Jay Chambers (1977) and Randall Eberts and Joe Stone (1984, 1985, 1986).

Finally, along similar lines, it is useful to consider the financing of local school systems. There are again a great many different financing schemes by which states support local schools. Beginning in the late 1960s, local reliance on property taxes and state distribution schemes that did not counteract differences in property tax bases became an active area of judicial attention, legislative concern, and academic research. The 1968 California court case of *Serrano* v. *Priest* opened a virtual outpouring of studies, legal suits, and legislative bills.[55] In simplest terms, a general equity argument was made that some districts, those with larger tax bases, found it easier to raise money for schools than districts with lower tax bases. As a result, expenditures per student tended to be quite unequal across jurisdictions. The research into expenditure variations across local school districts and their causes has been extensive (see, for example, Robert Berne and Leanna Steifel 1983; Martin Feldstein 1975; Robert Reischauer and Robert Hartman 1973; David Stern 1973; Robert Inman 1978; John Coons, William Clune, and Stephen Sugarman 1970; John Pincus 1974).

Much of this discussion appears motivated by an underlying assumption that poor districts (in terms of property tax bases) are the same as poor students. This, it turns out, is not uniformly the case.[56] But, more than that, the discussion is based entirely on a presumption that expenditures per student are the appropriate focus for policy. Without this presumption, an unwarranted one from the previous evidence, the line of argument—

[54] As noted above, such comparisons require some caution in interpretation because the age and experience of the teaching force has changed over time.

[55] The legal issues took several years to be sorted out. Originally suits were brought under the Fourteenth Amendment to the U.S. Constitution on the grounds that students were being discriminated against "on the basis of the wealth of their neighbors" because the size of the property tax base directly influenced how much money could be raised and spent in the local schools. The U.S. Constitution arguments were not accepted by the U.S. Supreme Court in its 1973 decision in *Rodriguez* v. *San Antonio Independent School District.* Subsequently, a number of successful (and a number of unsuccessful) suits were brought in state courts under education clauses of state constitutions. See John Coons, William Clune, and Stephen Sugarman 1970; John Pincus 1974; James Guthrie 1980; Walter Garms, Guthrie, and Lawrence Pierce 1978.

[56] As an example, New York City and the other large cities of New York State had to develop a new argument, "municipal overburden," in order to join the property-poor plaintiffs in their state school finance suit of *Levittown* v. *Nyquist.* This argument—that large cities had disproportionate other demands on their resources—allowed them to enter on the side of the plaintiffs in suing the state for increased state financing. This was necessary because these large cities tended to have the largest tax bases per student in the state; they also have a disproportionate share of poor people.

legal and academic—becomes quite peculiar.[57]

One might argue that altering existing financing formulae would have only distributional consequences, because expenditure variations do not relate to the performance of different school systems. But this is not the only effect. The politics of redistribution tend to promote increases in total spending on schools. States find it difficult to lower funding for one district in order to raise it for another, and therefore they tend to raise low spending districts up to the level of high spending districts. (This probably explains the general support by teachers unions for school finance "reform.") The responses of states to challenges to their funding of schools are thus frequently to increase the amount of economic inefficiency in the system.

A final policy area that is closely related but not precisely covered by the research discussed above is the public versus private school debate. All of the evidence presented previously relates to public school systems. Perhaps as a response to perceptions that public schools need improving, a variety of measures have been proposed to encourage further private school competition. The notion of educational vouchers, originally proposed by Milton Friedman (1962), has always had some appeal to economists because it would promote more individual choice and competition. A recent version of this—tuition tax credits—has received the endorsement of President Ronald Reagan and has appeared in his budgetary proposals; it would effectively encourage private schooling through the federal income tax system. Nevertheless, private schools have not been subjected to much direct analysis.

A recent study by Coleman, Thomas Hoffer, and Sally Kilgore (1982) has again brought the issue into discussion. This study basically contrasted the performance of students in public and private schools and concluded that private schools systematically performed better than public schools. This conclusion has been the subject of intense debate, one that remains unresolved. There are two basic questions: First, are the results simply a reflection of selectivity bias arising from parents' choice of school type? Second, does the control of schools (private versus public) identify the most important differences among the sampled schools? The study attempts to measure and to control for a series of background measures of students, but many critics have argued that it does so imprecisely (Arthur Goldberger and Cain 1982; Murnane 1984; Jay Noell 1982). Also, the study makes no attempt to describe the specific characteristics of schools and teachers in either the public or private setting. Therefore, the policy conclusions rest importantly on having a random sample of schools and being able to replicate the private school success through a policy of expanding the private sector. In this area, the evidence is very incomplete.

Individual studies of the educational production process frequently point to other specific conclusions about policies. Nevertheless, because many of these conclusions appear only once and are not replicated in other studies, it does not appear useful to develop them in detail. Instead, given the current state of research, it is appropriate to stop with these general observations.

[57] It is possible to interpret the issue as one of "taxpayer equity" instead of "educational equity." The tax rate that is needed to achieve any given funding level for schools does vary widely across districts, and it is particularly sensitive to the amount of nonresidential property in the local tax base. This line of argument has a distinct legal problem associated with it, because most judicial challenges to school funding formulae have arisen from specific mention of educational concerns in state constitutions. Tax equity does not enjoy the same legal status.

V. *Some Research Implications*

The analyses of schools obviously raise a number of unanswered questions that could profit from more research. Because many of these have been discussed elsewhere (see, for example, Hanushek 1979, 1981; Murnane and Nelson 1984), another set of research questions is raised here. Namely, what do these findings have to say about other lines of research by economists?

Through two decades of research, an enormous amount has been learned about the empirical application of production function notions to educational policy questions. Much of this clearly is transferable to other areas—for example, health programs or agricultural production. Perhaps most important is the lesson about evaluation of activities where the idiosyncratic nature of the actors can be key to the results. In a great many areas, particularly ones related to public policy matters, it is necessary to evaluate production efficiency and this, in turn, frequently calls for the analysis of individual skill differences. In these, the straightforward econometric design may yield quite misleading results.

But beyond such areas, one must also consider how the results of the educational analysis relate to analyses of the effects of schooling. In particular, a wide variety of public finance investigations implicitly or explicitly consider how differences in education and schools affect some other types of behavior.

Following the theoretical work of Charles Tiebout (1956) and the empirical work of Wallace Oates (1969), a number of studies have investigated how differences in the attractiveness of jurisdictions come to be capitalized into the price of houses. A substantial portion of these studies has focused on differences in the provision of public services, of which schooling is the most important local one. With few exceptions,[58] the level of schooling provided is given by expenditures, and this is contrasted with the local tax cost of providing such expenditures. But if expenditures per student are not an accurate index of educational provision, this does not adequately capture locational differences.[59] Studies of housing location have also tended to make similar presumptions.

Investigations of labor market performance of individuals have, in their quest to include individual quality differences, used a variety of measures of schooling such as expenditures or characteristics of teachers in given schools (see above). Again, these do not appear to be good indicators of schooling differences.[60]

Finally, the signaling versus production models of schools represents an area where the preceding analysis is most appropriate. Empirical analyses of screening have typically looked for labor market tests of the competing hypotheses. Both models, however, imply higher earnings by more educated people: the screening

[58] Examples of exceptions include Kain and Quigley (1975) and Harvey Rosen and D. J. Fullerton (1977).

[59] Of course, changing school finance formulae, which would have the effect of helping some jurisdictions and hurting others, would affect housing values immediately—unless there were offsetting changes in school quality. See, for example, John Hilley (1983) and Donald Jud and James Watts (1981).

[60] It is important, however, to distinguish the above evidence from the potential use of expenditures in such analyses. The above evidence indicates that expenditures per student do not do particularly well at indicating the value added of schools. On the other hand, by themselves (that is, without controlling for any other factors) expenditures per student are quite generally correlated with student achievement. This results from the fact that higher income families tend both to pay more for schools and to provide more educational input in the homes. Therefore, expenditure differences do tend to measure differences in student achievement, which is what is needed for labor market studies; they just cannot be interpreted as indicating the importance of schools per se. This is quite different from the preceding discussion of capitalization, where the conceptual factor is the value added of local schools and not just the overall performance of students.

model through the information provided about differential abilities, and the production model through changing the abilities of individuals. While some ingenious tests have been proposed (see, for example, Wolpin 1977; Riley 1979), these necessarily fail because the models predict that the observations of individuals in the labor market will have the same basic character. (In technical terms, they are generally unidentified.) The models differ significantly, however, when one looks at the schooling process itself. The signaling version assumes that individuals are basically unaffected by school experience—they simply wait and endure schooling until the information about abilities catches up with their actual abilities. The production model suggests that the schooling experience changes individuals. At least in the polar cases, the weight of available evidence on schools suggests that the production model is more appropriate, because where students end up is strongly affected by the schools they attend. This conclusion breaks down, of course, if one holds to a "mixed" model of schools, because there is no way to make judgments about the absolute differences that come out of the process. Nonetheless, evidence about school production seems most appropriate for addressing these hypotheses.

REFERENCES

AEROSPACE RESEARCH CENTER. *Meeting technology and manpower needs through the industry/university interface: An aerospace industry perspective.* Washington, DC: Aerospace Industries Assoc. of Amer., May 1983.

AIGNER, DENNIS J. AND CHU, S. F. "On Estimating the Industry Production Function," *Amer. Econ. Rev.*, Sept. 1968, *58*(4), pp. 826–39.

AKIN, JOHN S. AND GARFINKEL, IRWIN. "School Expenditures and the Economic Returns to Schooling," *J. Human Res.*, Fall 1977, *12*(4), pp. 460–81.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, AND NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION. *Standards for educational and psychological testing.* Washington, DC: American Psychological Assoc., 1985

ARMOR, DAVID ET AL. *Analysis of the school preferred reading program in selected Los Angeles minority schools.* R-2007-LAUSD. Santa Monica, CA: Rand Corp., 1976.

ASHENFELTER, ORLEY AND MOONEY, JOSEPH D. "Graduate Education, Ability, and Earnings," *Rev. Econ. Statist.*, Feb. 1968, *50*(1), pp. 78–86.

AVERCH, HARVEY A. ET AL. *How effective is schooling? A critical review and synthesis of research findings.* Englewood Cliffs, NJ: Educational Technology Publications, 1974.

BARNOW, BURT S. AND CAIN, GLEN G. "A Reanalysis of the Effect of Head Start on Cognitive Development: Methodology and Empirical Findings," *J. Human Res.*, Spring 1977, *12*(2), pp. 177–97.

BECKER, GARY S.; LANDES, ELISABETH M. AND MICHAEL, ROBERT T. "An Economic Analysis of Marital Instability," *J. Polit. Econ.*, Dec. 1977, *85*(6), pp. 1141–47.

BEHRMAN, JERE R. AND BIRDSALL, NANCY. "The Quality of Schooling: Quantity Alone Is Misleading," *Amer. Econ. Rev.*, Dec. 1983, *73*(5), pp. 928–46.

BEIKER, RICHARD F. AND ANSCHEK, KURT R. "Estimating Educational Production Functions for Rural High Schools: Some Findings," *Amer. J. Agr. Econ.*, Aug. 1973, *55*, pp. 515–19.

BEN-PORATH, YORAM. "The Production of Human Capital Over Time," in *Education, income, and human capital.* Ed.: W. LEE HANSEN. NY: NBER, 1970, pp. 129–47.

BERG, IVAR. *Education and jobs: The great training robbery.* NY: Praeger, 1970.

BERMAN, PAUL AND MCLAUGHLIN, MILBREY W. *Federal programs supporting educational change, Vol. IV: The findings in review.* R-1589/4-HEW. Santa Monica, CA: Rand Corp., 1975.

BERNE, ROBERT AND STIEFEL, LEANNA. "Changes in School Finance Equity: A National Perspective," *J. Educ. Finance,* Spring 1983, *8*, pp. 419–35.

BERRUETA-CLEMENT, JOHN R. ET AL. *Changed lives: The effects of the Perry preschool program on youths through age 19.* Ypsilanti, MI: The High/Scope Press, 1984.

BOARDMAN, ANTHONY; DAVIS, OTTO AND SANDAY, PEGGY. "A Simultaneous Equations Model of the Educational Process," *J. Public Econ.*, Feb. 1977, *7*(1), pp. 23–49.

BOARDMAN, ANTHONY AND MURNANE, RICHARD. "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement," *Sociology Educ.*, Apr. 1979, *52*, pp. 113–21.

BOWLES, SAMUEL. "Toward an Educational Production Function," in *Education, income and human capital.* Ed.: W. LEE HANSEN. NY: NBER, 1970, pp. 11–60.

BOWLES, SAMUEL AND GINTIS, HERBERT. *Schooling in capitalist America.* NY: Basic Books, 1976.

BOWLES, SAMUEL AND LEVIN, HENRY M. "The Determinants of Scholastic Achievement—An Appraisal of Some Recent Evidence," *J. Human Res.*, Winter 1968, *3*(1), pp. 3–24.

BRELAND, HUNTER M. *Population validity and col-*

lege entrance measures. NY: The College Board, 1979.

BRENEMAN, DAVID W. "The Ph.D. Production Process," in JOSEPH T. FROOMKIN, DEAN T. JAMISON AND ROY RADNER, 1976, pp. 3–52.

BRIDGES, R. GARY; JUDD, CHARLES M. AND MOOCK, PETER R. *The determinants of educational outcomes: The impact of families, peers, teachers, and schools.* Cambridge, MA: Ballinger, 1979.

BROWN, BYRON W. AND SAKS, DANIEL H. "The Production and Distribution of Cognitive Skills within Schools," *J. Polit. Econ.*, June 1975, *83*(3), pp. 571–93.

BURKHEAD, JESSE. *Input-output in large city high schools.* Syracuse, NY: Syracuse U. Press, 1967.

BUSINESS-HIGHER EDUCATION FORUM. *America's competitive challenge: The need for a national response.* Washington, DC: Business-Higher Education Forum, Apr. 1983.

CAIN, GLEN G. AND WATTS, HAROLD W. "Problems in Making Policy Inferences from the Coleman Report," *Amer. Soc. Rev.*, Apr. 1970, *35*(2), pp. 328–52.

CENTRAL ADVISORY COUNCIL FOR EDUCATION (ENGLAND). *Children and their primary schools.* London: Her Majesty's Stationery Office, 1967.

CHAMBERS, JAY G. "The Impact of Collective Bargaining for Teachers on Resource Allocation in Public School Districts," *J. Urban Econ.*, July 1977, *4*(3), pp. 324–39.

CHIZMAR, JOHN F. AND ZAK, THOMAS A. "Modeling Multiple Outputs in Educational Production Functions," *Amer. Econ. Rev.*, May 1983, *73*(2), pp. 18–22.

COHEN, DAVID K. AND MURNANE, RICHARD J. "The Merits of Merit Pay," *The Public Interest*, Summer 1985, *80*, pp. 3–30.

———. "Merit Pay and the Evaluation Problem: Understanding Why Most Merit Pay Plans Fail and a Few Survive," *Harvard Educ. Rev.*, Feb. 1986, *56*(1), pp. 1–17.

COHN, ELCHANAN. "Economies of Scale in Iowa High School Operations," *J. Human Res.*, Fall 1968, *3*(4), pp. 422–34.

COHN, ELCHANAN, WITH MILLMAN, STEPHEN D. *Input-output analysis in public education.* Cambridge, MA: Ballinger, 1975.

COLEMAN, JAMES S. ET AL. *Equality of educational opportunity.* Washington, DC: U.S. GPO, 1966.

COLEMAN, JAMES S.; HOFFER, THOMAS AND KILGORE, SALLY. *High school achievement: Public, Catholic, and private schools compared.* NY: Basic Books, 1982.

CONGRESSIONAL BUDGET OFFICE. *Trends in educational achievement.* Washington, DC: Congressional Budget Office, 1986a.

CONGRESSIONAL BUDGET OFFICE. *Educational achievement: Explanations and implications of recent trends.* Washington, DC: Congressional Budget Office, 1986b.

COONS, JOHN E.; CLUNE, WILLIAM H. AND SUGARMAN, STEPHEN D. *Private wealth and public education.* Cambridge, MA: The Belknap Press of Harvard U. Press, 1970.

CRONBACH, LEE J. AND FURBY, LITA. "How Should We Measure 'Change'—or Should We?" *Psych. Bull.*, July 1970, *74*(1), pp. 68–80.

DARLINGTON, RICHARD B. ET AL. "Preschool Programs and Later School Competence of Children from Low-Income Families," *Science*, April 11, 1980, *208*, pp. 202–04.

DENISON, EDWARD F. *Accounting for United States economic growth 1929–1969.* Washington, DC: Brookings Inst., 1974.

DUGAN, DENNIS J. "Scholastic Achievement: Its Determinants and Effects in the Education Industry," in JOSEPH T. FROOMKIN, DEAN T. JAMISON AND ROY RADNER, 1976, pp. 53–83.

EASTERLIN, RICHARD A. "What will 1984 Be Like? Socioeconomic Implications of Recent Twists in Age Structure," *Demography*, Nov. 1978, *15*(4), pp. 397–421.

EBERTS, RANDALL W. AND STONE, JOE A. *Unions and public schools: The effect of collective bargaining on American education.* Lexington, MA: Lexington Books, 1984.

———. "Wages, Fringe Benefits, and Working Conditions: An Analysis of Compensating Differentials," *Southern Econ. J.*, July 1985, *52*(1), pp. 274–79.

———. "On the Contract Curve: A Test of Alternative Models of Collective Bargaining," *J. Labor Econ.*, Jan. 1986, *4*(1), pp. 66–71.

EDUCATION COMMISSION OF THE STATES, TASK FORCE ON EDUCATION FOR ECONOMIC GROWTH. *Action for excellence: A comprehensive plan to improve our nation's schools.* Denver, CO: Education Commission of the States, June 1983.

EHRLICH, ISAAC. "On the Relation Between Education and Crime," in *Education, income, and human behavior.* Ed.: F. THOMAS JUSTER. NY: McGraw-Hill, 1975, pp. 313–37.

FELDSTEIN, MARTIN S. "Wealth Neutrality and Local Choice in Public Education," *Amer. Econ. Rev.*, Mar. 1975, *65*(1), pp. 75–89.

FREEMAN, RICHARD B. *The market for college trained manpower.* Cambridge, MA: Harvard U. Press, 1971.

———. "Unionism Comes to the Public Sector," *J. Econ. Lit.*, Mar. 1986, *14*(1), pp. 41–86.

FRIEDMAN, MILTON. *Capitalism and freedom.* Chicago: U. of Chicago Press, 1962.

FROOMKIN, JOSEPH T.; JAMISON, DEAN T. AND RADNER, ROY, eds. *Education as an industry.* Cambridge, MA: Ballinger, 1976.

FULLER, BRUCE. *Raising school quality in developing countries: What investments boost learning?* Report No. EDT7, Education and Training Series, The World Bank, Sept. 1985.

GARMS, WALTER I.; GUTHRIE, JAMES W. AND PIERCE, LAWRENCE C. *School finance: The economics and politics of public education.* Englewood Cliffs, NJ: Prentice-Hall, 1978.

GLASMAN, NAFTALY S. AND BINIAMINOV, ISRAEL. "Input-Output Analyses in Schools," *Rev. Educ. Res.*, Winter 1981, *51*(4), pp. 509–39.

GLASS, GENE V. AND SMITH, MARY LEE. "Meta-Analysis of Research on Class Size and Achievement," *Educ. Eval. and Policy Anal.*, 1979, *1*(1), pp. 2–16.

GOLDBERGER, ARTHUR S. AND CAIN, GLEN G. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgore Report," *Sociology Educ.*, Apr./Jul. 1982, *55*, pp. 103–22.

GREENBERG, DAVID AND MCCALL, JOHN. "Teacher Mobility and Allocation," *J. Human Res.*, Fall 1974, *9*(4), pp. 480–502.

GRILICHES, ZVI AND MASON, WILLIAM. "Education, Income, and Ability," *J. Polit. Econ.*, May/June 1972, *80*(3), Pt. 2, pp. S74–S103.

GROSSMAN, MICHAEL. "The Correlation Between Health and Schooling," in *Household production and consumption.* Ed.: NESTOR E. TERLECKYJ. NY: NBER, 1973, pp. 147–211.

GUTHRIE, JAMES W., ed. *School finance policies and practices.* Cambridge, MA: Ballinger, 1980.

GUTHRIE, JAMES W. ET AL. *Schools and inequality.* Cambridge, MA: MIT Press, 1971.

HANSEN, W. LEE; WEISBROD, BURTON AND SCANLON, WILLIAM J. "Schooling and Earnings of Low Achievers," *Amer. Econ. Rev.*, June 1970, *60*(3), pp. 409–18.

HANUSHEK, ERIC A. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," *Amer. Econ. Rev.*, May 1971, *61*(2), pp. 280–88.

———. *Education and race: An analysis of the educational production process.* Cambridge, MA: Heath-Lexington, 1972.

———. "Regional Differences in the Structure of Earnings," *Rev. Econ. Statist.*, May 1973, *55*(2), pp. 204–13.

———. "Comment," in JOSEPH T. FROOMKIN, DEAN JAMISON AND ROY RADNER, 1976, pp. 191–96.

———. "Ethnic Income Variations: Magnitudes and Explanations," in *American ethnic groups.* Ed.: THOMAS SOWELL. Washington, DC: Urban Inst., 1978, pp. 139–56.

———. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *J. Human Res.*, Summer 1979, *14*(3), pp. 351–88.

———. "Throwing Money at Schools," *J. Policy Anal. Manage.*, Fall 1981, *1*(1), pp. 19–41.

———. "The Trade-off Between Child Quantity and Quality: Some Empirical Evidence." Working Paper No. 45, Rochester Center for Economic Research, 1986.

HANUSHEK, ERIC A. AND JACKSON, JOHN E. *Statistical methods for social scientists.* NY: Academic, 1977.

HANUSHEK, ERIC A. AND KAIN, JOHN F. "On the Value of 'Equality of Educational Opportunity' as a Guide to Public Policy," in *On equality of educational opportunity.* Eds.: FREDERICK MOSTELLER AND DANIEL P. MOYNIHAN. NY: Random House, 1972, pp. 116–45.

HANUSHEK, ERIC A. AND QUIGLEY, JOHN M. "Life-Cycle Earning Capacity and the OJT Investment Model," *Int. Econ. Rev.*, June 1985, *26*(2), pp. 365–85.

HAUSE, JOHN C. "Earnings Profile: Ability and Schooling," *J. Polit. Econ.*, May/June 1972, *80*(3), Pt. II, pp. S108–38.

HAVEMAN, ROBERT H. AND WOLFE, BARBARA L. "Schooling and Economic Well-Being: The Role of Nonmarket Effects," *J. Human Res.*, Summer 1984, *19*(3), pp. 377–407.

HEIM, JOHN AND PERL, LEWIS. *The educational production function: Implications for educational manpower policy.* Institute of Public Employment Monograph 4. Ithaca, NY: Cornell U., 1974.

HENDERSON, VERNON; MIESZKOWSKI, PETER AND SAUVAGEAU, YVON. *Peer group effects and educational production functions.* Ottawa, Canada: Economic Council of Canada, 1976.

HEYNEMAN, STEPHEN P. AND LOXLEY, WILLIAM. "The Effect of Primary-School Quality on Academic Achievement Across Twenty-nine High and Low-Income Countries," *Amer. J. Sociology*, May 1983, *88*, pp. 1162–94.

HILLEY, JOHN. "The Distributive Impact of Education Finance Reform," *Nat. Tax J.*, Dec. 1983, *36*(4), pp. 503–9.

HOGAN, TIMOTHY D. "Faculty Research Activity and the Quality of Graduate Training," *J. Human Res.*, Summer 1981, *16*(3), p. 400–15.

HOLMES GROUP. *Tomorrow's teachers.* East Lansing, MI: The Holmes Group, 1986.

INMAN, ROBERT P. "Optimal Fiscal Reform of Metropolitan Schools: Some Simulation Results," *Amer. Econ. Rev.*, Mar. 1978, *68*(1), pp. 107–22.

JENCKS, CHRISTOPHER S. AND BROWN, MARSHA. "Effects of High Schools on Their Students," *Harvard Educ. Rev.*, Aug. 1975, *45*(3), pp. 273–324.

JENCKS, CHRISTOPHER ET AL. *Inequality: A reassessment of the effects of family and schooling in America.* NY: Basic Books, 1972.

JOHNSON, GEORGE E. AND STAFFORD, FRANK P. "Social Returns to Quantity and Quality of Schooling," *J. Human Res.*, Spring 1973, *8*(2), pp. 139–55.

JUD, G. DONALD AND WATTS, JAMES M. "Schools and Housing Values," *Land Econ.*, Aug. 1981, *57*, pp. 459–70.

KAIN, JOHN F. AND QUIGLEY, JOHN M. *Housing markets and racial discrimination.* NY: NBER, 1975.

KATZMAN, MARTIN. *Political economy of urban schools.* Cambridge, MA: Harvard U. Press, 1971.

KEISLING, HERBERT. "Measuring a Local Government Service: A Study of Efficiency of School Districts in New York State." Unpublished PhD thesis, Harvard U., 1965.

———. "Measuring a Local Government Service: A Study of School Districts in New York State," *Rev. Econ. Statist.*, Aug. 1967, *49*, pp. 356–67.

KERSHAW, JOSEPH A. AND MCKEAN, ROLAND N. *Teacher shortages and salary schedules.* NY: McGraw-Hill, 1962.

KLITGAARD, ROBERT E. AND HALL, GEORGE R. "Are There Unusually Effective Schools?" *J. Human Res.*, Winter 1975, *10*(1), pp. 90–106.

KOLLODGE, RICHARD AND HORN, ROBIN. *Education research and policy studies at the World Bank: A bibliography.* Report No. EDT23, Education and Training Series, The World Bank, Dec. 1985.

LAYARD, RICHARD AND PSACHAROPOULOS,

GEORGE. "The Screening Hypothesis and the Returns to Education," *J. Polit. Econ.*, Sept./Oct. 1974, *82*(5), pp. 985–98.

LEIBENSTEIN, HARVEY. "Allocative Efficiency vs. 'X-Efficiency,'" *Amer. Econ. Rev.*, June 1966, *56*(3), pp. 392–415.

LEIBOWITZ, ARLEEN. "Home Investments in Children," *J. Polit. Econ.*, Mar./Apr. 1974, *82*(2), Pt. II, pp. S111–31.

LEVIN, HENRY M. "A New Model of School Effectiveness," in *Do teachers make a difference?* U.S. Office of Education. Washington, DC: U.S. Dept. of Health, Education, and Welfare, 1970, 55–78.

______. "Economic Efficiency and Educational Production," in JOSEPH T. FROOMKIN, DEAN JAMISON AND ROY RADNER. Cambridge, MA: NBER, 1976, pp. 149–90.

LIMA, ANTHONY K. "An Economic Model of Teaching Effectiveness," *Amer. Econ. Rev.*, Dec. 1981, *71*(5), pp. 1056–59.

LINK, CHARLES R. AND RATLEDGE, EDWARD C. "Social Returns to Quantity and Quality of Education: A Further Statement," *J. Human Res.*, Winter 1975, *10*(1), pp. 78–89.

______. "Student Perceptions, I.Q., and Achievement," *J. Human Res.*, Winter 1979, *14*(1), pp. 98–111.

LUECKE, DANIEL F. AND MCGINN, NOEL F. "Regression Analyses and Education Production Functions: Can They Be Trusted?" *Harvard Educ. Rev.*, Aug. 1975, *45*(3), pp. 325–50.

MAYNARD, REBECCA AND CRAWFORD, DAVID. "School Performance," in *Rural income maintenance experiment final report.* Madison, WI: Inst. for Research on Poverty, U. of Wisconsin, 1976.

MCDONNELL, LORRAINE AND PASCAL, ANTHONY. *Organized teachers in American Schools*, R-2407-NIE. Santa Monica, CA: Rand Corp., 1979.

MICHAEL, ROBERT T. "Measuring Non-monetary Benefits of Education: A Survey," in *Financing education: Overcoming inefficiency and inequity.* Ed.: WALTER W. MCMAHON AND TERRY G. GESKE. Urbana, IL: U. of Illinois Press, 1982, pp. 119–49.

MICHELSON, STEPHEN. "The Association of Teacher Resources with Children's Characteristics," in U.S. Office of Education. *Do Teachers Make a Difference?* Washington, DC: U.S. GPO, 1970, pp. 120–68.

______. "For the Plaintiffs—Equal School Resource Allocation," *J. Human Res.*, Summer 1972, *7*(3), pp. 283–306.

MINCER, JACOB. "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," *J. Econ. Lit.*, Mar. 1970, *8*(1), pp. 1–26.

MURNANE, RICHARD J. *Impact of school resources on the learning of inner city children.* Cambridge, MA: Ballinger, 1975.

______. "Teacher Mobility Revisited," *J. Human Res.*, Winter 1981a, *16*(1), pp. 3–19.

______. "Interpreting the Evidence on School Effectiveness," *Teachers College Record*, Fall 1981b, *83*(1), pp. 19–35.

______. "How Clients' Characteristics Affect Organization Performance: Lessons from Education," *J. Policy Anal. Manage.*, Spring 1983, *2*(3), pp. 403–17.

______. "A Review Essay—Comparisons of Public and Private Schools: Lessons from the Uproar," *J. Human Res.*, Spring 1984, *19*(2), pp. 263–77.

______. "An Economist's Look at Federal and State Education Policies," in *American domestic priorities: An economic appraisal.* Eds.: JOHN M. QUIGLEY AND DANIEL L. RUBINFELD. Berkeley: U. of California Press, 1985, pp. 118–47.

MURNANE, RICHARD J.; MAYNARD, REBECCA A. AND OHLS, JAMES C. "Home Resources and Children's Achievement," *Rev. Econ. Statist.*, Aug. 1981, *63*(3), pp. 369–77.

MURNANE, RICHARD J. AND NELSON, RICHARD R. "Production and Innovation when Techniques Are Tacit," *J. Econ. Behavior Org.*, 1984, *5*, pp. 353–73.

MURNANE, RICHARD J. AND PHILLIPS, BARBARA. "What do Effective Teachers of Inner-City Children Have in Common?" *Social Sci. Res.*, Mar. 1981, *10*(1), pp. 83–100.

NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION. *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. GPO, Apr. 1983.

NELSON, RICHARD R. AND PHELPS, EDMUND. "Investment in Humans, Technology Diffusion, and Economic Growth," *Amer. Econ. Rev.*, May 1966, *56*(2), pp. 69–75.

NELSON, RICHARD R. AND WINTER, SIDNEY G. *An evolutionary theory of economic change.* Cambridge, MA: The Belknap Press of Harvard U. Press, 1982.

NIEMI, RICHARD AND SOBIESZEK, BARBARA I. "Political Socialization," *Annual Rev. Sociology*, 1977, *3*, pp. 209–33.

NOELL, JAY. "Public and Catholic Schools: A Reanalysis of Public and Private Schools," *Sociology Educ.*, Apr./Jul. 1982, *55*, pp. 123–32.

OATES, WALLACE. "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis," *J. Polit. Econ.*, Nov./Dec. 1969, *77*(6), pp. 957–71.

O'NEILL, DAVE M. AND SEPIELLI, PETER. *Education in the United States: 1940–1983*, CDS-85-1, Special Demographic Analysis. Washington, DC: U.S. Bureau of the Census, 1985.

PERL, LEWIS J. "Family Background, Secondary School Expenditure, and Student Ability," *J. Human Res.*, Spring 1973, *8*(2), pp. 156–80.

PERL, LEWIS J. "Graduation, Graduate School Attendance, and Investments in College Training," in JOSEPH T. FROOMKIN, DEAN T. JAMISON AND ROY RADNER, 1976, pp. 95–135.

PINCUS, JOHN, ED. *School finance in transition.* Cambridge, MA: Ballinger, 1974.

PRESTON, SAMUEL H. "Children and the Elderly: Divergent Paths for America's Dependents," *Demography*, Nov. 1984, *21*(4), pp. 435–57.

RAYMOND, RICHARD. "Determinants of the Quality of Primary and Secondary Public Education in West Virginia," *J. Human Res.*, Fall 1968, *3*(4), pp. 450–70.

REISCHAUER, ROBERT O. AND HARTMAN, ROBERT W. *Reforming school finance.* Washington, DC: Brookings Inst., 1973.

RIBICH, THOMAS I. AND MURPHY, JAMES L. "The Economic Returns to Increased Educational Spending," *J. Human Res.*, Winter 1975, *10*(1), pp. 56–77.

RILEY, JOHN G. "Informational Equilibrium," *Econometrica*, Mar. 1979a, *47*(2), pp. 331–59.

———. "Testing the Educational Screening Hypothesis," *J. Polit. Econ.*, Oct. 1979b, Pt II, *87*(5), pp. S227–52.

ROGOSA, DAVID R. AND WILLETT, JOHN B. "Understanding Correlates of Change by Modeling Individual Differences in Growth," *Psychometrika*, June 1985, *50*(2), pp. 203–28.

ROGERS, DANIEL P. "Private Rates of Return to Education in the United States: A Case Study," *Yale Econ. Essays*, Spring 1969, *9*(1), pp. 89–134.

ROSEN, HARVEY S. AND FULLERTON, DAVID J. "A Note on Local Tax Rates, Public Benefit Levels, and Property Values," *J. Polit. Econ.*, Apr. 1977, *85*(2), pp. 433–40.

ROSEN, SHERWIN. "Human Capital: A Survey of Empirical Research," in *Research in labor economics*, Vol. I. Ed.: RONALD G. EHRENBERG. Greenwich, CT: JAI Press, 1977, pp. 3–39.

ROSENSHINE, BARAK. "The Stability of Teacher Effects upon Student Achievement," *Rev. Educ. Research*, Dec. 1970, *40*(5), pp. 647–62.

SEBOLD, FREDERICK D. AND DATO, WILLIAM. "School Funding and Student Achievement: An Empirical Analysis," *Public Finance Quart.*, Jan. 1981, *9*(1), pp. 91–105.

SIEGFRIED, JOHN J. AND FELS, RENDIGS. "Research on Teaching College Economics: A Survey," *J. Econ. Lit.*, Sept. 1979, *17*(3), pp. 923–69.

SMITH, MARSHALL. "Equality of Educational Opportunity: The Basic Findings Reconsidered," in *On equality of educational opportunity.* Eds.: FREDERICK MOSTELLER AND DANIEL PATRICK MOYNIHAN. NY: Random House, 1972, pp. 230–342.

SPENCE, A. MICHAEL. "Job Market Signaling," *Quart. J. Econ.*, Aug. 1973, *87*(3), pp. 355–74.

———. *Market signalling: Informational transfer in hiring and related screening processes.* Cambridge, MA: Harvard U. Press, 1974.

STERN, DAVID. "Effects of Alternative State Aid Formulas on the Distribution of Public School Expenditures in Massachusetts," *Rev. Econ. Statist.*, Feb. 1973, *55*(1), pp. 91–97.

STRAUSS, ROBERT P. AND SAWYER, ELIZABETH A. "Some New Evidence on Teacher and Student Competencies," *Econ. Educ. Rev.*, 1986, *5*(1), pp. 41–48.

SUMMERS, ANITA AND WOLFE, BARBARA. "Do Schools Make a Difference?" *Amer. Econ. Rev.*, Sept. 1977, *67*(4), pp. 639–52.

TASK FORCE ON TEACHING AS A PROFESSION. *A nation prepared: Teachers for the 21st century.* Carnegie Forum on Education and the Economy, 1986.

TAUBMAN, PAUL AND WALES, TERENCE. *Higher education and earnings.* NY: McGraw-Hill, 1974.

TIEBOUT, CHARLES M. "A Pure Theory of Local Expenditures," *J. Polit. Econ.*, Oct. 1956, *64*, pp. 416–24.

TUCKMAN, HOWARD P. "High School Inputs and Their Contribution to School Performance," *J. Human Res.*, Fall 1971, *6*(4), pp. 490–509.

TWENTIETH CENTURY FUND, TASK FORCE ON FEDERAL ELEMENTARY AND SECONDARY EDUCATION POLICY. *Making the grade.* NY: Twentieth Century Fund, 1983.

U.S. BUREAU OF THE CENSUS. *Statistical abstract of the United States, 1985.* Washington, DC: U.S. GPO, 1985.

U.S. BUREAU OF THE CENSUS. Current Population Reports. *Money income of households, families, and persons in the United States: 1983*, Ser. P-60, No. 146. Washington, DC: U.S. GPO, 1985.

U.S. DEPARTMENT OF EDUCATION. *The nation responds: Recent efforts to improve education.* Washington, DC: U.S. Department of Education, 1984.

———. *What works: Research about teaching and learning.* Washington, DC: U.S. Department of Education, 1986.

WACHTEL, PAUL. "The Effect on Earnings of School and College Investment Expenditures," *Rev. Econ. Statist.*, Aug. 1976, *58*(3), pp. 326–31.

WEISS, ANDREW. "A Sorting-cum-Learning Model of Education," *J. Polit. Econ.*, June 1983, *91*(3), pp. 420–42.

WEISS, RANDALL D. "The Effect of Education on the Earnings of Blacks and Whites," *Rev. Econ. Statist.*, May 1970, *52*(2), pp. 150–59.

WELCH, FINIS. "Measurement of the Quality of Schooling," *Amer. Econ. Rev.*, May 1966, *56*(2), pp. 379–92.

———. "Education in Production," *J. Polit. Econ.*, Jan./Feb. 1970, *78*(1), pp. 35–59.

———. "Black-White Differences in the Returns to Schooling," *Amer. Econ. Rev.*, Dec. 1973, *63*(5), pp. 893–907.

WINKLER, DONALD. "Production of Human Capital: A Study of Minority Achievement." Unpublished PhD thesis, U. of California, Berkeley, 1972.

———. "Educational Achievement and School Peer Group Composition," *J. Human Res.*, Spring 1975, *10*(2), pp. 189–204.

WIRTZ, WILLARD ET AL. *On further examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline.* NY: College Entrance Examination Board, 1977.

WOLPIN, KENNETH I. "Education and Screening," *Amer. Econ. Rev.*, Dec. 1977, *67*(5), pp. 949–58.